The University Of Sheffield.

Structure-function studies on 5' nucleases

Emma L. Brudenell

A thesis submitted in partial fulfilment of the requirements for the degree of

Doctor of Philosophy

The University of Sheffield
Faculty of Medicine, Dentistry, and Health
Department of Infection, Immunity, and Cardiovascular Disease

January, 2022

Declaration

I, the author, confirm that the Thesis is my own work. I am aware of the University's Guidance on the Use of Unfair Means (www.sheffield.ac.uk/ssid/unfair-means). This work has not been previously been presented for an award at this, or any other, university.

Abstract

The 5′ nucleases are members of the Flap endonucleases (FENs) family of structure-specific DNA-processing metalloenzymes. They fulfil essential functions DNA replication and repair. Genome integrity relies on their inherent 5′ to 3′ exonuclease and 5′ flap endonuclease activities. A variety of bifurcated nucleic acids can be processed *in vitro,* hence these enzymes have become important component of many molecular biology techniques.

Catalytic parameters for the T7 gene product 6 exonuclease (T7 gp6), a FEN family member, were determined for various substrates using a fluorescent real-time assay. Binding affinity was determined by electrophoretic mobility shift assay. In attempts to modulate relative and absolute levels of endonuclease and exonuclease activity, the impact of buffer composition and site-directed mutagenesis of residues in and around the active site was explored. The effect on catalytic activity and DNA binding affinity of mutations in the T7 gp6 active-site was studied. Three mutations retained DNA binding affinity comparable to the wild-type enzyme. Two of which were catalytically inert (Asp160Lys, Asp162Lys) whilst the third (Asp202Lys) exhibited severely reduced nuclease activity. Mutation in the potassium ion-binding helix-3-turn-helix motif (Ile200Arg) increases the specificity constant (k_{cat}/K_M) compared to wild-type for all substrates tested by at least 2-fold. Additionally, T7 gp6 Ile200Arg exhibited enhanced binding affinity for single-flap substrate but not a nicked (exonuclease) substrate, compared to wild type, suggesting the helix-3-turn helix motif contributes to differential modulation of endonuclease and exonuclease activity.

Attempts to determine the structure of T7 gp6 failed. However, a high-resolution structure (1.44 Å) of T5FEN active-site mutant Asp155Lys and two structures of *Taq polymerase* FEN domain with DNA were solved (1.82 –2.18 Å). The latter represent novel structural insights into how this important enzyme interacts with DNA.

Acknowledgements

First and foremost, I would like to thank my supervisor, Professor Jon Sayers. Thank you for providing the opportunity to do this project and for your endless help and support throughout my PhD and truly inspiring me during every conversation.

Thank you to the members of the Sayers' lab. A special thank you to Dr Sarbendra Pradhananga, for helping get me to get started at the beginning of my project and demonstrating countless new techniques to me. Thank you to Mrs Janine Phipps, for being an amazing technician. Thank you to Dr Domen Zafred, for all thing's crystallography related. Thank you to Dr Cyril Sanders, for training me in the ^{32}P work. Thank you to Dr Patrick Baker, for your assistance with my Taq structures.

Thank you to the industrial partner Atlas Genetics for providing funding towards this project. Thank you to my industrial supervisor's, Dr David Pearce and Dr Anna Dixson. Whilst my placement did not go as planned, I enjoyed the week I spent in Trowbridge regardless.

Thank you to all my friends always being there for me. A special thank you to Natalia and Nathan whom I met during my PhD and went on an amazing trip to Vietnam with, you are truly friends for life. Thank you to Kiran for being the best friend I could ask for. Thank you to Diana for always checking in on me and providing the best baked goods.

Lastly, thank you to all my family: Mum, Dad, Spike, Sarah and Teresa. Thank you for always being only a phone call away whenever my 'PhD panic' would set in and bringing Lu-Lu for cuddles when I needed it.

Table of Contents

List of figures

List of tables

Abbreviations

6-FAM	6-carboxyfluorescin
AI	Autoinduction
AIM	Autoinduction media
AEBSF	4-(2-aminoethyl)benzenesulfonyl fluoride hydrochloride
Ala (A)	Alanine
AP	apurinic/ apyrimidinic
APS	Ammonium persulfate
Arg (R)	Arginine
Asp (D)	Aspartate
AU	Asymmetric unit
Bicine	N,N-bis[2-hydroxyethyl]glycine
BER	Base excision repair
bp	Base pair
BSA	Bovine serum albumin
Cas	CRISPR associated protein
ccc	Covalently closed circular
cDNA	Complementary DNA
CRISPR	Clustered regularly interspaced short palindromic repeats
CV	Column volume
Cy3	Cyanine 3
Cys (C)	Cysteine
Da	Daltons
DMSO	Dimethyl Sulfoxide
dNMP	Deoxyribonucleotide monophosphate
ds	Double-stranded
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
EMSA	Electrophoretic mobility shift assay
EXO1	Exonuclease I
FEN	Flap endonuclease
FITC	Fluorescein isothiocyanate
FLU	Fluorescein
FRET	Förster Resonance Energy Transfer
GEN	Gap endonuclease
Gly (G)	Glycine
Gp6	Gene protein 6
H2/3TH	Helix-2/3-turn-helix
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HPLC	High-performance liquid chromatography

HR	Homologous repair
IPTG	Isopropyl β- d-1-thiogalactopyranoside
Ile (I)	Isoleucine
ITC	Isothermal titration calorimetry
k_{cat}	Turnover number
K_D	Dissociation constant
kDa	Kilodaltons
KLD	Kinase-Ligase-*Dpn1*
K_M	Michaelis- Menten constant
KP	Potassium phosphate buffer
KP7	Potassium phosphate buffer pH 7
KP8	Potassium phosphate buffer pH 8
Leu (L)	Leucine
LP-BER	Long patch base excision repair
Lys (K)	Lysine
MALDI-TOF	Matrix assisted laser desorption/ionization time-of-flight
MD	Molecular dynamics
MIB	Sodium malonate, imidazole, boric acid
MPD	2-Methyl-2,4-pentanediol
MMR	Mismatch repair
MR	Molecular replacement
MST	Microscale thermophoresis
MW	Molecular weight
MWCO	Molecular weight cut-off
NER	Nucleotide excision repair
NHEJ	Non-homologous end joining
NLS	Nuclear localisation signal
nt	Nucleotide
PCNA	Proliferating cell nuclear antigen
PCR	Polymerase chain reaction
PEI	Polyethylenimine
P_{lac}	Lactose promoter
PMSF	Phenylmethylsulfonyl fluoride
PNK	Polynucleotide Kinase
PoC	Point-of-Care
Pol I	DNA polymerase I
Q column	Q Sepharose column
qPCR	Quantitative polymerase chain reaction
RER	Ribonucleotide excision repair
RMSD	Root-mean-square deviation
rNMP	Ribonucleotide monophosphate

RPM	Revolutions per minute
RT-PCR	Reverse transcriptase polymerase chain reaction
SB	SuperBroth
SD	Standard deviation
SDS	Sodium dodecyl sulphate
SDS-PAGE	Sodium dodecyl sulphate polyacrylamide gel electrophoresis
SEC	Size-exclusion chromatography
SEM	Standard error of the mean
Ser (S)	Serine
SNP	Small nucleotide polymorphism
SP-BER	Short patch base excision repair
SP column	SP Sepharose column
SPR	Surface plasmon resonance
ss	Single-stranded
T5	Bacteriophage T5
T7	Bacteriophage T7
T_a	Annealing temperature
TAE	Tris-acetate-EDTA
Taq	*Thermus aquaticus*
TCEP-HCl	tris(2-carboxyethyl)phosphine hydrochloride
TEMED	Tetramethylethylenediamine
Thr (T)	Threonine
Tris	Tris (hydroxymethyl) aminomethane
Trp (W)	Tryptophan
U	Enzyme units
UV	Ultraviolet
V_M	Matthews coefficient
XPG	Xeroderma pigmentosa complementation group G

Chapter 1 - Introduction

1.1. Flap endonucleases

Flap endonucleases (FENs) are a group of conserved metalloenzymes that possess structure-specific 5′ endonuclease activity and 5′ to 3′ exonuclease activity. They have been identified across all domains of life and even in some viruses, where they possess an essential function for viability (Kucherlapati *et al.*, 2002; Tishkoff *et al.*, 1997; Bayliss *et al.*, 2005; Díaz *et al.*, 1992; Fukushima *et al.*, 2007). Bacterial 5′ nucleases are members of the FEN superfamily that share several conserved structural elements, whose members include FEN-1, Exonuclease I (EXO1), Xeroderma Pigmentosa complementation group G (XPG) and Gap Endonuclease I (GEN1). Activities of the superfamily span similar but complementary DNA replication, repair and recombination pathways, through cleavage of a diverse range of substrate topologies that arise as nucleic acid intermediates during these processes as summarised in Figure 1-1.

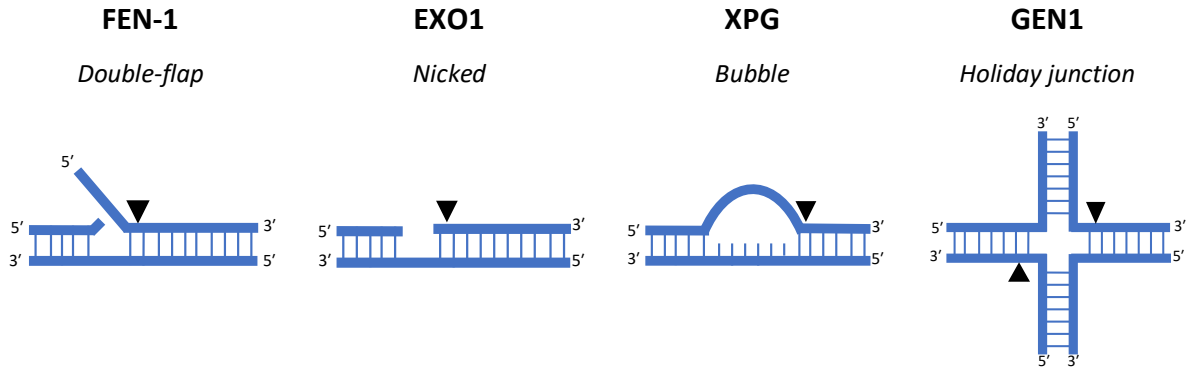


Figure 1-1: FEN superfamily substrate specificity

The nucleases FEN-1, EXO-1, XPG and GEN-1 are members of the FEN superfamily. In eukaryotes the preferred cellular substrate for FEN-1 is double-flap DNA, which arises during DNA synthesis by displacement of RNA primers. The preferred substrate for EXO-1 is a nicked DNA, for XPG is a bubble structure and GEN-1 it is a Holiday junction.

Additionally, a large number of structure-specific substrates are cleaved by FENs *in vitro,* indiscriminate of nucleotide sequence. This has led to the exploitation of the prokaryotic and viral enzymes within a myriad of molecular biology and biotechnology techniques. This

literature review will examine the structure, function, mechanism and uses of FENs with a greater emphasis on the prokaryotic and viral homologues.

1.2. FEN activities and substrates *in vitro*

A variety of substrates, as shown in Figure 1-2, have been shown to be cleaved by FENs *in vitro* although substrate specificity and catalytic efficiency of cleavage is highly dependent on FEN homologue and reaction conditions. It is important to note that laboratory-designed substrates are used for convenience and to provide insight into activity, rather than being particularly biologically relevant.

1.2.1. 5′ flap endonuclease activity

The flap endonuclease activity of FENs is specific for 5′ single-stranded (ss) DNA or ssRNA at a duplex junction or bifurcation. Examples of substrates cleaved by this activity *in vitro* include single-flap (composed of a 5′ ssDNA flap), double-flap (composed of a 5′ ssDNA flap and a single 3′ unpaired nucleotide), pseudo-Y and 5′ overhangs (Sayers and Eckstein, 1990; Lyamichev *et al.*, 1993; Harrington and Lieber, 1994; Lyamichev *et al.*, 1999). Compared to a single 5′ flap substrate, binding and cleavage efficiency in mammalian, yeast, archaeal and bacterial FENs increases for double-flap substrates containing a single 3′ nucleotide (Finger *et al.*, 2009; Kao *et al.*, 2002; Lyamichev *et al.*, 1993; Murante *et al.*, 1995). These double-flap substrates are hypothesized to be the natural intermediates of DNA replication and repair *in vivo* (Harrington & Lieber, 1995). The 3′ flap is also important for fixing the cleavage site one nucleotide into the double-stranded (ds) DNA region to leave a nick which can be directly ligated by DNA ligase without any gap filling (Lyamichev *et al.*, 1993). Finger *et al.* (2009) found that human FEN-1 can even tolerate a hair-pin loop in the 5′ flap providing the 3′ flap was present. However, this substrate tolerance was not found with *Escherichia coli* DNA polymerase I FEN domain (Xu *et al.*, 2001), which may represent species-specific substrate specificity or could have been due to different experimental conditions used in the studies. In contrast to this, the presence of a 3′ flap does not fix the cleavage site or improve cleavage efficiency in bacteriophage T5FEN (Williams *et al.*, 2007), which is likely due to the absence of a 3′ flap binding pocket in bacteriophage FENs.

Pseudo-Y substrates, which lack downstream dsDNA, can be cleaved by T5FEN and mammalian FEN-1 albeit with greater variation in cleavage site and a reduced cleavage efficiency (Harrington and Lieber, 1994; Murante *et al.*, 1995; Williams *et al.*, 2007). For example, mouse FEN-1 cleaves a pseudo-Y substrate with 100-fold lower efficiency compared to when there is downstream dsDNA present (Harrington and Lieber, 1994). However, archaeal FENs require larger substrates that contain two duplex regions so cannot cleave pseudo-Y and 3′ overhang substrates (Williams *et al.*, 2007).

1.2.2. 5′ to 3′ exonuclease activity

FENs exhibit 5′ to 3′ exonuclease activity on recessed 5′ ends, nicks and gaps. Bacterial, bacteriophage and mammalian FENs can cleave blunt ended dsDNA, albeit with different levels of activity whilst some bacteriophage and bacterial FENs can also cleave ssDNA (Frenkel and Richardson, 1971; Sayers *et al.*, 1988; Bhagwat *et al.*, 1997; Williams *et al.*, 2007; Allen *et al.*, 2009).

1.2.3. Gap endonuclease activity

A subset of FENs have also been shown to act on gapped DNA through a separate endonucleolytic activity known as gap endonuclease (GEN) activity. Originally observed in T5FEN, through experiments measuring nuclease degradation of plasmid DNA, this activity results in the cleavage of gapped DNA to create a double-stranded nick which is then subsequently completely degraded through 5′ to 3′ exonuclease activity (Sayers and Eckstein, 1991). Later, Parrish *et al.*, (2003), showed that the *C. elegans* FEN-1 homologue, CRN-1 cleaves gapped dsDNA during DNA fragmentation in apoptosis. This GEN activity is also present in human FEN-1, acting at the 3′ end of gaps and has been implicated in the processing of stalled replication forks (Zheng *et al.*, 2005).

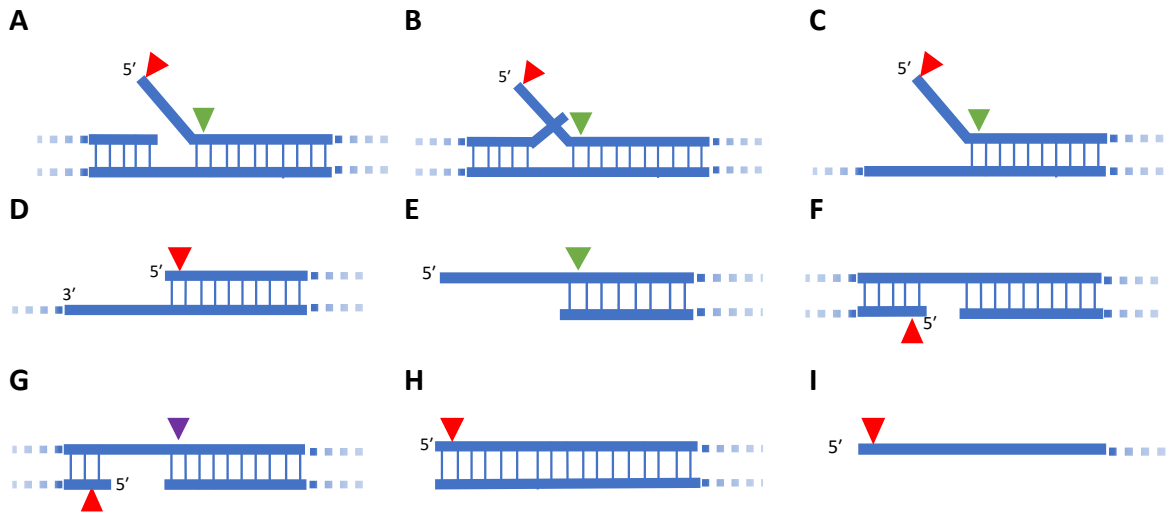


Figure 1-2. Substrates cleaved by FENs

Examples of DNA substrates that are cleaved by FENs include: (A) single-flap, (B) double-flap, (C) pseudo Y, (D) recessed 5′ end, (E) 5′ overhang, (F) nicked, (G) gapped, (H) blunt 5′ end of double stranded DNA, (I) 5′ end of single stranded DNA. The sites of exonuclease (red triangle), endonuclease (green triangle) or gap endonuclease (purple triangle) are shown. There appears to be some variability in the substrate specificity between FENs and not all FENs are able to cleave all of these substrates.

1.2.4. Divalent metal ion co-factors

FENs have a requirement for divalent metal ions to support catalysis and two or three metal ions have been shown to be co-ordinated in the active site by conserved residues. Although Mg^{2+} is the preferred cofactor for most FENs, many have been shown to utilize a range of divalent ions as cofactors *in vitro,* with various levels of efficiency. Examples of cofactors utilized by a variety of FENs are summarised in Table 1-1. Studies of the roles of different metal ions in bacteriophage T5FEN catalysed DNA cleavage show that divalent ions regulate the type of nuclease cleavage in a metal ion and concentration dependant manner (Feng *et al.*, 2004). For example, endonuclease cleavage predominates at low Mg^{2+}, Mn^{2+}, Co^{2+}, Zn^{2+} and Ni^{2+} ion concentrations whilst exonuclease activity is stimulated at higher concentrations (Feng *et al.*, 2004). This dependence can possibly be explained by the presence of two distinct metal ion binding sites of differing affinity, which have been shown to contribute differently to endonuclease and exonuclease activity (Feng *et al.*, 2004).

Interestingly, under certain reaction conditions, T5FEN can stimulate the cleavage of covalently closed circular double stranded DNA (cccDNA). This is supported at a low

monovalent salt concentration by Mn^{2+} and to a lesser extent with Co^{2+}, Zn^{2+} and Mg^{2+}, under specific reaction conditions (Garforth *et al.*, 2001). However, the mechanism of this reaction and relevance, if any, to the *in vivo* functions of FENs is a point of discussion.

Table 1-1:Range of cofactors utilized by FENs in vitro

FEN	Divalent metal ions	Reference(s)
Bacteriophage T5FEN	Catalytic: Mg^{2+} (+++), Mn^{2+} (+++), Co^{2+}(++), Fe^{2+} (++), Ni^{2+} (++), Cu^{2+} (+), Zn^{2+} (+) Non-catalytic: Ca^{2+}	(Garforth *et al.*, 2001; Feng *et al.*, 2004)
Bacteriophage T4 RNase H	Catalytic: Mg^{2+} (+++), Mn^{2+} (+++)	(Bhagwat *et al.*, 1997)
Murine FEN-1	Catalytic: Mg^{2+} (+++), Mn^{2+} (+++), Co^{2+} (++), Ca^{2+}(+), Ni^{2+} (+) Non-catalytic: Cu^{2+}, Zn^{2+}	(Harrington and Lieber, 1994)
Mycobacterium smegmatis FenA	Catalytic: Mn^{2+} (+++), Mg^{2+} (++), Co^{2+} (++), Ni^{2+} (+), Zn^{2+}(+) Non-catalytic: Ca^{2+}, Cu^{2+}	(Uson *et al.*, 2017)
Thermococcus barophilus Ch5 FEN-1	Catalytic: Mg^{2+} (+++), Mn^{2+} (+++), Ni^{2+} (++), Co^{2+} (++), Cu^{2+} (+), Zn^{2+} (+), Ca^{2+} (+)	(Lin *et al.*, 2022)

+++, greatest level of activity; ++, moderate level of activity; +, low level of activity.

1.3. FENs in molecular biology and biotechnology

As outlined above, FENs possess structure-specific and sequence independent nuclease activity. This feature makes them attractive enzymes for use in molecular biology and biotechnology and various FENs have been exploited in these fields since as far back as the 1980s. Early techniques such as phosphorothioate-based site-directed mutagenesis (SDM) utilized the phage FENs from bacteriophages T5 and T7. Whilst techniques developed later such as Third Wave's Invader © assay and the revolutionary TaqMan assay, include archaeal and bacterial FENs, respectively. The use of bacterial and viral FENs within the molecular biology field still remains popular to this day.

1.3.1. Phosphorothioate-based site-directed mutagenesis

Phosphorothioate-based site-directed mutagenesis (SDM) is an early site-directed mutagenesis technique developed in the Eckstein laboratory. Built on the earlier work of Zoller and Smith (1982), this method combined T7 gene product 6 (gp6) and T5FEN into a highly

efficient technique (Sayers *et al.*, 1988; Sayers *et al.*, 1992; Nakamaye and Eckstein, 1986). This was packaged and sold until the early 2000s within Amersham International's commercially available SDM kit before being outcompeted by PCR-based methods. In this technique, as summarised in Figure 1-3, a mutagenic primer is annealed to a single-stranded circular phage (+) DNA template and polymerised using a phosphorothioate deoxynucleoside triphosphate analogue. T5FEN is used to selectively cleave and degrade all products of this reaction which are not closed circular dsDNA, such as partially polymerised DNA (which contains nicks or gaps) or template DNA. A nick is introduced on the template strand of the fully double stranded polymerised product by a restriction endonuclease and then gapped and fully degraded by T7 gp6, leaving the single-stranded (-) DNA carrying the mutation. This strand is used as the template for a second polymerisation reaction to create a double-stranded DNA plasmid carrying the mutation in both strands.

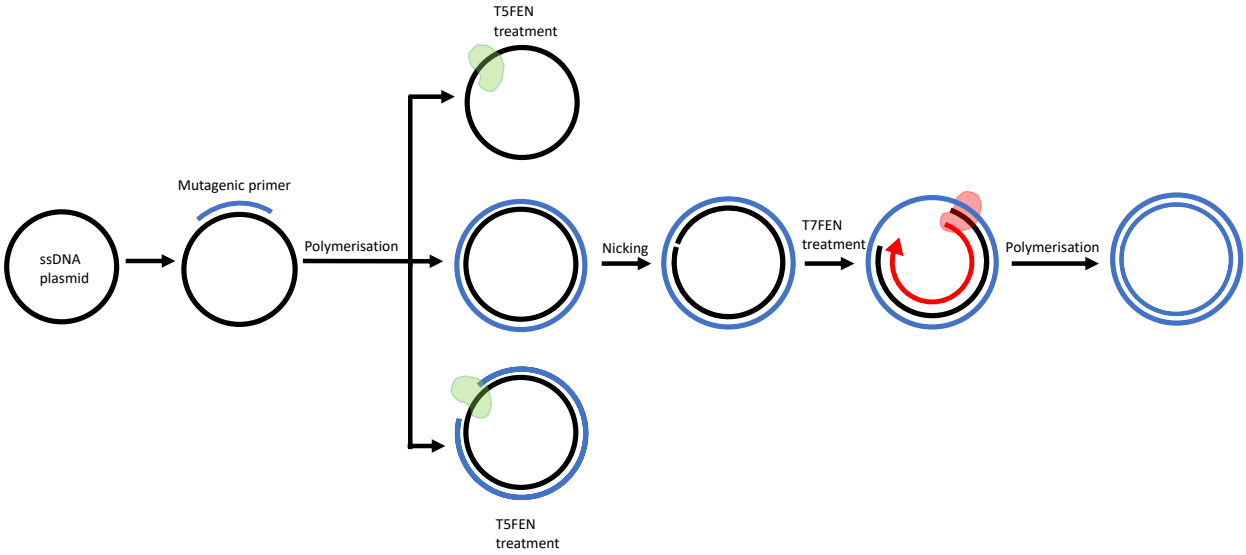


Figure 1-3: Phosphorothioate-based site-directed mutagensis

A mutagenic primer is annealed to the single-stranded DNA template plasmid and is polymerised in the presence of 3 natural nucleoside triphosphates and one nucleoside thiophosphate. After polymerisation, the products are treated with T5FEN which is able to selectively digest the by-products of this (ssDNA template plasmid and gapped product as a result of incomplete polymerisation), leaving intact the covalently closed double-stranded DNA containing the mutation on one strand. The template strand is nicked and subsequently digested by T7 gp6, leaving the mutagenic ssDNA plasmid. Polymerisation of this produces a double-stranded DNA bacteriophage genome which can be used for transformation of *E. coli.*

1.3.2. Plasmid preparations

T5FEN (also referred to a T5 D15 or T5 exonuclease in the early literature (Moyer and Rothe, 1977)) can completely degrade all DNA except closed circular DNA under certain conditions (Sayers and Eckstein, 1991). This property has made it an attractive enzyme for use for preparation of plasmids for molecular cloning. Alkaline lysis is a technique used to isolate plasmid DNA from bacteria using alkaline conditions. Due to the conditions used, preparations isolated via this method can be contaminated by denatured plasmids (Vinograd and Lebowitz, 1966; Sayers *et al.*, 1996) which migrate close to the covalently closed circular plasmid DNA on an agarose gel as so called 'ghost bands.' These contaminating plasmids pose an issue as they are resistant to endonuclease cleavage so can contribute to the background of non-recombinant transformants (Hengen, 1996; Sayers *et al.*, 1996). However, treatment of plasmid isolations with T5FEN removes this contamination, as well as nicked and linearized plasmid and reduces the background by two orders of magnitude (Sayers *et al.*, 1996). T5FEN treatment has been shown to produce higher transfection rates of plasmid complementary DNA (cDNA) library minipreps in eukaryotic hosts due to the removal of contaminating bacterial DNA (Kiss-Toth *et al.*, 2001).

Even to this day, T5FEN is used to increase the efficiency of DNA preparations. T5FEN has been used for preparations of the low copy number replication template in hepatitis B virus-infected human cells. This replication template is composed of cccDNA and is isolated by T5FEN digestion of other forms of DNA and replicative intermediates in DNA extractions of infective cells which can have downstream uses in a drug development and infection modelling setting (Qu *et al.*, 2018).

1.3.3. Gibson cloning and DNA assembly techniques

The versatile T5FEN is also used within more modern molecular cloning techniques such as Gibson's assembly method. Removing the need for restriction endonucleases for molecular cloning, Gibson's assembly can combine multiple overlapping DNA fragments up to several hundred kilobases within a single reaction (Gibson *et al.*, 2009), and utilizes three enzymes: T5FEN, DNA polymerase and DNA ligase, as summarised in Figure 1-4. Specifically, T5FEN removes nucleotides from the ends of overlapping dsDNA fragments, leaving complementary

3’ overhangs that anneal to each other. Gaps generated by T5FEN digestion are filled with DNA polymerase whilst DNA ligase ligates the strands into a continuous covalently joined DNA strand. Prevention of over digestion of the DNA fragments by T5FEN is controlled by the suboptimal reaction conditions of 50°C and pH 7.5 (Gibson *et al.*, 2009).

More recent studies have examined the key three enzyme-catalysed reactions in more detail (FEN, polymerase, ligase) and have found that the only protein required is T5FEN, at low concentrations, whilst the gap filling and ligation can take place after transformation, *in vivo*, utilizing *E. coli* host DNA replication machinery. The initial *in vitro* reaction takes place at 30°C and costs less than 1 pence per reaction (Xia *et al.*, 2019).

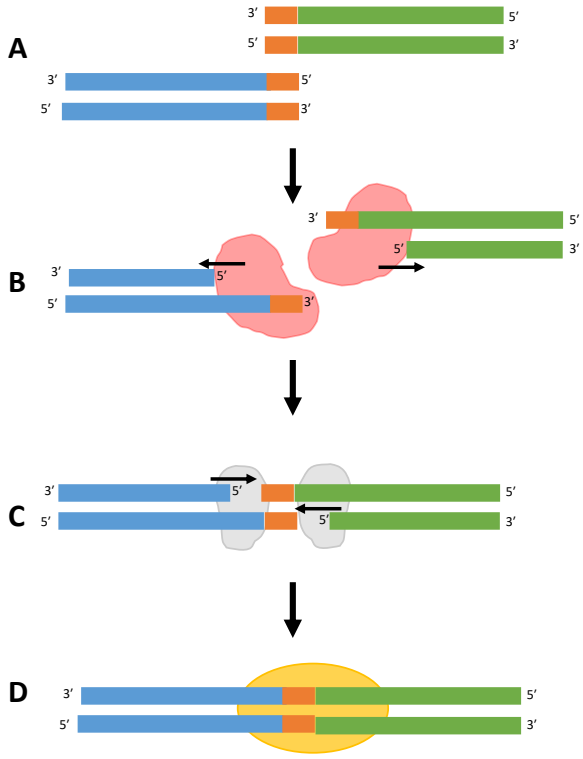


Figure 1-4: T5FEN mediated joining of large DNA fragments.

(A) DNA fragments with overlapping terminal sequences (orange). (B) T5FEN (red) cleaves 5’ ends of DNA fragments resulting in ssDNA overhangs. At a high temperature, T5FEN activity is inhibited. (C) Complementary ssDNA strands anneal and DNA polymerase (grey) fills the gaps. (D) DNA ligase seals the resulting nick to create a continuous strand of DNA. The Gibson cloning system consists of the three enzymes (T5FEN, DNA polymerase and DNA ligase) as a single reagent. Other T5FEN mediated joining of DNA fragments consists of only the T5FEN stage (A-B) *in vitro*, whilst gap filling and ligation occurs *in vivo*, after transformation into *E. coli*.

1.3.4. CRISPR-Cas gene editing chimeric proteins

The clustered regularly interspaced short palindromic repeat (CRISPR) and CRISPR-associated protein (Cas) system has revolutionised genome editing, enabling targeted insertions and deletions within genomes (Jinek et al., 2012). In multiple studies, T5FEN has been used to generate Cas fusion proteins with enhanced genome editing capabilities. For example, a technique called TEXT (**T**ethering **Ex**onuclease **T**5 with FnCas12a) employs a fusion strategy in

which a chimeric protein is generated from T5FEN and *Francisella novicida* Cas12a. This system has been shown to significantly increase the knockout efficiency of *Fn*Cas12a in human cells at multiple genomic loci in three different cell lines by increasing both size and frequency of deletions. This is mediated by T5FEN cleavage of 5' overhangs generated by *Fn*Cas12a (Wu et al., 2020). In another study, T5FEN fusion to either Cas9 or Cas12a increased the frequency and size of deletions in rice genomes to improve the editing efficiencies of this enzymes (Zhang *et al.*, 2020).

1.3.5. Molecular therapeutic screening

Aptamers are short single-stranded oligonucleotides that can fold into defined structures enabling them to bind to various molecules, from small inorganic molecules, to proteins, with a high affinity and selectivity comparable to antibodies. They possess therapeutic potential through binding to proteins and disrupting protein interactions (Keefe *et al.*, 2010) and have a broad range of therapeutic applications such as in the prevention of thrombosis, modulation of inflammation and inhibition of viral replication (Bock *et al.*, 1992; Green *et al.*, 1995; Pan *et al.*, 1995). Aptamers are discovered through an iterative screening of oligonucleotide libraries in a process named system evolution of ligands by exponential enrichment (SELEX) and once detected, they must be characterised through various biophysical techniques such as isothermal titration calorimetry (ITC), surface plasmon resonance (SPR) or microscale thermophoresis (MST), to yield binding parameters (Ruscito and DeRosa, 2016). This can be limited by the low throughput nature of these techniques. A recent paper has described a novel use of T5FEN in selection of aptamers that bind to small molecules (Alkhamis *et al.*, 2020). The authors have shown that the binding of ligands to aptamers prevents their digestion by T5FEN (Figure 1-5). The level of inhibition to digestion correlates with the affinity of the ligand to the aptamer and was comparable to the techniques listed previously. This assay has been used in a high-throughput microplate format compared to low throughput of other DNA binding analysis techniques (Alkhamis *et al.*, 2020).

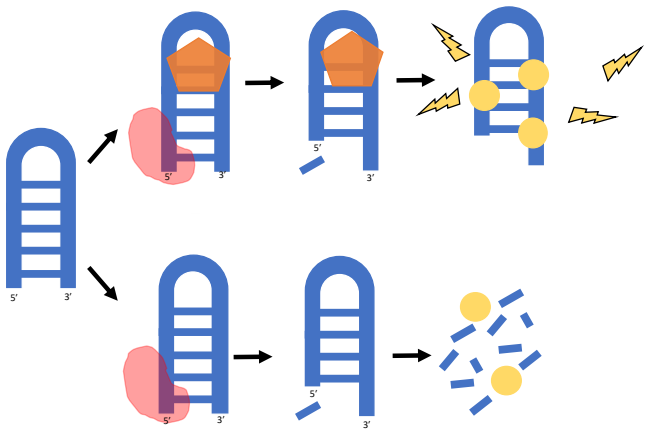


Figure 1-5: Profiling of DNA aptamer-small molecule binding using T5FEN

DNA aptamers bound to small molecule ligands (orange) inhibits aptamer digestion by T5FEN (red) and the extent of the inhibition is correlated with the strength of aptamer-ligand binding. The level of DNA digestion is measured through staining with a DNA dye (yellow), where the dye is not able to bind to digestion products of small oligonucleotides.

T5FEN has also been used in a different screening approach to identify inhibitors of DNA topoisomerases (Deng and Leng, 2021). This high-throughput assay utilizes a plasmid that has been engineered to form an AT hairpin upon supercoiling. This hairpin provides a substrate for initial cleavage by T5FEN to produce a nick, which is subsequently completely degraded by the exonuclease activity of the enzyme. Whilst, relaxation of the plasmid by topoisomerases yields a structure that is resistant to T5FEN cleavage. Plasmid degradation is detected by the use of a double-stranded DNA-binding dye and therefore, inhibitors are detected based on a change is dye signal. This technique has been validated by a 50-compound library (Deng and Leng, 2021).

1.3.6. FEN activity with molecular diagnostic applications

1.3.6.1. Third Wave's invader assay

Third Wave Technologies' Invader ® assay (Figure 1-6) was developed and used for single nucleotide polymorphism (SNP) genotyping and the detection of specific DNA and RNA sequences (Victor Lyamichev *et al.*, 1999; Hall *et al.*, 2000; Allawi *et al.*, 2004; Schutzbank *et al.*, 2007). It exploits the minimal substrate requirements of archaeal FENs (a double-flap) to provide a highly sensitive and specific detection system (Lyamichev *et al.*, 1999; Hall *et al.*, 2000). The assay has been successfully used in the detection of factor V (Leiden) SNP mutations and the detection of high-risk human papilloma virus DNA (Ledford *et al.*, 2000; Schutzbank *et al.*, 2007). When multiplexed with PCR, it has been used for high-throughput

SNP analysis. For example, a study that screened 768 patients for >92,000 SNPs to identify specific cardiovascular risk factor markers (Ozaki *et al.*, 2002).

In the assay, three oligonucleotide probes are used: an invader, primary probe and Förster resonance energy transfer (FRET) probe. In the first stage of the reaction, the invader and primary probe anneal to target DNA. In order to detect a specific SNP for example, the invader and primary probe anneal to target DNA and form an invasive structure, composed of a 5' flap and single unpaired 3' nucleotide, whereas absence of the SNP does not result in an invasive structure forming. The archaeal FENs used can only cleave the double-flap substrate formed, due substrate specificity (Lyamichev *et al.*, 1993). After cleavage, the released flap can subsequently form a second invasive structure by annealing to the FRET probe for fluorescent signal detection after cleavage by FEN.

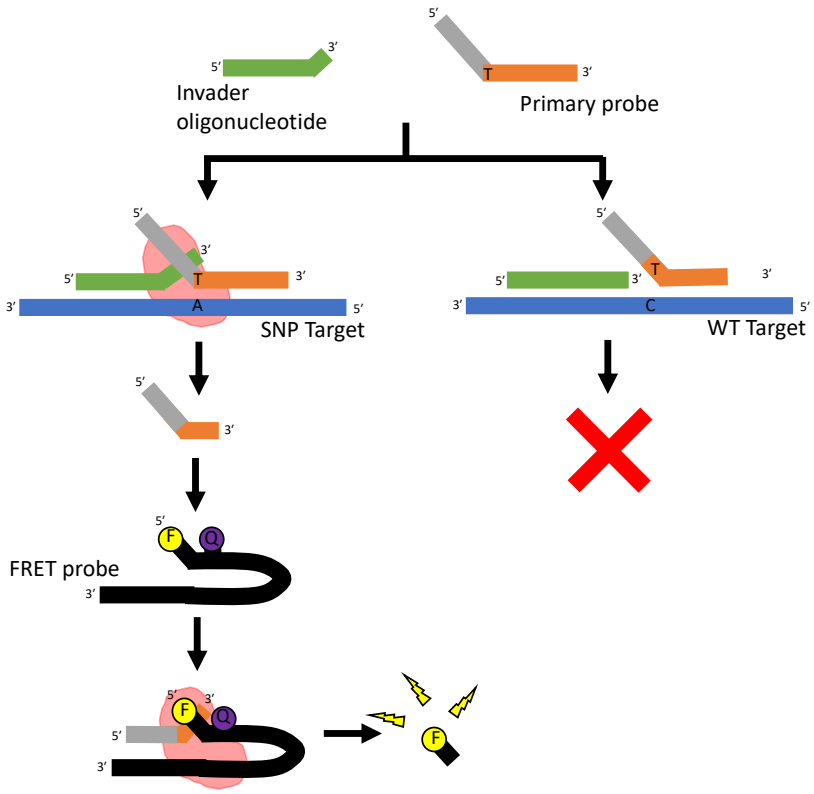


Figure 1-6: Third Wave's Invader © assay for identification of SNP

Two oligonucleotides (invader oligonucleotide and primary probe) anneal to SNP or WT target DNA. Annealing of probe to SNP target DNA results in the invader oligonucleotide annealing with a 3' flap, which creates a substrate which can be cleaved by FEN-1 (red). This cleavage product anneals to a dual-labelled FRET probe and is also cleaved by FEN-1. The fluorophore is separated from the quencher dye and an increase in emission fluorescence is recorded. Alternatively, annealing of probe to WT target DNA does not result in a FEN-1 substrate forming and therefore no downstream reaction takes place.

1.3.6.2. TaqMan Quantitative-PCR

Perhaps the most widely used commercial application of FENs is in the polymerase chain reaction (PCR) field. PCR involves the selective amplification of DNA sequences and has a broad range of applications, not only in basic research such as molecular cloning but also in areas of medical diagnostics, forensics and agriculture with PCR-based tests fast becoming the new 'gold standard' of molecular diagnostics (Zauli, 2019). Early PCR protocols utilized the thermostable *Thermus aquaticus* DNA polymerase I (Taq Pol) for DNA amplification. The TaqMan assay was first described by Livak *et al.* (1995) and can provide quantitation of amplified DNA as depicted in Figure 1-7. It relies on the FEN domain of Taq polymerase to cleave a dual-labelled probe in techniques such as real-time PCR (RT-PCR). TaqMan probes are target-specific and labelled with a fluorophore on one side and a quencher molecule on the other and the signal is quenched, through FRET. During PCR, the probe anneals between the forward and reverse primers. As the polymerase domain of Taq Pol extends the primers it encounters the probe and the probe is partially displaced and then degraded *via* 5' to 3' exonuclease cleavage by the FEN domain, separating the reporter from the quencher. The change in fluorescent intensity from the probe can be measured in real-time to determine quantity of DNA. Nucleic acid based detection systems such as described above for the TaqMan © system are widely used in applications and have been successfully used in the detection of broad range viral and bacterial infections (Speers, 2006). Developments using new polymerases with differing levels of processivity, fidelity, extension rates and thermal stability have also been made. Furthermore, the size of this market is huge and the global polymerase chain reaction global market is predicted to be worth USD 25.3 billion by 2028 (Grand View Research, 2021).

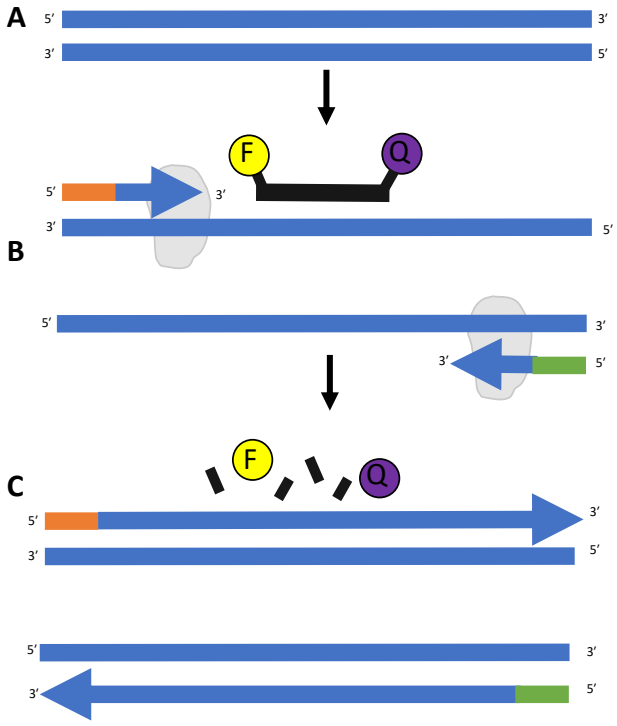


Figure 1-7: QPCR utilizing TaqMan probes and FEN activity

(A) Template DNA for PCR. (B) The TaqMan probes anneals to the region of template DNA that carries the probe-specific target DNA and the fluorescence from the fluorophore is quenched when the probe is intact. Forward (red) and reverse (green) primers anneal to template DNA and are extended by polymerase. (C) During extension by polymerase, the TaqMan probe is cleaved by FEN activity of the polymerase to complete the extension cycle. This releases the fluorophore from the quencher and the change in fluorescence is detected.

1.3.6.3. Point-of-care diagnostics

Much of PCR-based nucleic acid detection takes place at highly specialised and expensive laboratories at hospitals and consequently require the infrastructure for this. Furthermore, although PCR protocols for diagnosis can take 45 minutes or less, the transportation of samples, sample processing and relaying of results mean that diagnosis can still take a few days for results to be presented. Rapid-point-of care (PoC) detection offers a solution to this issue and PCR techniques have been exploited into so called 'lab on a chip' devices, which can offer sample to result in less than 45 minutes directly at the PoC. Once a sample has been inserted into the device, all intermediate stages between sample collection and diagnosis, including sample preparation, amplification of nucleic acids and detection, are carried out within the device itself, relying on microfluidics for transitioning between the stages (Kong et al, 2016). Binx health (the industrial partner of this PhD project) have developed an FDA-approved PoC nucleic acid amplification testing device known at the IO © platform. This device is used for specific and sensitive detection of *Chlamydia trachomatis* and *Neisseria gonorrhoeae* in less than 30 minutes (Van Der Pol *et al.*, 2020). Described in Figure 1-8, target DNA is amplified from a sample and detected using an electrochemically labelled ferrocene probe. Specific exonuclease activity of the T7 gp6 cleaves the annealed probe, which oxidises

at a known potentiation to produce a change in current that is detected. A benefit of a PoC test like this is that it can be carried out by a non-specialist after minimal training.

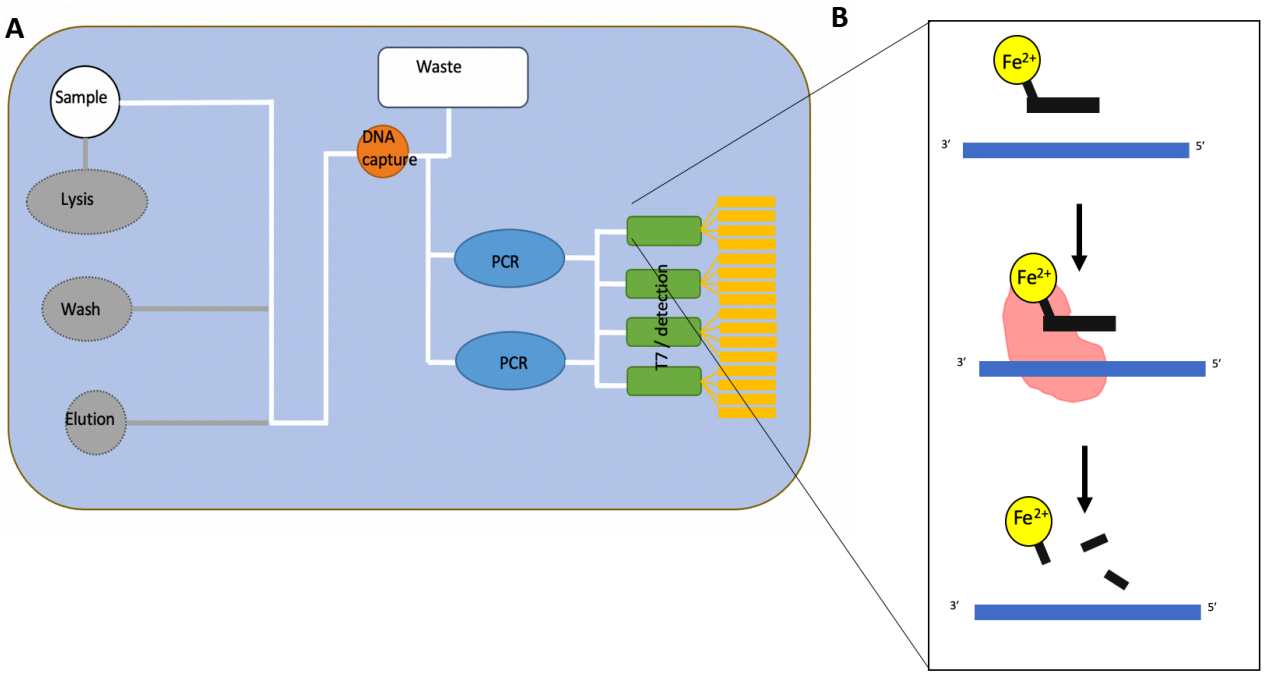


Figure 1-8: Binx IO © point-of-care testing device utilizes FEN activity of bacteriophage T7 gp6

The Binx IO © technology is a cartridge-based nucleic acid amplification test used for the point-of-care detection of bacterial infections. (A) Once a test sample is applied to the cartridge and is inserted into the reader, microfluidics are used to carry out all stages of sample preparation, DNA extraction, PCR amplification and detection in a single test. (B) During the detection stage, a ferrocene-labelled probe anneals to a target region within amplified DNA. The probe is cleaved by T7 gp6 (red) and the ferrocene label is released. The released ferrocene is oxidised at a known voltage to produce a current which is measured.

1.4. Physiological function of flap endonucleases

1.4.1. DNA replication

DNA replication is an essential but complex process, utilizing the concerted actions of many different replication proteins. It occurs in a semi-conservative manner with each strand of the parental double helix used as a template for the complementary strand. Initiation occurs at an origin of replication, where the replication machinery is loaded onto the DNA. Helicases unwind the two parental DNA strands and single-stranded (ss) DNA-binding proteins stabilize the DNA and hold the single-stranded DNA template strands open. This active site of

replication is known as a replication fork (Figure 1-9). The process of DNA replication is highly conserved between prokaryotes and eukaryotes.

In prokaryotes, RNA primase synthesizes RNA primers that serve as the starting point for DNA synthesis and these are extended by the holoenzyme DNA polymerase III to generate the DNA daughter strand. The sliding clamp (β clamp) is a subunit of DNA polymerase III and ensures the processivity of DNA replication and interacts with other replication proteins (Saro and O'Donnell, 2001). Due to the uni-directionality of DNA synthesis in the 5' to 3' direction, the leading strand is synthesized as a continuous fragment whilst the lagging strand is synthesized as discontinuous Okazaki fragments which range in size from 100 -200 base pairs (bp) (Ogawa and Okazaki, 1980). On the lagging strand, DNA polymerase I replaces the RNA primers with DNA, through the FEN domain, whilst DNA ligase seals the DNA to generate a continuous strand of DNA.

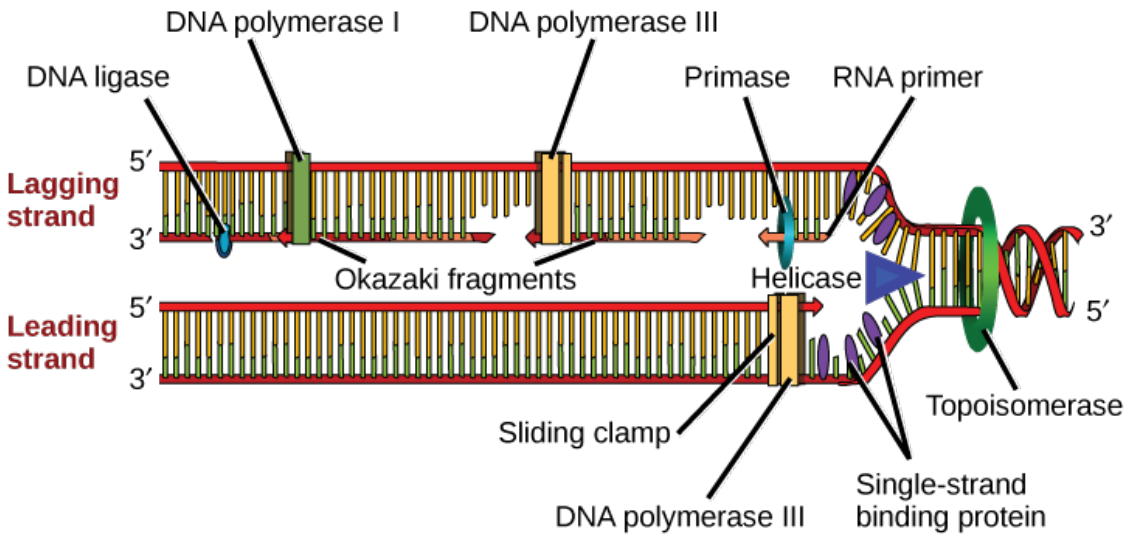


Figure 1-9: Prokaryotic replication fork

Leading and lagging strand synthesis during DNA replication. Due to the unidirectionality of DNA polymerase, the lagging strand is synthesized as discontinuous Okazaki fragments which are subsequently processed into continuous DNA. Diagram by Rye et al., 2016. Distributed under the Creative Commons Attribution 4.0 International License. [Access for free at https://openstax.org/books/biology/pages/1-introduction]

In prokaryotes, two pathways for RNA primer removal exist as summarised in Figure 1-10: (A) FEN short-flap pathway and (B) FEN-RNase H pathway. In the FEN short-flap pathway, the RNA primer is displaced into a 5' ss flap by DNA polymerase I, through strand displacement

synthesis and this flap is subsequently cleaved by the FEN domain of the protein (Balakrishnan and Bambara, 2013). In the FEN-RNase H pathway, RNA primers are not displaced into flaps and are cleaved, 1 nucleotide at a time by RNase H and the last nucleotide is cleaved by the FEN domain of DNA polymerase I (Balakrishnan and Bambara, 2013).

The FEN-RNase H pathway is suggested to be the primary pathway for Okazaki fragment processing in prokaryotes. Mutagenesis studies in various bacterial strains show that although FEN activity is required for viability, RNase H is not (Fukushima *et al.*, 2007; Bayliss *et al.*, 2005; Ogawa and Okazaki, 1984). However, RNase H knockout bacteria generally exhibit temperature sensitive and growth deficient phenotypes, indicating that although DNA replication can occur through an RNase H independent mechanism, efficient replication requires cooperation between FEN and RNase H (Fukushima *et al.*, 2007; Bayliss *et al.*, 2005; Ogawa and Okazaki, 1984).

FENs also play essential roles in the replication of bacteriophage DNA in phage-infected cells, through the processing of Okazaki fragments generated during DNA replication (Studier, 1969; Hobbs and Nossal, 1996). For example, bacteriophage T4 RNase H (a FEN homologue) is essential for replication of T4 DNA in *E. coli* (Hobbs and Nossal, 1996). Although this function can be partially fulfilled by the host's own FEN activity, replication where T4 RNase H has been knocked out is reduced, less accurate and an increased frequency of mutations are observed (Hobbs and Nossal, 1996).

In bacteriophage T7 DNA synthesis, the polymerase, gene 5 protein (gp5), does not normally perform strand displacement synthesis, due to 3' to 5' proof-reading exonuclease activity and therefore, the main route of primer removal from Okazaki fragments is via the exonuclease activity of T7 gp6 (Lechner *et al.*, 1983). However, flaps may arise under certain circumstances, for example, due to binding of the single-stranded DNA binding protein from gene 2.5 (gp2.5), so flap endonuclease activity is still important to T7 DNA replication (Nakai and Richardson, 1988; Mitsunobu *et al.*, 2014b). Additionally, gp6 plays an essential role in the degradation of host DNA to generate nucleoside 5'-monophosphate precursors for bacteriophage DNA replication (Sadowski and Kerr, 1970). In this process, summarised in Figure 1-11, the endonuclease gene 3 protein (gp3) cleaves the host chromosome to generate

duplex DNA fragments, which are subsequently degraded by gp6 to generate nucleoside 5'-monophosphates. Up to 85% of cellular DNA is rendered acid-soluble and used for DNA transcription and studies show that both gp3 and gp6 are required for *E. coli* host DNA degradation following bacteriophage T7 infection (Sadowski and Kerr, 1970).

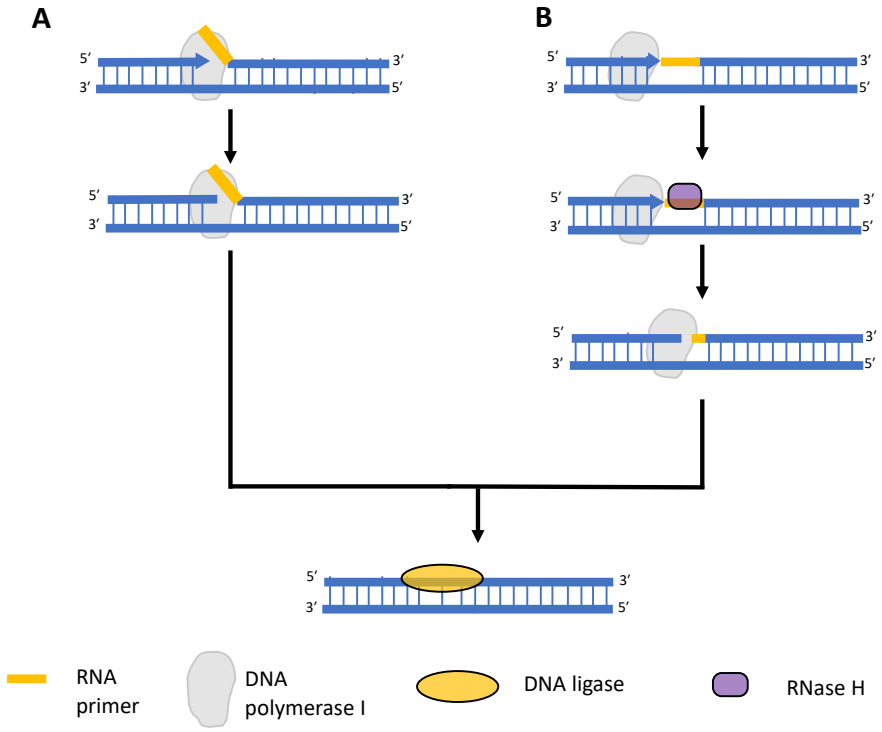


Figure 1-10: Two pathways of Okazaki fragment processing in prokaryotes

(A) FEN short flap pathway. RNA primers are displaced by DNA polymerase I, resulting in a short 5' single-stranded flap, that is removed by the FEN activity of DNA polymerase I. (B) FEN-RNase H pathway. RNA primers are not displaced into flaps by DNA polymerase I, The RNA primers are cleaved, one nucleotide at a time by RNase H and the last nucleotide is cleaved by via the FEN activity of DNA polymerase I. These pathways result in nicked DNA which is ligated together by DNA ligase to form continuous double stranded DNA.

1.4.2. DNA repair

Damage to DNA can be caused by endogenous and exogenous sources. Endogenous sources of DNA damage arise from cellular metabolic processes and include DNA base-pair mismatch, oxidation and alkylation. Exogenous sources of DNA damage arise from environmental factors and include ultraviolet (UV) radiation, ionizing radiation and various chemical agents (Hakem, 2008). To protect genome integrity and maintain cell viability, these lesions must be repaired.

Cells have developed several sophisticated repair mechanisms to achieve this through various DNA repair pathways, including: base excision repair (BER), nucleotide excision repair (NER), ribonucleotide excision repair (RER), mismatch repair (MMR), homologous recombination (HR) and non-homologous end joining (NHEJ). There is a high degree of conservation between prokaryotes and eukaryotic DNA repair pathways and FEN superfamily members function in several of these pathways (Kisker *et al.*, 2013).

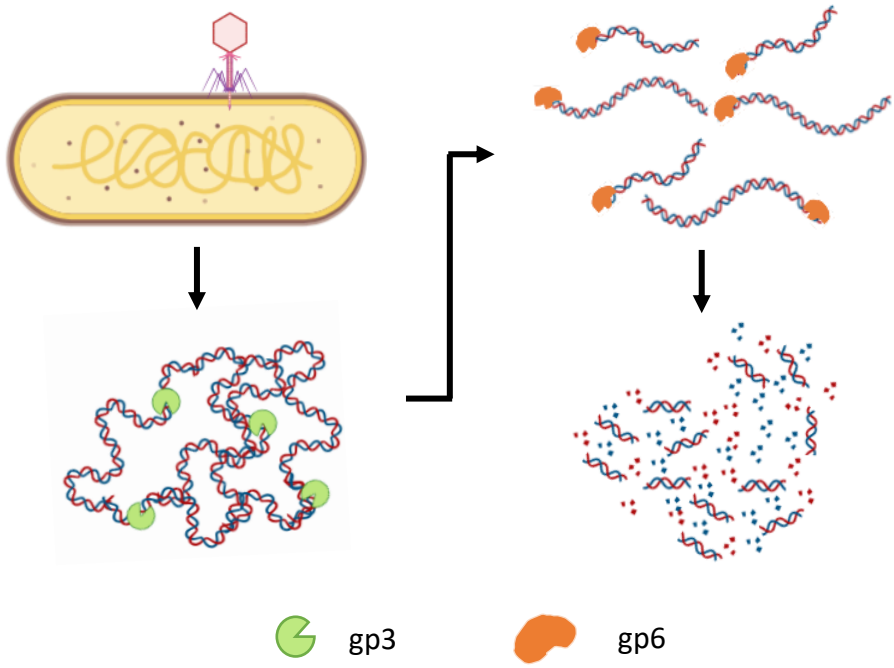


Figure 1-11: T7 gp6 degredation of host *E. coli* DNA

The DNA of a T7 bacteriophage-infected *E. coli* cell is initially cleaved by the endonuclease gene 3 protein (gp3), generating duplex DNA products. This DNA is subsequently cleaved via the exonucleolytic activity of gene 6 protein (gp6) to release nucleoside 5'-monophosphates which contribute to the nucleotide pool for new T7 DNA synthesis.

Ribonucleotide excision repair (RER) is a mechanism for removing ribonucleotides that have been mis-incorporated into a DNA chain. In this pathway in *E. coli*, RNase H2 incises the bond 5' to the ribonucleotide monophosphate (rNMP) to generate a single-stranded nick. DNA polymerase I performs strand displacement synthesis, which results in the rNMP being displaced into a 5' flap. This flap is cleaved by the FEN domain of DNA polymerase I and resulting nick is sealed by DNA ligase (Figure 1-12). In archaea, the pathway is similar, except DNA polymerase and FEN are separate enzymes (Marshall and Santangelo, 2020).

Mutagenesis studies in *E. coli* have determined that there are multiple backup systems in place, for example, in the absence of DNA polymerase I, degredation may be facilitated by ExoIX, the backup flap endonuclease encoded by the *xni* gene (Sayers, 1994; Vaisman *et al.*, 2014).

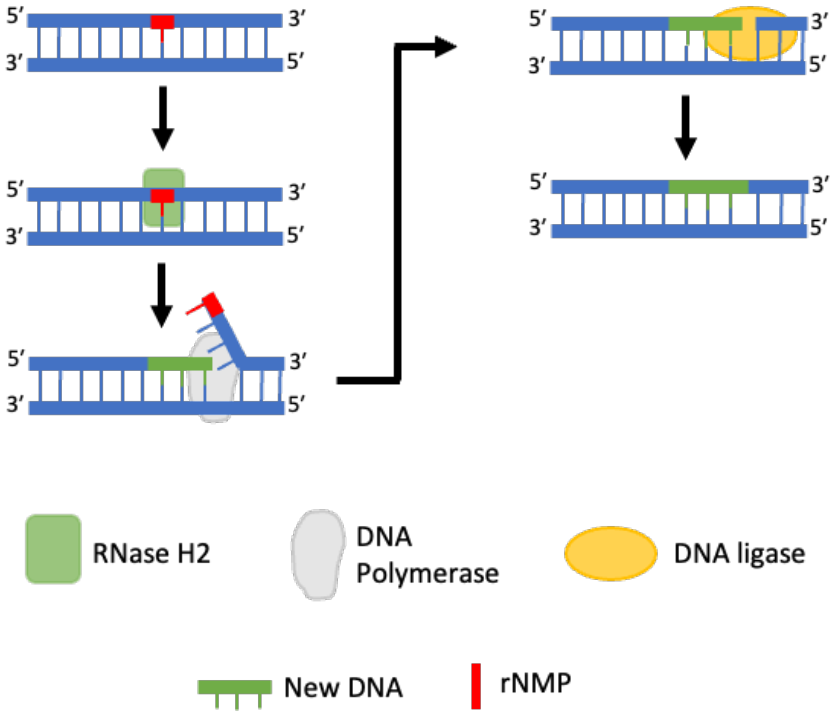


Figure 1-12: Ribonucleotide excision repair in *E. coli*

The ribonucleotide excision repair process removes ribonucleotide monophosphates (rNMP) which have been aberrantly incorporated into the DNA chain. RNase H2 incises the bond 5′ to the rNMP at the junction between rNMP and dNMP, generating a single-stranded nick. DNA polymerase I performs strand displacement synthesis resulting in a 5′ flap forming. The 5′ flap is cleaved by DNA polymerase FEN domain and resulting nick is sealed by DNA ligase.

Base excision repair (BER) is a mechanism used for correcting damage to a single base that occurs through factors such as oxidation, deamination and alkylation, where the DNA helix has not been distorted. DNA glycosylase cleaves the N-glycosylic bond to remove the damaged base and generate an apurinic/ apyrimidinic (AP) site and DNA AP endonuclease cleaves the DNA backbone to create a single-stranded nick, which is processed and resolved through short-patch BER (SP-BER) or long patch BER (LP-BER) (Dianov and Lindahl, 1994). SP-BER is used for repair if the sugar moiety has not been damaged, whilst LP-BER is used if this is the

case (Liu *et al.*, 2004). During SP-BER, the nicked DNA is filled by DNA polymerase and the DNA strand sealed by DNA ligase. In LP-BER, summarised in Figure 1-13, FEN activity is utilized and has been reconstituted *in vitro* using purified proteins in human, bacterial and archaeal systems (Dianov and Lindahl, 1994; Klungland and Lindahl, 1997; Schomacher *et al.*, 2009). In this pathway, repair patches of 2 - 6 nt are generated by DNA polymerase through strand displacement synthesis, to generate a 5′ flap (Klungland and Lindahl, 1997). The flap is cleaved by FEN activity, from DNA polymerase or a separate FEN protein, and the resultant nick is sealed by DNA ligase (Dianov and Lindahl, 1994; Klungland and Lindahl, 1997; Marshall and Santangelo, 2020).

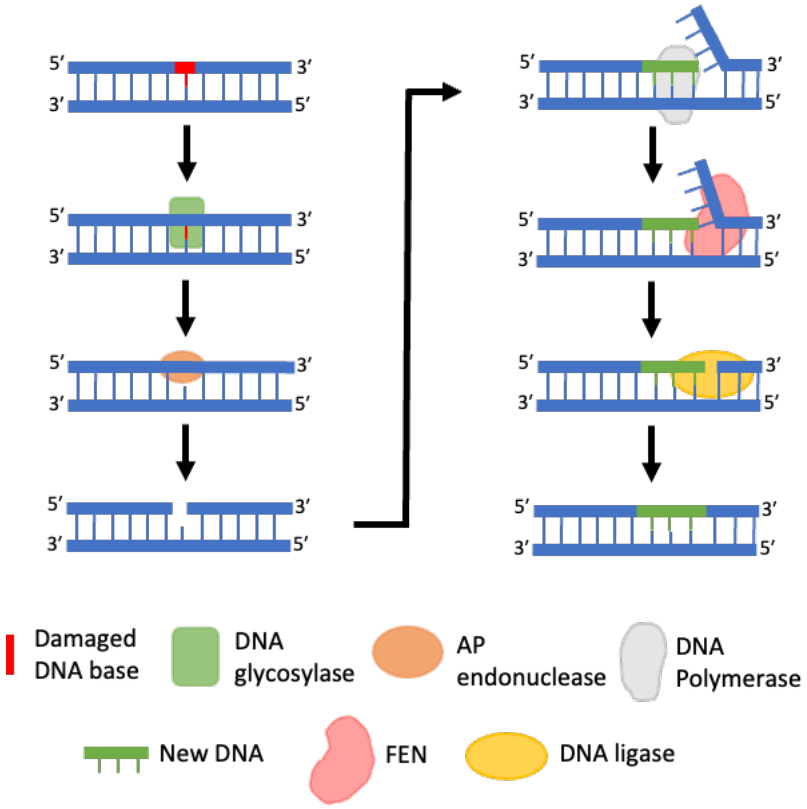


Figure 1-13: Long-patch base excision repair pathway

In the LP-BER DNA damage pathway, the damaged DNA base is recognised and cleaved by a DNA glycosylase and DNA AP endonuclease cleaves the DNA backbone to create a nick. Whilst filling the gap, DNA polymerase displaces the DNA strand to create a 5′ flap. The flap is cleaved by FEN activity (a domain of DNA polymerase or a separate protein) and sealed by DNA ligase to result in a continuous DNA strand.

Other DNA repair mechanisms utilize the activity of other 5′ nuclease superfamily members. The NER pathway removes bulky lesions that distort the DNA helix such as damage induced

by UV radiation or by bulky chemical adducts (de Laat *et al.*, 1999). In this pathway in eukaryotes, DNA repair proteins involved open up the lesion-containing DNA to form intermediate 'bubble' structures of around 24 - 32 nt (Wood, 1997). The bubble is cleaved by a XPG and ERCC1/XPF, at the 3' and 5' ends of the bubble, respectively (de Laat *et al.*, 1999) and the resulting gap is processed by DNA polymerase δ/ε and DNA ligase I (Shivji *et al.*, 1995). However, in bacteria including *E. coli* this pathway is carried out by the UvrABC protein system (Sancar, 1996).

1.5. The essential nature of flap endonucleases

FENs are essential enzymes for viability in all living organisms tested to date. However, early studies in bacteria gave conflicting answers regarding this essential nature. In *E. coli,* a defect in the FEN domain of DNA polymerase I (encoded by the *polA* gene) results in enhanced DNA damage sensitivity and conditional lethality that is permissive at 30°C but lethal at 43°C (Konrad and Lehman, 1974). Later studies showed that deletion of *polA* is viable on minimal media but lethal on rich media, whilst adding back just the FEN domain restores full viability (Joyce and Grindley, 1984). However, mutational studies of *Haemophilus influenza* (Bayliss *et al.*, 2005) and *Streptococcus pneumoniae* (Díaz *et al.*, 1992) have shown that DNA polymerase I and specifically this FEN domain are essential for viability, supporting the essential nature of FENs.

Saturation transposon mutagenesis of the *Mycobacterium tuberculosis* genome indicated that the DNA polymerase I FEN domain is essential. In this study, transposable genetic elements are transfected into the bacteria and the transposons randomly integrates into the genome. The transposon carries a stop codon which can introduce in all six possible reading frames. Individual viable clones are recovered and the transposons sequenced to identify the position of integration within the genome. In this study, no clones were recovered with transposons in the gene encoding the FEN domain of DNA polymerase I, whilst 67 viable clones were recovered when transposons were inserted within the polymerase (C-terminal) domain of the gene and the N-terminal domain would remain intact (DeJesus *et al.*, 2017).

Bioinformatic analysis of the *E. coli* DNA sequence led to the hypothesis that some bacteria, possess an alternative enzyme that provides backup FEN activity and hence rescue viability (Sayers, 1994). This was later proven by Fukushima *et al.* (2007), who demonstrated that backup FEN proteins in *E. coli* and *Bacillus subtilis* are able to compensate when DNA polymerase I is knocked out, albeit with a temperature sensitive phenotype and only on specific growth media. The complementary functions between the FEN activity from DNA polymerase I and the backup FEN protein is demonstrated by the synthetic lethality observed when both activities are knocked out (Fukushima *et al.*, 2007). In fact, all bacteria have at least one type of FEN activity and this is required for viability; 90% of bacteria have FEN activity from DNA polymerase I and 37% have a separate FEN-like protein, whilst a proportion of these have both (Allen *et al.*, 2009; Fukushima *et al.*, 2007).

1.6. Structural conservation of the flap endonuclease family

1.6.1. Conserved structural architecture of FENs

FENs are a family of structurally and functionally conserved proteins. Sequence comparison between species reveal a number of conserved residues, which contribute to important functional domains within the protein, as well as conserved active site residues. The structure of several FENs has been solved and these crystal structures reveal a common architecture between homologues, which has a central β-sheet surrounded by α-helices and loops. The crystal structures of six FEN homologues are shown in Figure 1-14. In some bacteria such as *Thermus aquaticus*, FENs exist as a domain of DNA polymerase I (Figure 1-14). This core FEN structure consists of several conserved domains and motifs, which will be discussed in more detail below and include a β-pin, helix-2/3-turn-helix (H2/3TH), hydrophobic wedge and a helical arch or loop, some of which are completely conserved with all FEN homologues. These are highlighted on the structure of bacteriophage T5FEN in Figure 1-15A and a close up of the active site is shown in Figure 1-15B. Certain motifs and domains, such as the 3' flap binding pocket and C-terminal domain for protein: protein interactions are only observed in FENs from higher organisms (eukaryotes and archaea), whilst a nuclear localisation signal (NLS) is only present in FENs from eukaryotes. FENs can be further subdivided into four groups based on the presence of these various structural features: 1) eukaryotic and archaeal FENs, 2) bacterial

DNA polymerases, 3) bacterial FEN and 4) viral FENs. Examples from each group are provided in Table 1-2.

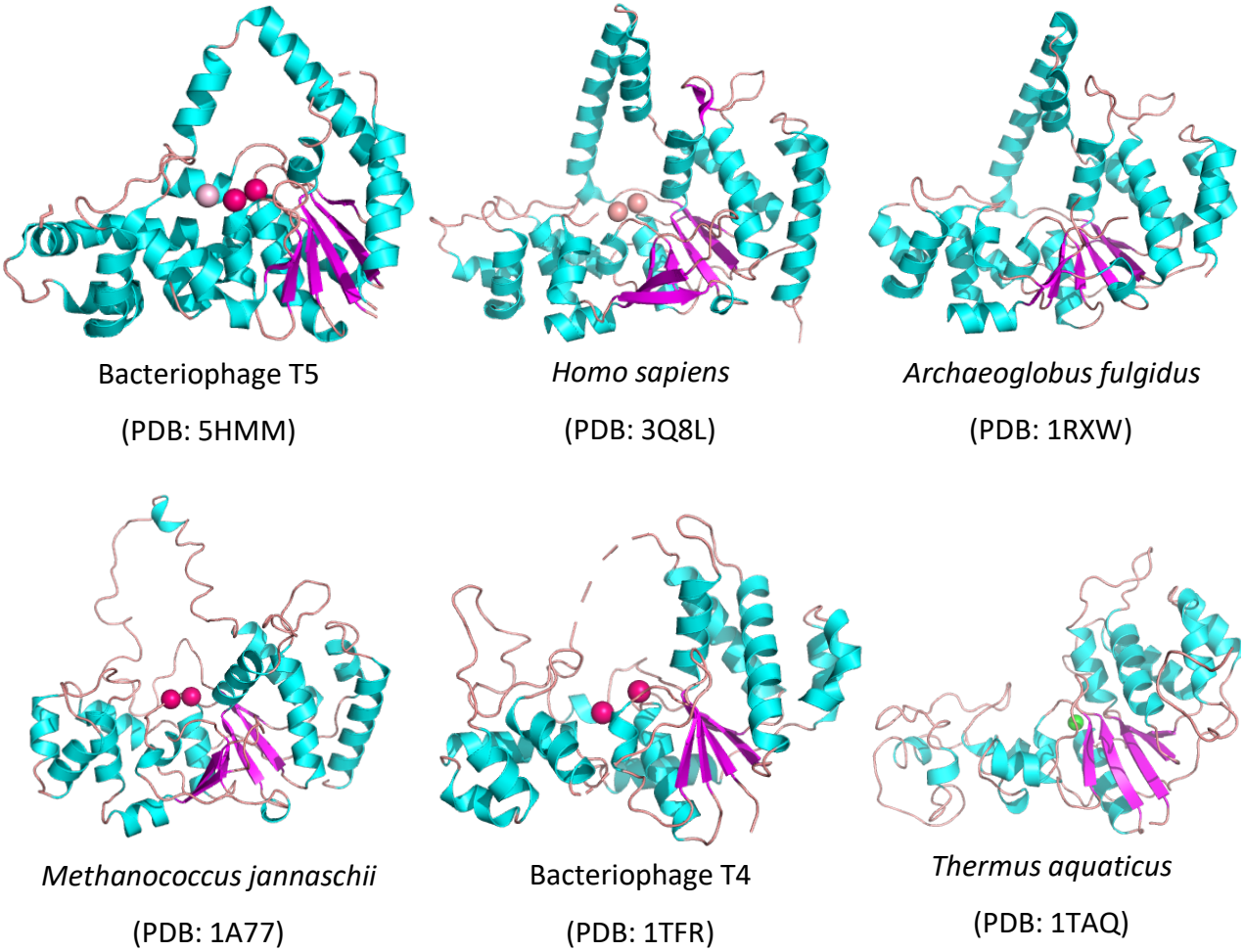


Figure 1-14: Crystal structures of FEN homologues

The structure of six FENs solved by x-ray crystallography are displayed as cartoons and highlight a common structural architecture observed between homologues. Typically, β-sheets (magenta ribbons) are surrounded by α-helices (cyan) and loops (wheat). All structures are of individual FEN proteins from the organisms listed, except Thermus aquaticus, where the FEN domain of the DNA polymerase structure has been extracted and shown here. The helical arch region of bacteriophage T5FEN, Homo sapiens FEN-1 and *Archaeoglobus fulgidus* is seen as two ordered α-helices, whereas the arch region of *Methanococcus jannaschii* FEN adopts a loop-out conformation. In bacteriophage T4 RNase H and the FEN domain of Thermus aquaticus DNA polymerase this region could not be fully modelled.

A

B

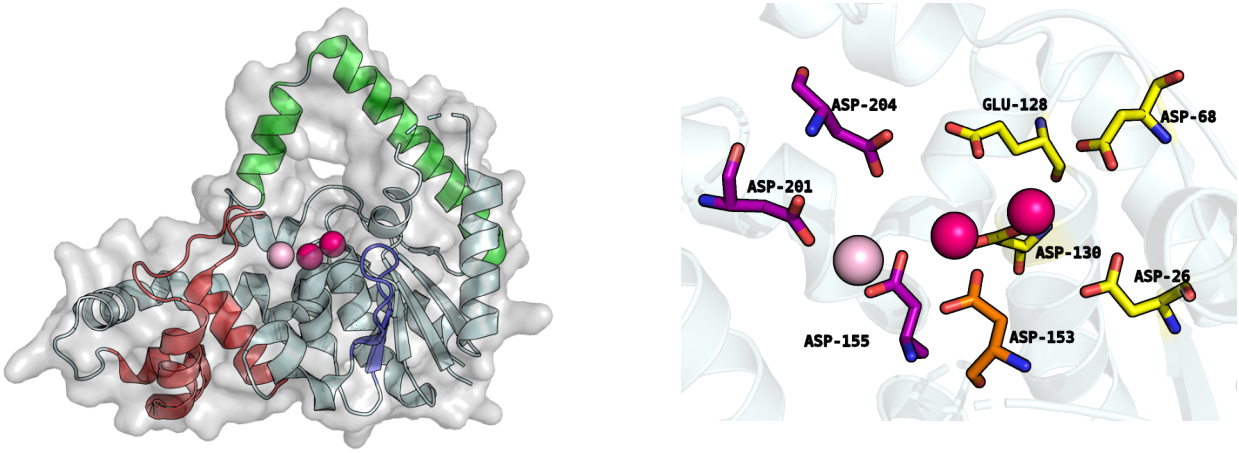


Figure 1-15: Structural features of T5FEN

(A) Crystal structure of T5FEN (PDB: 5HMM) with conserved motifs highlighted: helical arch (green), helix-3-turn-helix motif (red) and β- pin (dark blue). (B) The active site sits at the base of the helical arch and consists of eight conserved carboxylate residues that mediate binding of catalytic Mg^{2+} ions (in pink). In this structure, three Mg^{2+} ions are found bound to two catalytic sites (Site I, dark pink; Site II, light pink). Residues that interact with Site I Mg^{2+} ions are yellow, Site II Mg^{2+} ions are purple and both Site I and Site II Mg^{2+} ions are orange (AlMalki *et al.*, 2016).

1.6.2. Conserved structural features of FENs

1.6.2.1. Helical arch

Perhaps the most apparent structural feature observed in FENs is the loop region that sits above the active site that forms a small but flexible opening. This highly variable and flexible motif has several names in the literature including helical arch, helical gateway and helical clamp, based on structural observations of FEN homologues and proposed function. As seen in Figure 1-14, in several crystal structures, such as T5FEN, *Homo sapiens* FEN-1 and *Archaeoglobus fulgidus* FEN, this motif is seen as two ordered helices, one containing positively charged residues and the other hydrophobic residues (Tsutakawa *et al.*, 2011; AlMalki *et al.*, 2016; Chapados *et al.*, 2004). These are involved in DNA interactions, where the positively charged residues interact with the negatively charged backbone and hydrophobic residues interact with DNA bases. When ordered, the helical arch is only large enough to accommodate ssDNA (Ceska *et al.*, 1996). However, the arch region in other FEN structure in a looped out conformation, such as *Methanococcus jannaschii* FEN-1 (Hwang *et al.*, 1998),

whist in other structures this motif could not be modelled, such as T4 RNase H (Mueser *et al.*, 1996), and Taq DNA polymerase I (Kim *et al.*, 1995). Furthermore, in a recent structure of T5 FEN, one asymmetric unit had an arch with an ordered conformation, whilst the other adopted a looped-out structure indicating a highly variable and flexible motif within individual FENs (Figure 1-16) (AlMalki *et al.*, 2016).

Table 1-2: Distinct classes of FENs

FEN subfamily	FEN name
Eukaryotic FEN-1	
Homo sapiens FEN-1	FEN-1
Mus musculus FEN-1	FEN-1
Saccharomyces cerevisiae RAD27	RAD27/ FEN-1
Schizosaccharomyces pombe Rad2	Rad2/ FEN-1
Archaeal FEN-1	
Archaeoglobus fulgidus FEN	FEN-1
Methanococcus jannaschii FEN	FEN-1
Pyrococcus furiosus	FEN-1
Bacterial DNA polymerase	
Escherichia coli	DNA polymerase I FEN
Staphylococcus aureus	DNA polymerase I FEN
Thermus aquaticus	DNA polymerase I FEN
Streptococcus pneumoniae	DNA polymerase I FEN
Haemophilus influenza	DNA polymerase I FEN
Bacterial FEN	
Escherichia coli	ExoIX / FEN Xni
Bacillus subtilis	YpcP
Staphylococcus aureus	*Sa*FEN
Bacteriophage FEN	
T5 bacteriophage	T5FEN
T7 bacteriophage	T7 gp6
T4 bacteriophage	RNase H

1.6.2.2. Active site

The active site sits within a conserved globular domain and co-ordinates divalent metal ions essential for FEN activity, through the side chains of seven or eight conserved carboxylate-containing amino acids. Two metal-ion binding sites bind at least 2- 3 catalytic metal ions (Feng *et al.*, 2004; Syson *et al.*, 2008). As determined through isothermal titration calorimetry (ITC) experiments in T5FEN, these binding sites have different affinities for divalent metal ions (Feng *et al.*, 2004). Mutational studies of active site residues also indicate that these binding sites may contribute to endonuclease and exonuclease activity differently. For example, in T5FEN, loss of binding site I (high affinity binding site) results in a complete loss of nuclease activity, whilst loss of binding site II (low affinity binding site) causes a selective loss of exonuclease activity (Feng *et al.*, 2004).

As discussed above, several bacteria possess an independent FEN protein in addition to or in the place of the FEN domain of DNA polymerase I. Within this class, two further discrete classes exist and are separated by the number of conserved active site residues. *Staphylococcus aureus* FEN has a fully conserved active site with two independent metal ion binding sites whilst *E. coli* ExoX1, encoded by the *xni* gene, lacks the aspartates for metal binding site II, although two Mg^{2+} ions are found bound, 2.5 Å apart to site I in a crystal structure of this (Anstey-Gilbert *et al.*, 2013; Allen *et al.*, 2009).

1.6.2.3. Helix-2/3-turn helix

Another conserved domain, the helix-2/3-turn helix motif (H2/3TH) is a potassium ion binding motif (Figure 1-15) that interacts and binds to the phosphodiester backbone of dsDNA downstream of the cleavage site (Hosfield *et al.*, 1998; Tsutakawa *et al.*, 2011).

1.6.2.4. Beta pin

This motif contains arginine, asparagine and lysine residues which form hydrogen bonds with the DNA (Figure 1-15). The residues are orientated such as to impose a bend of 100° in a nicked dsDNA double-flap substrate (Tsutakawa *et al.*, 2011).

1.6.2.5. 3’ binding pocket

Absent from bacteriophage FENs, this motif is composed of negatively charged acidic residues that provide binding for a 1 nt 3’ flap (Tsutakawa *et al.*, 2011). Mutagenesis of specific residues in this motif demonstrate the importance of this for substrate recognition and stringent cleavage precisely 1 nt into the dsDNA region (Friedrich-Heineken and Hübscher, 2004).

1.6.2.6. C-terminal interaction domain

The C-terminal interaction domain is present in eukaryotic and archaeal FENs to permit interactions with other proteins to enable a tight regulation of their functions to better protect genome integrity. Human FEN-1 has been shown to interact with more than 30 proteins including PCNA, WRN-1, RPA and the Rad9-Rad1-Hus1 heterotrimeric complex (Zheng *et al.*, 2011).

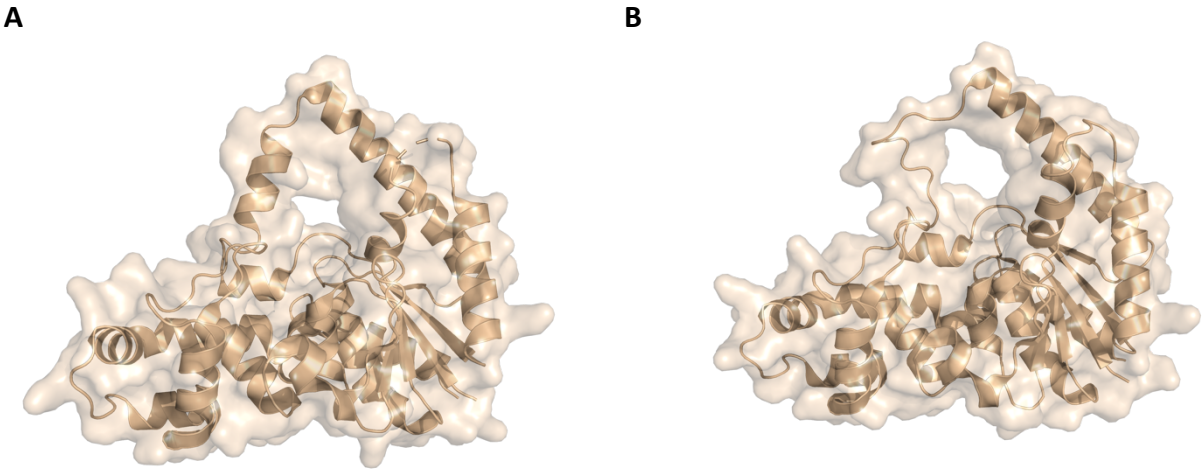


Figure 1-16: Alternative arch conformations observed in T5FEN

The most recent crystal structure of T5FEN (PDB: 5HMM) has two protein molecules per asymmetric unit. The arch of each of these molecules is in a different conformation. The surface of each molecule is shown to visualise the effect of these arch structures on the overall protein size (A) Ordered helical arch composed of two α-helices. This leaves a small aperture above the base of this region. (B) Looped-out arch composed of one α-helix and a number of residues that adopt a looped conformation. This results in a larger aperture through the protein in this region.

1.7. Mechanism of FEN: DNA interaction

There has been much debate regarding the mechanism by which FENs are able to recognize and bind to their specific substrates and precisely cleave the DNA, usually leaving a nick that can be directly ligated by DNA ligase. Several models have been proposed for the mechanism of FEN: DNA interaction based on biochemical and structural studies. The two predominating models are the tracking model and threading model. The tracking model proposes that FENs specifically recognize and bind the end of a 5′ ss arm and then slide or track down via an open cleft in a bi-directional manner to the cleavage site as shown in Figure 1-17A (Murante *et al.*, 1995). Whilst the threading model proposes that FENs recognise and bind the dsDNA junction at the base of the flap and the flap is then threaded through an opening in the protein as shown in Figure 1-17B. Both of these models have a requirement of a free 5′ end although much of the evidence used to corroborate one model could also be used in support of the other.

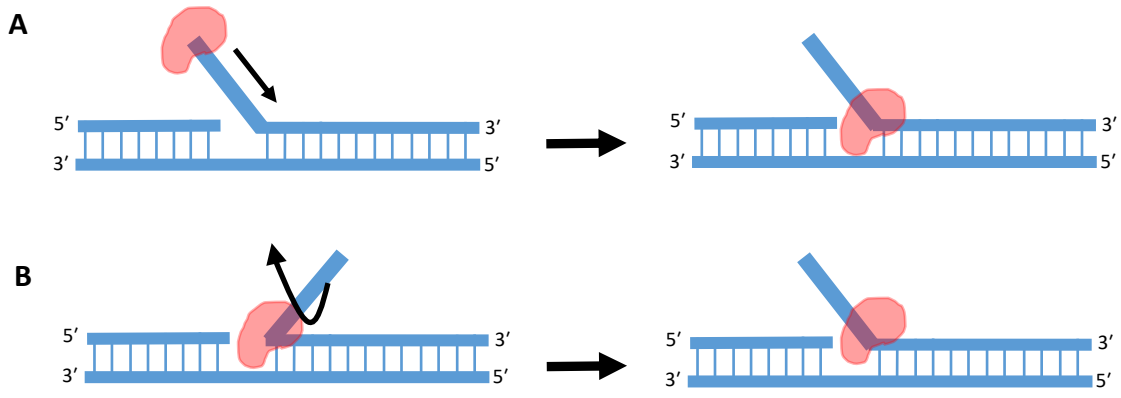


Figure 1-17: Proposed models of the mechanism of the interaction of FEN and DNA

(A) In the tracking model, FEN recognises and binds to the end of the 5′ single-stranded flap. After binding, FEN tracks down to the cleavage site for flap cleavage. (B) In the threading model, FEN recognises and binds to the base of the 5′ flap. The flap is threaded through the arch for cleavage. Based on recent crystallographic FEN studies, the threading model is now widely accepted as the mechanism for FEN: DNA interaction.

Early experiments examining FEN: DNA interactions looked at the ability of FENs to cleave substrates with modified 5′ flaps. They showed that FEN cleavage is inhibited when the ends of the 5′ flaps are blocked with biotin-streptavidin complexes, primers or hairpin formation

(Murante *et al.*, 1995; Xu *et al.*, 2001; Bornarth *et al.*, 1999). However, there is still a degree of tolerability in these 5′ flaps as cleavage could still occur with biotinylated flaps alone or when small adducts mimicking DNA damage were placed at various points along the flap (Murante *et al.*, 1995; Barnes *et al.*, 1996), an important feature regarding the roles of FENs in DNA damage removal. In support of the threading model, Xu *et al*., (2001) examined the kinetics of the cleavage of 5′ flaps of different lengths by the FEN domain of *E. coli* DNA polymerase I. They found that cleavage of a 30 nt flap was considerably less efficient compared to a 10 nt flap and argued that sliding from the 5′ end could not result in this dramatic difference and thus the rate limiting stage was the threading of the tail through the enzyme once it had bound (Xu *et al.*, 2001). Furthermore, human FEN-1 binds to substrates with the similar affinity regardless of whether the end of the 5′ flap is accessible (Gloor et al, 2010), including to bubble structure (Hohl et al, 2007), which FEN-1 itself is unable to cleave, pointing to the idea that FENs first recognise the dsDNA at the base of the flap as proposed in the threading model.

In support of both models, the crystal structure of bacteriophage T4 RNase H co-crystallized with a pseudo-Y substrate was solved in 2007 (Figure 1-18A) (Devos *et al.*, 2007). The structure implies that the 5′ flap is enclosed by a disordered loop of protein, corresponding to the arch above the active site, trapping the DNA. Unfortunately, the arch itself was too disordered to model and residues 91-96 (amino acid sequence: GKAREE) were not accounted for in the model (Devos *et al.*, 2007). Although the inability to model this region could indicate a transient nature of the arch structure.

In 2012, Patel *et al.* updated the more widely accepted threading model into a disorder-thread-order model. This model proposes that the FEN recognizes and binds to the substrate at the dsDNA junction and the 5′ flap threads through a disordered helical arch, which becomes ordered when threading is completed. They showed that FENs were able to cleave biotin-streptavidin modified flaps regardless of flap length if the FEN had been able to bind first before streptavidin addition implying the flaps had already been threaded (Patel *et al.*, 2012).

More recent structural data provides the final answers to this long running debate. In 2016, ground-breaking co-crystal structures with T5FEN and DNA were obtained through the generation of active-site mutants which were capable of binding DNA but not cleavage (AlMalki *et al.*, 2016). In the first 'pre-threaded' complex, the 5' end of overhang DNA is seen poised to enter through a gap in the helical arch where it interacts with residues around this gap. A fully threaded structure was also obtained which clearly shows the 5' end of the overhang substrate threaded through an ordered helical arch Figure 1-18B. Furthermore, the authors obtained crystal structures of T5FEN alone that show clearly that the helical arch can adopt both an ordered helical arch conformation or a looped out conformation as shown in Figure 1-16, hence providing structural evidence on the disorder-thread-order model (AlMalki *et al.*, 2016). The structure of human FEN-1 with a double-flap DNA substrate was solved soon after and provided more answers regarding the energetics of this interaction (Figure 1-18C) (Tsutakawa *et al.*, 2017). The structure showed that the backbone of the 5' flap was inverted after the scissile phosphodiester and positively charged steering residues help move flap through the arch to position the first dsDNA base pairs at the active site. Mutational studies have found that these phosphate steering residues exhibit conservation in FEN homologues (Tsutakawa et al, 2017).

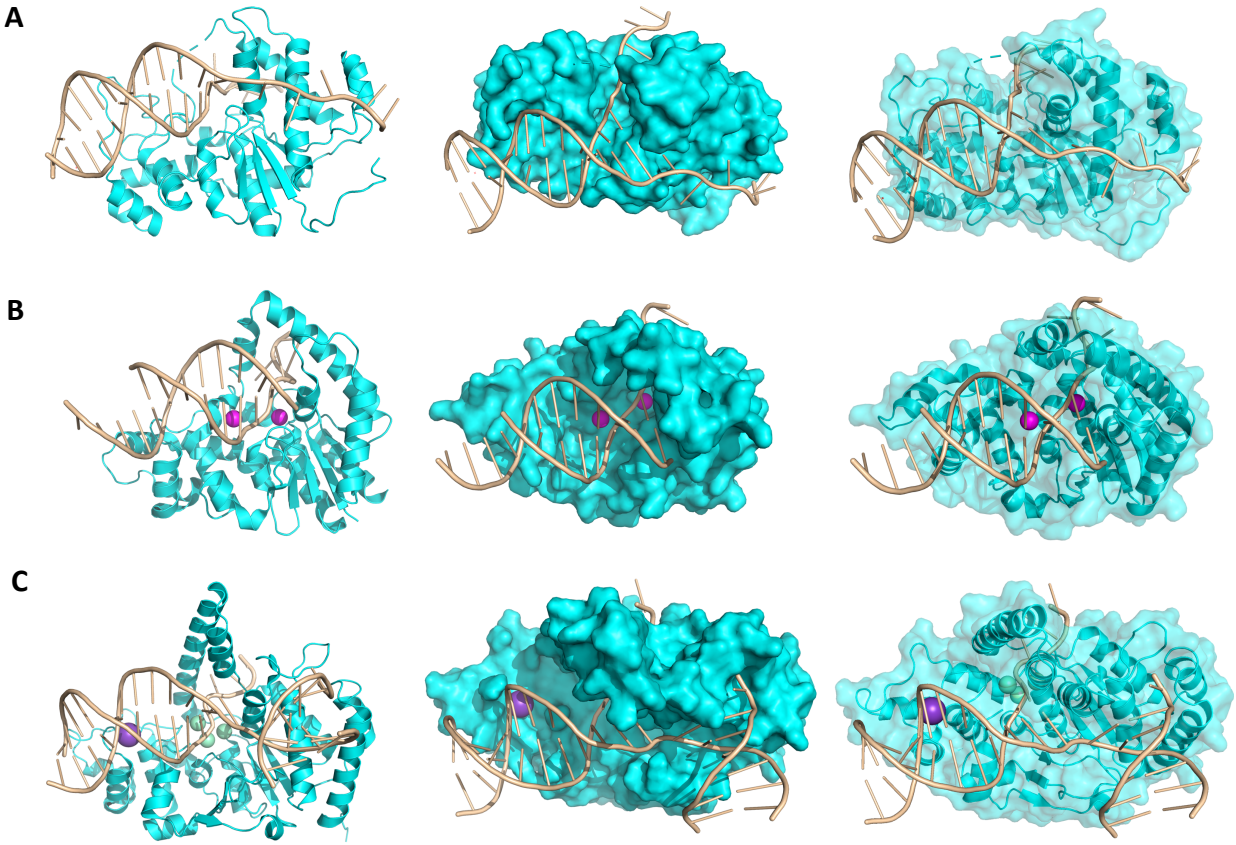


Figure 1-18: Interaction of FENs with 5' flap DNA substrates

Co-crystal structures of FEN homologues with DNA. (A) T4 RNase H with pseudo–Y DNA (PDB: 2IHN)(Devos *et al.*, 2007). (B) T5FEN with 5' over-hang DNA (PDB: 5HNK)(AlMalki *et al.*, 2016). (C) Human FEN-1 with 5' flap DNA (PDB: 5KSE) (Tsutakawa *et al.*, 2017). In the T4 RNase H crystal structure, the DNA looks like it could pass through the arch region, although this density could not be modelled. In T5FEN and human FEN-1 crystal structures, the 5' flap can clearly be seen threading through an ordered helical archway.

1.8. PhD project

1.8.1. Project rationale

This PhD project is run in collaboration with an industrial partner, Binx Health, who have developed the point-of-care testing device, the IO © device presented in section 1.3.6.3. The device is FDA-approved and provides detection of *Chlamydia trachomatis* and *Neisseria gonorrhoeae* in less than 30 minutes (Van Der Pol *et al.*, 2020). This PCR-based device utilizes Taq DNA polymerase I for DNA amplification. A redox-active oligonucleotide probe (5' ferrocene-labelled) is used to facilitate detection due to cleavage by bacteriophage T7 gp6 resulting in release of the ferrocene-labelled mononucleotide. This can be distinguished from the un-cleaved probe by electrochemical methods. However, little optimization of the

conditions within the device has been carried out for the detection stage using T7 gp6 (Personal communication, David Pearce, Binx Health).

As described in section 1.3, FENs have been shown to have extensive uses in molecular biology and biotechnology applications. Prokaryotic and bacteriophage FENs are the most widely used homologues and this is probably due to the higher levels of exonuclease activity and greater substrate tolerance (Lyamichev *et al.*, 1993; Kaiser *et al.*, 1999; Williams *et al.*, 2007; Oates, 2016; Lau, 2017). T7 gp6 was first identified over 50 years ago where it was shown to have an essential function in the processing of Okazaki fragments and breakdown of host DNA within phage-infected cells as described in section 1.4.1, as well as other documented roles within T7 genetic recombination and maturation of concatemeric T7 DNA (Center *et al.*, 1970; Sadowski and Kerr, 1970; Lee and Miller, 1974; Serwer *et al.*, 1990).

T7 gp6 was first purified in 1972, where the exonuclease activity was characterized (Kerr and Sadowski, 1972). This work demonstrated that the enzyme hydrolyses duplex DNA from the 5' terminus to generate predominately 5'-mononucleotides. Furthermore, activity of T7 gp6 was shown to be dependent on Mg^{2+} or Mn^{2+} ions and greatly stimulated by KCl (Kerr and Sadowski, 1972). The enzyme has been commercially exploited within phosphorothioate site-directed mutagenesis as well as several more recently published electrochemical detection methods (Nakamaye and Eckstein, 1986; Sayers *et al.*, 1988; Sayers *et al.*, 1992; Xia *et al.*, 2020; Li *et al.*, 2021; Zhou *et al.*, 2021). However, in the past 20 years, only one primary research article has been published which characterizes both the exonuclease and flap endonuclease activity of the protein through gel-based assays (Mitsunobu *et al.*, 2014). In that study, whilst the catalytic parameters (K_M, k_{cat}) were determined for blunt-ended (exonuclease) and overhang (endonuclease) substrates, other substrates such as nicked and flapped DNA were shown to be cleaved by gp6 but catalytic parameters were not determined (Mitsunobu *et al.*, 2014). Additionally, the authors demonstrated that the exonuclease activity was inhibited by NaCl, however, the effect on the endonuclease activity as well the characterization of monovalent and divalent cation dependence was not investigated (Mitsunobu *et al.*, 2014).

In contrast, another bacteriophage FEN, T5FEN, has been extensively biochemically and structurally characterized, becoming the archetypal bacteriophage FEN. Whilst the two proteins only share 15% sequence identity and 32% sequence similarity, there is predicted to be a high level of structural conservation between the proteins (Figure 5-24). As shown in section 1.3, T5FEN has extensive and well-characterized uses within molecular biology and biotechnology. It could be argued that this greater adoption of T5FEN within these techniques is due to its greater prominence in the literature.

Therefore, this project was developed to improve the characterization of FENs used within the IO device, with an emphasis of T7 gp6. Improvements in the understanding in the enzymology of T7 gp6 and other FENs, may contribute in the future to improved biotechnological use of FENs, both within the IO device and other applications.

1.8.2. Project aims

1. Biochemical characterisation of T7 gp6 cleavage specificity (Chapter 3).

Using a heat-shock *E. coli* expression system, the wild-type enzyme was overexpressed and purified to >95% purity. The endonuclease and exonuclease activity of the purified protein was assessed for various substrates using UV, FRET and gel-based assays, including determination of catalytic parameters (K_M, k_{cat}) with the FRET assay. Monovalent and divalent salt concentrations were optimized as well as pH. The DNA binding affinity was determined using the electrophoretic mobility shift assay.

2. Produce an active site T7 gp6 mutant that is catalytically inert but can bind DNA (Chapter 4).

Currently, there are no published experimental structures of T7 gp6, alone or bound to DNA. In this study, three active site point mutants were constructed in T7 gp6, in which aspartic acid was substituted with lysine, were cloned, overexpressed and biochemically characterized for nuclease activity and DNA binding affinity. Active site mutants that are inert but facilitate DNA binding could be suitable for co-crystallization trials with DNA and catalytically relevant Mg^{2+} ions.

3. Determine the effect of mutations in the helix-3-turn helix motif of nuclease activity and DNA binding affinity of T7 gp6 (Chapter 4).

In T5FEN, chemical modification of residue 202 with a thiol-modifying compound results in a defect in endonuclease but not exonuclease activity (Zhang, 2012). The same principal was applied to T7 gp6, but with amino acid modification. Two T7 gp6 mutants were cloned and overexpressed. These mutants had a point mutation in the helix-3-turn-helix motif which introduced an amino acid with a bulky side chain (Ile200Arg and Thr201Trp). Furthermore, the corresponding T5FEN mutants were generated (Leu202Arg and Gly203Trp) for comparison. The H3TH mutants were biochemically characterized for nuclease activity and DNA binding affinity. Enzymes with altered ratios of nuclease activity are of interest as they may represent improved enzymes in some applications.

4. Structurally characterize relevant 5' nuclease homologues (Chapter 5)

Crystal trials were set with the WT and active site mutants alone and with DNA. Crystallization was unsuccessful and the predicted structure of T7 gp6 was determined instead. Structural characterization of relevant FEN homologues was performed and three novel structures were obtained. These were the (1,2) N-terminal FEN domain of *Thermus aquaticus* polymerase I carrying an active site mutation co-crystallized with double-flap DNA substrates and (3) a T5FEN active site mutant.

Chapter 2 - Materials and methods

2.1. *E. coli* growth media

Growth media used for plating and growth of *E. coli* was prepared and used according to Table 2-1. To make agar plates, 1.5% (w/v) agar was added prior to autoclaving. For antibiotic selection, carbenicillin (100 μg/mL) was used throughout.

Table 2-1: Growth media utilized

Media	Components or manufacturer	Use
LB	10 g/L tryptone, 5 g/L yeast extract, 5 g/L NaCl	pTTQ18-T5 plates and starter cultures
2YT	16 g/L tryptone, 10 g/L yeast extract, 5 g/L NaCl	pJONEX4 plates and starter cultures
5YT	40 g/L tryptone, 25 g/L yeast extract, 5 g/L NaCl	pJONEX4 protein expression
AIM-2YT	Commercially produced by Formedium	pTTQ18-T5 protein expression
AIM- SB	Commercially produced by Formedium	pTTQ18-T5 protein expression

AIM, autoinduction media; SB, SuperBroth.

2.2. Cloning techniques

2.2.1. Plasmids, bacterial strains and growth conditions

E. coli strains and genotypes used in this work are summarised in Table 2-2. The plasmids used and generated in this work are summarised in Table 2-3.

Table 2-2: *E. coli* strains used with genotype

E. coli strain	Genotype	Source or reference
BL21	B, *F⁻ ompT hsdS(rb⁻mb⁻)gal dcm*	Agilent #200133
XL1Blue	*recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac [F´ proAB lacIqZΔM15 Tn10 (Tetr)]*	Agilent #200249
M72(λ) / ΔH1+	K-12 *F⁻ lac⁻ amSm^R*(cI857N7N53ΔH1*bio⁻*)	Castellazzi *et al.*, 1972
W3110	K-12 *F⁻ λ⁻ mcrA mcrB IN(rrnD, rrnE)*	ATCC #27325

Table 2-3: Plasmids used in this project

Vector	Gene	Provided or generated?
pJONEX4	T7 gp6	Provided
	T7 gp6 Asp160Lys	Generated by site-directed mutagenesis of pJONEX4.T7gp6
	T7 gp6 Asp162Lys	
	T7 gp6 Ile200Arg	
	T7 gp6 Thr201Trp	
	T7 gp6 Asp202Lys	
	T5FENΔ19	Provided
	T5FENΔ19 Asp153Lys	Provided
	T5FENΔ19 Asp155Lys	Provided
	T5FENΔ19 Leu202Arg	Generated by site-directed mutagenesis of pJONEX4.T5FENΔ19
	T5FENΔ19 Gly203Trp	
PTTQ18-T5	T7 gp6	Generated by sub-cloning from pJONEX4.T7gp6

2.2.1.1. pJONEX4

Expression from the heat-inducible pJONEX4 (Sayers and Eckstein, 1991) is under the control of λP_L (leftward promoter) and was carried out in the *E. coli* strains M72(λ) or ΔH1+. These strains are isogenic with each other (Castellazzi *et al.*, 1972) and encode a thermolabile repressor of the λ promoter, cI857 (Remaut *et al.*, 1981). The antisense messenger RNA (mRNA) from the downstream lactose promoter (P_{lac}) in the vector counters any leakage from λP_L to give a tight control of expression. Cells transformed with pJONEX4 were grown at 28°C with media supplemented with 0.1 mM isopropyl β- d-1-thiogalactopyranoside (IPTG) and protein expression was induced at 42°C. The β-lactamase gene is used as the selectable marker.

2.2.1.2. pTTQ18-T5

Expression from the lactose inducible pTTQ18-T5 is under the control of the bacteriophage T5 promoter and was carried out in the *E. coli* strains BL21, W3110 and Xl1blue. Cells transformed with pTTQ18-T5 were grown at 37°C in media which utilized vegetable derived tryptone. The β-lactamase gene is used as the selectable marker.

2.2.2. Preparation of calcium chloride chemically competent cells

A glycerol stock of *E. coli* was streaked onto a LB agar plate and grown for 48 h at 28°C to obtain single colonies. A single colony was used to inoculate 10 mL LB media and grown at 28°C overnight, until cloudy. The 10 mL culture was used to inoculate 100 mL LB and grown until A_{600} = 0.4. The culture was centrifuged at 3000 x *g* for 20 minutes at 4°C to pellet the cells. The pellet was resuspended in 40 mL ice cold 0.1 mM $CaCl_2$ and incubated overnight on ice at 4°C. The culture was centrifuged at 1000 x *g* for 20 minutes at 4°C and the resulting pellet was gently resuspended in 6 mL ice cold 85 mM $CaCl_2$/ 15% (v/v) glycerol before aliquoting 100 μL stocks for storage at -80°C.

2.2.3. Transformation of plasmid into cells

Aliquots of 100 μL chemically competent cells were thawed on ice and 100 ng of plasmid DNA was added. Cells were incubated on ice for 45 minutes. Cells were placed at room temperature for 5 minutes (pJONEX4) or were heat shocked at 42°C for 30 seconds (pTTQ18-T5) before plating on appropriate agar plates as summarised in Table 2-1. Plates were incubated at 28°C (pJONEX4) or 37°C (pTTQ18-T5) to observe growth of single colonies. This took between 24 to 48 hours. If required, glycerol stocks were made by selecting single colonies to inoculate a 3 mL starter culture of appropriate media which was incubated, whilst shaking until visible growth had occurred. Glycerol was added to make give a final 30% (v/v) concentration and aliquots were stored in cryogenic vials at -80°C.

2.2.4. Plasmid DNA preparation

Minipreps were carried out to obtain plasmid DNA stocks for long term storage at -20°C. These were carried out using the Monarch® Plasmid Miniprep kits (New England Biolabs) with the manufacturer's standard protocol using 5 mL cultures in appropriate media. After plasmid preparation, the quantity and quality of DNA was assessed spectrophotometrically using a NanoDrop spectrophotometer.

2.2.5. Sub-cloning

The gene insert from pJONEX4 was sub-cloned into pTTQ18-T5 using the restriction sites *Hin*dIII and *Eco*RI with NEB high fidelity enzymes. The restriction digest was carried out in

CutSmart buffer according to the NEB standard protocol with 1 μg DNA and 20 units of each enzyme. Reactions were incubated at 37°C for 15 minutes and heat inactivated at 80°C for 20 minutes. Digestion products were run on a 1% agarose gel and bands were excised and purified from the gel using a gel extraction kit (New England Biolabs). The vector and insert were ligated together at a molar ratio of 3 insert: 1 vector using T4 DNA ligase and 100 ng DNA in total was ligated. The ligated DNA was transformed into *E. coli* XL1blue cloning strain before plating on LB agar plates and incubated at 37°C until single colonies were observed. The ligation of insert was confirmed by selecting 5 colonies for sequencing after extraction of plasmid DNA with a MiniPrep kit (NEB).

2.2.6. Cloning of mutant FENs by site-directed mutagenesis

Various point mutations were introduced through site-directed mutagenesis, to the corresponding wild-type pJONEX4 derivative as the template. Mutagenesis was carried out using the Q5® site-directed mutagenesis kit (New England Biolabs) with the standard protocol with the PCR stage summarised in

Table 2-4. Overlapping primers were purchased from Eurofins and the sequences and annealing temperatures (T_a) are summarised in Table 2-5. After the PCR stage, a 1% agarose gel was run to visualise the presence of the amplified DNA at the correct size before proceeding to the Kinase-Ligase-*Dpn1* (KLD) stage for ligation of PCR product and removal of template DNA. After transformation, 5 colonies were selected for miniprep and sequencing for the presence of the desired mutation.

Table 2-4: PCR stage of site-directed mutagenesis

PCR step	Temperature	Time
Initial denaturation	98°C	30 seconds
25 cycles	98°C	10 seconds
	T_a	30 seconds
	72°C	2.5 minutes
Final extension	72°C	5 minutes
Hold	10°C	Hold

2.2.7. Sequencing plasmids

Sequencing of plasmid inserts was carried out by The Genomics Core Facility, University of Sheffield. The sequencing primers used are summarised in Table 2-6. Electropherogram results were visualised and where possible, any omissions manually interpreted. Results were compared to the published DNA sequences for T7 gp6 (NC_001604.1) or T5FEN (NC_005859.1) obtained from their complete genome in the NCBI database.

Table 2-5. Primers used for site-directed mutagenesis

FEN	Mutation	Primer sequence (5' to 3')	T_a (°C)
T7 gp6	Asp160Lys	F: CATCTCTTGCAAAAAGGACTTTAAGACC	55
		R: ATTACAGCCTTACGAGCAC	
T7 gp6	Asp162Lys	F: TTGCGATAAGAAATTTAAGACCATC	55
		R: GAGATGATTACAGCCTTAC	
T7 gp6	Ile200Arg	F: CAAGGGTGACCGGACTGATGGTTAC	58
		R: ATGGTCTGGAAGAGGTGC	
T7 gp6	Thr201Trp	F: GGGTGACATCTGGGATGGTTACTCAG	56
		R: TTGATGGTCTGGAAGAGG	
T7 gp6	Asp202Lys	F: TGACATCACTAAAGGTTACTCAGGG	58
		R: CCCTTGATGGTCTGGAAG	
T5 FENΔ19	Leu202Arg	F: TATGGGAGATCGGGGAGATAATATTCGTG	56
		R: ATTGCTTTCAGGGAGATAAAC	
T5 FENΔ19	Gly203Trp	F: GGGAGATCTATGGGATAATATTCGTG	53
		R: ATAATTGCTTTCAGGGAG	

2.3. Overexpression of FEN proteins

Single colonies were selected from plates of freshly transformed cells or streaked glycerol stocks to inoculate a 5 mL starter culture of appropriate media as summarised in Table 2-1. The starter cultures were incubated whilst shaking at 28°C (pJONEX4) or 37°C (pTTQ18-T5) for a few hours until they appeared lightly turbid. These were used at a 1:1000 dilution to inoculate a culture for small-scale expression trials or large-scale protein expression. Large scale expression was carried out from pJONEX4 in a 5 L fermenter in rich 5YT media. This media was supplemented with antifoam-204 (0.01% v/v) and the culture was maintained at

23°C overnight, whilst stirring at 300 revolutions per minute (RPM), until absorbance (600 nm) was 2.0, indicating the cells were in mid-log phase. Heat shock protein expression was induced by increasing the temperature to 42°C for 2 hours. After this, cells were either harvested immediately or the temperature was decreased to 20°C overnight, depending on the protein. Small-scale expression studies have shown that the relative level of expression for some proteins studied (e.g. WT T7 gp6) drops after overnight incubation whilst other proteins the relative level of expression remains the same or increases whilst biomass also increases. After induction, cultures were centrifuged at 12,000 x *g* for 30 minutes at 4°C and resultant cell pellets were stored at -80°C in 10 g aliquots.

Table 2-6: Sequencing primers for plasmid inserts

Primer	Primer sequence (5' to 3')	Plasmid sequenced
M13 forward	TGTAAAACGACGGCCAGT	pJONEX4, pTTQ18-T5
M13 reverse	CAGGAAACAGCTATGACC	pJONEX4
T5 promoter	TAATACGACTCACTATAGGG	pTTQ18-T5

2.4. Purification of protein from cell pellet

All proteins were purified from the soluble fraction and all stages of purification were carried out at 4°C except for TaqFEN, which was purified at room temperature. Buffers used for lysis and purification are summarised in Table 2-7.

2.4.1. Lysis of cell pellet

Cell pellets were re-suspended in 5 mL/g of lysis buffer. Lysozyme was added at a final concentration of 1 mg per mL lysate and the cell suspension was stirred at room temperature for an hour until viscous. The suspension was placed on ice and 500 μg per mL lysate of sodium deoxycholate, 10 μg per mL of 4-(2-aminoethyl)benzenesulfonyl fluoride hydrochloride (AEBSF) and 1 mM DTT were added as concentrated solutions. The cell lysate was stirred on ice for a further hour. Sonication was performed, whilst incubating on ice, for 3 X 10 seconds and the cell suspension was spun at 30,000 x *g* for 30 minutes to separate soluble and insoluble material. An SDS-PAGE gel was run to confirm location of protein in the soluble or insoluble fraction.

Table 2-7: Buffers used for lysis and purification

Buffer name	Components
Lysis buffer	25 mM Tris-HCl pH 8, 100 mM NaCl, 5% (v/v) glycerol, 2 mM ethylenediaminetetraacetic acid (EDTA)
$KP_n/0$	25 mM potassium phosphate pH n, 5% (v/v) glycerol, 1 mM dithiothreitol (DTT), 1 mM EDTA
$KP_n/100$	25 mM potassium phosphate pH n, 5% (v/v) glycerol, 1 mM dithiothreitol (DTT), 1 mM EDTA, 100 mM NaCl
$KP_n/1000$	25 mM potassium phosphate pH n, 5% (v/v) glycerol, 1 mM DTT, 1 mM EDTA, 1 M NaCl
Tris8/0	25 mM Tris-HCl pH 8, 5% (v/v) glycerol, 1 mM DTT, 1 mM EDTA
Tris8/1000	25 mM Tris-HCl pH 8, 5% (v/v) glycerol, 1 mM DTT, 1 mM EDTA, 1 M NaCl
SEC buffer	25 mM Tris-HCl, 5% (v/v) glycerol, 500 mM NaCl, 2 mM DTT

n represents pH; SEC, size exclusion chromatography

2.4.2. Removal of nucleic acids

Nucleic acids were removed by precipitation with polyethyleneimine (PEI), pH 8. For this, 500 mM ammonium sulphate was added to the soluble lysate to prevent precipitation of proteins by PEI and 70 μL per mL of lysate of 5% PEI (v/v) was added whilst stirring. This solution was centrifuged at 30,000 x *g* for 20 minutes. To remove residual soluble PEI in the resultant supernatant, proteins were precipitated by adding solid ammonium sulphate to a final concentration of 3.5 M and this was centrifuged at 30,000 x *g* for 20 minutes. The pellet was re-suspended in 3 mL per g of starting cell pellet of KP7/100 buffer and dialysed twice against a 20-fold excess of this buffer to remove the ammonium sulphate before purification.

2.4.3. Ion exchange and affinity chromatography

Purification of all proteins was carried out by a combination of affinity (HiTrap heparin HP column), anion exchange (HiTrap Q HP Sepharose column) and cation exchange (HiTrap SP HP Sepharose column) chromatography. Pre-packed columns (Cytivia), with a column volume (CV) ranging from 5 mL – 20 mL, and the AKTA Prime Plus and AKTA pure systems were used for purification. Purification using Heparin and SP Sepharose columns was carried out using a potassium phosphate buffer, KP7 or KP8, whilst the Q Sepharose column was performed with buffer Tris8 (Table 2-7). Protein was always dialyzed into an appropriate buffer containing 50

- 100 mM NaCl to prevent precipitation. This did not affect the ability of the protein of interest to stick to the column.

Before purification, columns were washed with the appropriate buffer containing 1 M NaCl for 5 CV to remove any previously bound protein, followed by a 5 X CV wash with no salt buffer to equilibrate the column. Protein samples were centrifuged and filtered to remove any precipitation and loaded onto the column with a flowrate <0.5 CV/ min. After loading, a low salt wash (in appropriate buffer containing 0 - 100 mM NaCl) removed any unbound and weakly bound proteins. Proteins were eluted using an appropriate NaCl gradient, typically 100 - 800 mM and 2 - 10 mL fractions were collected. To locate protein, an SDS-PAGE gel was run of the peak fractions, with highest absorbance at 280 nm. The purest factions were pooled and dialyzed into >20-fold excess of appropriate buffer for the next column.

2.4.4 Size exclusion chromatography

Size exclusion chromatography (SEC) was performed using a 120 mL 16/600 Superdex 200 manually packed column. The column was equilibrated overnight with 2 CV of SEC buffer (Table 2-7). Approximately 10 – 20 mg protein was loaded per run, loaded in a volume of 500 – 1000 μL. The column was run at a flow rate of 0.5 mL per minute and 1 mL fractions were collected, whilst monitoring absorbance at 280 nm. Peak fractions were analysed by SDS-PAGE and the purest were pooled.

2.4.5. Protein quantification and storage

Protein concentration was calculated using the extinction co-efficient and confirmed using the Bradford assay. Purified proteins were concentrated using a VIVASPIN centrifugal concentrator with a 10,000 Daltons (Da) molecular weight cut off (MWCO) to about 500 μL. For storage at -20°C, glycerol was added to give a 50% (v/v) final concentration and the protein solution maintained at 1 mM DTT.

2.5. Exonuclease UV activity assay

This spectrophotometric activity assay was used to measure the release of acid-soluble nucleotides from high-molecular-weight DNA and was first described by (Fraser, 1980) and

later modified by (Sayers & Eckstein, 1990). A typical reaction mixture (800 μL) contained 667 μg.mL^{-1} herring sperm type XIV DNA (dissolved in 25 mM potassium glycinate pH 9.3), 25 mM potassium glycinate pH 9.3, 100 mM KCl, 10 mM $MgCl_2$ and 1 mM DTT. Reaction conditions were varied for salt concentration and pH, with buffers used selected based on their appropriate pH ranges.

The reaction was performed at 37^{o}C and the reaction mixture was pre-warmed for 10 minutes prior to the assay. FENs were diluted in the reaction buffer so that the total volume added was 4 μL. At various time points, 100 μL aliquots were removed and added to 100 μL of 6% (v/v) perchloric acid on ice, to terminate the reaction. Samples were incubated on ice for at least 10 minutes and centrifuged in a microfuge at 16,800 x *g* for 5 minutes. A 150 μL sample of the supernatant was diluted with 850 μL of H_2O in a quartz cuvette with a 1 cm path length and the absorbance, relative to a no-enzyme negative control, was measured at 260 nm (A_{260}), where A_{260}= 1.2 corresponds to 100 nmole of released nucleotides. One unit (U) is defined as the amount of enzyme required to release 1 nmole of nucleotides in 30 minutes at 37°C and the specific activity (U.μg^{-1}) was determined graphically by fitting the initial rate section of the reaction with a linear regression in GraphPad Prism 8.

2.6. Oligonucleotides for DNA substrates

Oligonucleotides were custom made by Eurofins and were supplied lyophilised. Oligonucleotide A3 was alternatively supplied by ADTBio Ltd, UK. Oligonucleotides were dissolved at 100 μM in TE buffer (10 mM Tris pH 8, 0.1 mM EDTA) and stored at -20°C. They were annealed to generate the various substrates studied in this project. The oligonucleotide sequences are shown in Table 2-8. The spectral properties of dye labels used are shown in Table 2-9.

2.7. FRET assay

The nuclease activity of the purified proteins was monitored in real-time through a Förster resonance energy transfer (FRET) assay using a variety of dual-labelled FEN substrates: single-flap, double-flap, nicked, overhang and dsDNA. The experiment was carried out with the Hitachi F-2500 FL spectrophotometer.

Table 2-8: Oligonucleotide sequences of substrates used

Name	Sequence (5'-3')	Assay(s)
A3	Cy3-TTTTCGCTGTCTCGCTGAGT-FAM	FRET
A3_UL	TTTTCGCTGTCTCGCTGAGT	EMSA
Invader	ACTCAGCGAGACAGCGCCGGAACACACGCTGCGTGTGTTCCGGT	FRET, EMSA
Flap	ACTCAGCGAGACAGCGCCGGAACACACGCTGCGTGTGTTCCGG	FRET, EMSA
Nick_1	ATACGCATACCTGTGATATCTGGCTAAAAGCACACGCACGGAGAC	FRET, EMSA
Nick_2	FITC-TAGCCAGATATCACAGGTATGCGTAT	FRET
Nick_2_UL	TAGCCAGATATCACAGGTATGCGTAT	EMSA
Nick_3	GTCTCCGTGCGTGTGCT-DAB	FRET
Nick_3_UL	GTCTCCGTGCGTGTGCT	EMSA
T_12	FAM-TTTTTTTTTTTT-BHQ	FRET
G16_A12	GGGGGGGGGGGGGGGAAAAAAAAAAAA	EMSA
OHP	Cy3-CTCTGTCGAACACACGC-FLU-T-GCGTGTGTTC	FRET
OHP1	Cy3-CTCTGTCCGAACACACGC-FLU-T-GCGTGTGTTCG	FRET
OHP2	Cy3- CTCTGTCACGAACACACGC-FLU-T-GCGTGTGTTCGT	FRET
OHP4	Cy3- CTCTGTCTGACGAACACACGC-FLU-T-GCGTGTGTTCGTCA	FRET
OHP4_UL	CTCTGTCTGACGAACACACGC-FLU-T-GCGTGTGTTCGTCA	EMSA
OHP10	Cy3-CTCTGTCCTGCACTGACGAACACACGC-FLU-T-GCGTGTGTTCGTCAGTGCAG	FRET
OHP_S	Cy3-TGTCGAACACACGC-FLU-T-GCGTGTGTTC	FRET
3OV6	GATCTATATGCCATCGG	Structural
5OV4	AAAAGCGTACGC	Structural
JT1	ACCGTCC	Structural
JT2	TTGAGGCAGAGT	Structural
JT2(+2)	AATTGAGGCAGAGT	Structural
JT2(+3)	AAATTGAGGCAGAGT	Structural
JT2(+4)	AAAATTGAGGCAGAGT	Structural
JT2(+5)	AAAAATTGAGGCAGAGT	Structural
JT3	ACTCTGCCTCAAGACGGT	Structural

BHQ, black hole quencher-1; Cy3, cyanine 3; DAB, DABCYL; FAM, 6-Carboxyfluorescein; FITC, fluorescein isothiocyanate; FLU, fluorescin.

2.7.1. Substrate preparation

Substrates were prepared from oligonucleotides (Table 2-8) in FRET base buffer (25 mM HEPES pH 7.4, 100 mM KCl, 0.5 mM EDTA). The single-flap and double-flap substrates were prepared at 4 μM by annealing A3, at a 25% molar excess, with flap (single-flap) or invader (double-flap) oligonucleotides. The nicked substrate was prepared at 4 μM, by annealing equimolar amounts of nick_1, nick_2 and nick_3 oligonucleotides. The dsDNA substrate was

prepared at 10 μM by annealing equimolar amounts of T_12 and G16_A12 oligonucleotides. The overhang substrate was prepared at 1 μM by self-annealing a labelled OHP oligonucleotide to generate a hairpin structure. All substrates were annealed by heating to 95°C for 5 minutes and allowing to cool slowly to ambient room temperature, except the overhang substrate, which was immediately placed at -20°C after heating. Diagrams of the structure of the annealed substrates are shown in Figure 2-1.

2.7.2. Reaction conditions

A stock of 5 x FRET base buffer was made with 125 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES) pH 7.4, 500 mM KCl and 2.5 mM EDTA. This buffer was used to freshly prepare the FRET reaction buffer. Although reaction conditions were varied for salt concentration and buffers, a general reaction buffer was composed of 25 mM HEPES pH 7.4, 100 mM KCl, 0.5 mM EDTA, 10 mM $MgCl_2$, 2 mM DTT and 0.1 mg mL^{-1} acetylated bovine serum albumin (BSA, produced in house by the Sayers' laboratory).

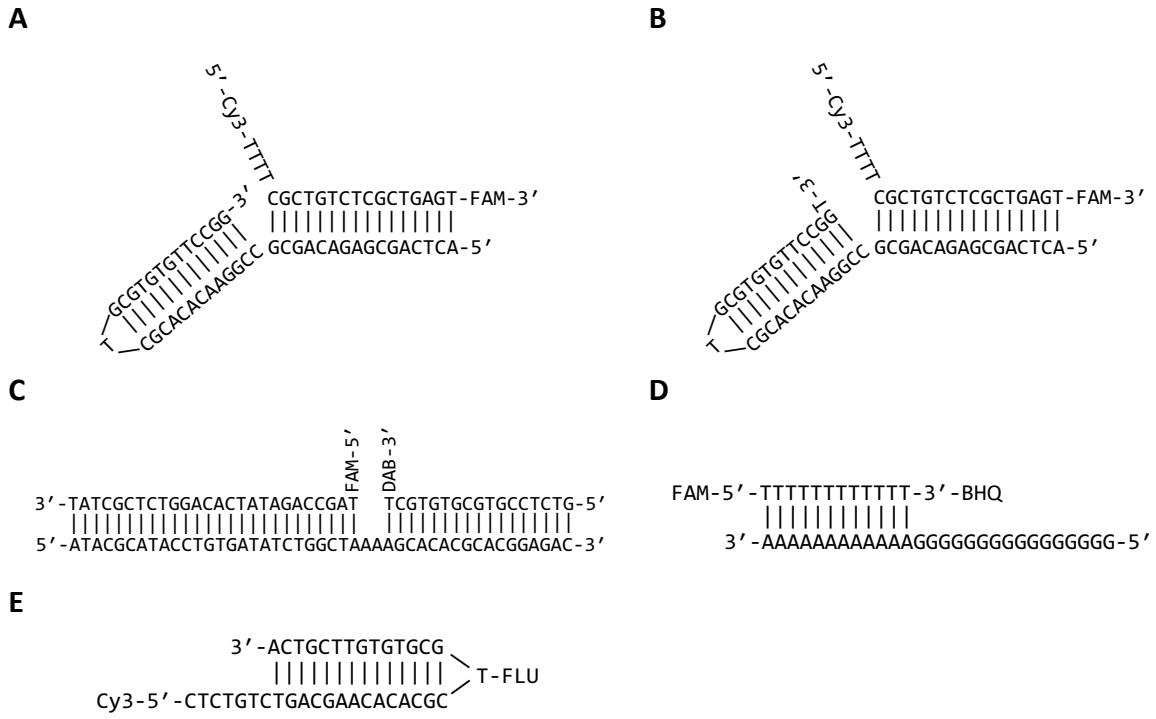


Figure 2-1: Annealed FRET substrates.

(A) Single-flap, (B) double-flap, (C) nicked, (D) dsDNA and (E) Overhang, with the structure of OHP4 as the example shown here.

Table 2-9: Dyes used in FRET assay

Dye label	Spectral profile (nm)
Cy3	Excitation: 500 Emission: 568
FLU	Excitation: 494 Emission: 522
FAM	Excitation: 495 Emission: 520
BHQ1	Absorption: 480 - 580
DAB	Absorption: 400 - 550

2.7.3. Cuvette assay

Enzymes and substrates were diluted to the required concentration using serial dilutions in FRET reaction buffer. For each reaction, 80 μL reaction buffer was mixed with 40 μL of enzyme in an Eppendorf tube and pre-warmed to 37°C in a water bath for 10 minutes. The reaction was initiated by addition of 40 μL substrate (also pre-warmed to 37°C), vortexed and 150 μL of this was immediately placed into a 4- walled glass cuvette. The cuvette was placed into the Hitachi F-2500 FL spectrophotometer and data was collected for 495 nm excitation / 520 nm emission, except the overhang substrate, which was 494 nm excitation / 522 nm emission. Therefore, the FAM or FLU is excited and the emission from the same fluorophore is monitored. Data was collected every 2 seconds for up to 400 seconds or until a visible endpoint had been reached. A no-enzyme negative control was also run to measure the effect of photobleaching.

2.7.4. Analysis of results

To account for photobleaching, the negative control was subtracted and the results were normalized to give an overall change in fluorescence over time. Initial rates (V_0) were calculated from the linear initial rate period and plotted against substrate concentration [S]. These results were fitted to the Michaelis-Menten non-linear regression in GraphPad Prism 8. This regression analysis uses the following model to calculate the Michaelis-Menten constant, K_M, and the maximum velocity, V_{max}:

$$V_0 = \frac{V_{max}[S]}{K_M + [S]} \quad \textbf{(1)}$$

To calculate the catalytic constant, $k_{cat,}$ the data was also fitted with a regression analysis using a variation of the Michaelis-Menten equation that takes in to account the concentration of enzyme sites, [E]. The following model was used:

$$V_0 = \frac{k_{cat}[S][E]}{K_M + [S]} \quad \textbf{(2)}$$

The data was also transformed and fitted to the Hanes-Woolf plot, to enable visualisation in a linear plot, using the following model:

$$\frac{[\mathrm{S}]}{V_0} = \frac{1}{V_{max}}[\mathrm{S}] + \frac{K_m}{V_{max}} \quad \textbf{(3)}$$

2.8. Electrophoretic mobility shift assay

An electrophoretic mobility shift assay (EMSA) was used to determine equilibrium dissociation constant (K_D) of FEN: DNA interactions using radio-labelled substrates.

2.8.1. Substrate preparation

Substrates were prepared from unlabelled oligonucleotides with the same sequences as those used in the FRET assay (Table 2-8). The substrates investigated in this assay were single-flap, double-flap, nicked and overhang (OHP4). The oligonucleotides A3, Nick_1 and OHP4_UL were 5' labelled with ^{32}P ATP prior to annealing. For this, each 150 pmol oligonucleotide was labelled using 1.5 μL [γ -^{32}P] ATP (6000 Ci/ mmol) and 5 units T4 polynucleotide kinase (PNK) (New England Biolabs) in T4 PNK buffer as 5 μL reactions. Reactions were left at room temperature overnight and heat inactivated at 95°C for 5 minutes.

The single-flap (A3 + flap), double-flap (A3 + invader) and nicked (Nick_1 + Nick_2_UL + Nick_3_UL substrates were prepared at 200 nM by annealing equimolar amounts of 5' labelled with unlabelled oligonucleotides. The overhang substrate was prepared at 200 nM by self-annealing labelled OHP4. Oligonucleotides were combined in annealing buffer (10 mM Tris pH 8, 100 mM NaCl, 1 mM EDTA, 12% (v/v) glycerol), heated to 95°C for 5 minutes in a water bath

and slowly cooling to 25°C, whilst stirring. The annealed substrates were purified from a 30 cm 8% native polyacrylamide gel (Tris-Borate-EDTA, TBE (89 mM Tris, 89 mM boric acid, 2 mM EDTA) 9% (v/v) acrylamide/ bisacrylamide (19:1), 0.1% (v/v) ammonium persulfate (APS) and 0.1% (v/v) tetramethylethylenediamine (TEMED)) using the crush and soak technique (Green and Sambrook, 2019). For this, bands excised from the polyacrylamide gel were crushed against the wall of an Eppendorf tube and soaked overnight in 150 μL extraction buffer (10 mM Tris pH 8, 100 mM NaCl, 1 mM EDTA) whilst shaking at 4°C. The extracted substrates were separated from the acrylamide in a 0.2 μm centrifugal filter. Purified substrates were quantified by comparing the intensity of annealed substrate with that of a known concentration of labelled unannealed oligonucleotide by autoradiography.

2.8.2. Assay and results analysis

Reactions of 20 μL were carried out in the presence of 500 pM DNA substrate in EMSA reaction buffer (25 mM potassium glycinate pH 9.3, 100 mM KCl, 5% (v/v) glycerol, 5 mM $CaCl_2$, 1 mM DTT, 0.5 mM EDTA, 1 $mg.mL^{-1}$ acetylated BSA). Reactions were incubated on ice for 1 hour to ensure equilibrium had been reached before loading onto a 10 cm 17% native polyacrylamide gel (50 mM Tris-Bicine pH 8.3 9% (v/v) acrylamide/ bisacrylamide (19:1), 1 mM EDTA, 1 mM DTT, 0.1% (v/v) APS and 0.1% (v/v) TEMED) with the running buffer 50 mM Tris-Bicine pH 8.3, 1 mM EDTA, 1 mM DTT at 4°C. The gel had been pre-run at 200 V for 60 minutes before loading samples and samples were loaded at 250 V whilst the running the gel, to ensure that all samples took the same amount of time to enter the gel. The gel was run for 80 minutes at 200 V until the bromophenol blue marker was 0.5 cm away from the end. Gels were dried for 30 minutes in a vacuum gel dryer at 80°C and exposed to phosphor imager plates (Fujifilm) for imaging and quantification (Fuji FLA3000, Image Guage V3.3 software). The fraction bound for each enzyme concentration was calculated using the formula, fraction bound= (intensity of bound band)/(bound + unbound). Results were plotted on GraphPad Prism and fitted to the 'one site-specific binding model.'

2.9. Electrophoresis methods

2.9.1. DNA agarose gel

Separation of DNA by size was achieved using a 1% agarose gel. Agarose powder was boiled with Tris-acetate-EDTA (TAE) buffer. The solution was allowed to cool for a few minutes and 0.05 $\mu L.mL^{-1}$ Midori Green advance stain (Nippon genetics) was added. The gel was cast into a prepared plate, comb inserted and allowed to cool to room temperature. Samples were mixed with 6X loading dye (New England Biolabs) and loaded into wells of the set gel along with an appropriate ladder (New England Biolabs). The gel was run in an electrophoresis tank with TAE buffer and 7 V/ cm was applied until the dye front had reached ~3/4 of the way down the gel and the gel was visualised using a BIO-RAD EZ gel documentation system on the SYBR green setting.

2.9.2. SDS-PAGE gel

Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) was used to separate and visualize proteins according to their molecular weight. A 10% gel was made by casting a resolving gel (100 mM Tris-Bicine pH 8, 10% (v/v) acrylamide/ bisacrylamide, 37.5:1, 0.1% (w/v) SDS, 0.05% (v/v) APS and 0.0016% (v/v) TEMED) topped with stacking gel (125 mM Tris-HCl pH 6.9, 5% (v/v) acrylamide/ bisacrylamide, (37.5:1), 0.01% (w/v) SDS, 0.066% (v/v) APS, 0.033% (v/v) TEMED). Samples were mixed 1:1 with 2X loading dye (2% (w/v) SDS, 20% (v/v) glycerol, 125 mM Tris-HCl, 0.05% (w/v) bromophenol blue, 2 mM DTT) and heated at 95^{o}C for 5 minutes. Samples were then loaded into the gel and electrophoresis was performed at 35 mA per gel in SDS running buffer (100 mM Tris, 100 mM bicine, 2 mM EDTA and 1% (w/v) SDS) for approximately 33 minutes, until the dye front was about 0.5 cm from the end of the gel. The gel was stained with Coomassie stain (2 $mg.mL^{-1}$ Coomassie Brilliant Blue dye, 40% (v/v) methanol and 10% (v/v) acetic acid) for about 30 minutes and de-stained in de-staining solution (40% (v/v) methanol, 10% (v/v) acetic acid) for several hours until bands became clear.

2.9.3. DNA-substrate PAGE gel

DNA-substrate PAGE gels, referred to as a zymogram, were cast in the same way as SDS-PAGE gels except 300 μg type XIV DNA (Sigma, in 25 mM Tris-HCl, pH 8) was incorporated into the resolving gel before casting. Samples were also prepared and run the same as for SDS-PAGE and after running, the gel was washed (3 X 20 minutes) with Tris-Bicine-glycerol (TBG) buffer (100 mM Tris, 100 mM Bicine, 10% (v/v) glycerol), to renature proteins separated in the gel. The gel was then incubated in a reaction buffer (100 mM Tris, 100 mM bicine, 10% (v/v) glycerol, 100 mM KCl, 50 mM NaCl, 10 mM $MgCl_2$, and 1 mM DTT) for 15 minutes, allowing nucleases to interact with and cleave DNA within the gel. The gel was then washed with TBG buffer, stained for 10 minutes with 0.5 $\mu L.mL^{-1}$ Midori Green Advance stain and imaged using the BIO-RAD EZ gel documentation system under the SYBR green setting. Nuclease activity was visualised by the presence of a dark band.

2.10. Mass spectrometry analysis

Proteins were analysed with Matrix-assisted Laser Desorption/Ionisation time-of-flight (MALDI-TOF) by the University of Sheffield Mass Spectrometry service (Department of Chemistry, UoS) to obtain the molecular weight and relative abundance of any proteins in the sample.

2.11. X-ray crystallography

2.11.1. Preparation of protein for crystallization

Protein used for crystallization had been purified using size exclusion chromatography as the final purification stage. Protein was exchanged into crystallization buffer (25 mM Tris-HCl pH 8, 50 mM KCl, 2 mM DTT) and concentrated to at least 30 $mg.mL^{-1}$ using a VIVASPIN centrifugal concentrator with a 10,000 Da MWCO. Protein was either used immediately for crystallization or small aliquots were flash frozen in liquid nitrogen and stored at -80°C.

2.11.2. Substrate preparation for co-crystallization with DNA

Various DNA substrates were used for co-crystallization and were prepared from oligonucleotides summarised in Table 2-8. Oligonucleotides were dissolved to 12 μM in

annealing buffer (10 mM Tris-HCl pH 8 and 50 mM KCl). The structures of the annealed substrates are shown in Figure 2-2. The 5OV4 (5′ overhang) and 3OV6 (3′ overhang pseudo product) were self-annealed at 6 μM to form a palindromic substrate. The JT2+n (double-flap) substrate was annealed at 4 μM by combining JT1, JT2+n and JT3 oligonucleotides at equimolar amounts, where n is the length of the 5′ flap. Oligonucleotides were annealed by heating to 95°C for 5 minutes and allowing to cool gradually to room temperature. Prepared substrates were stored at 4°C for up to 1 week.

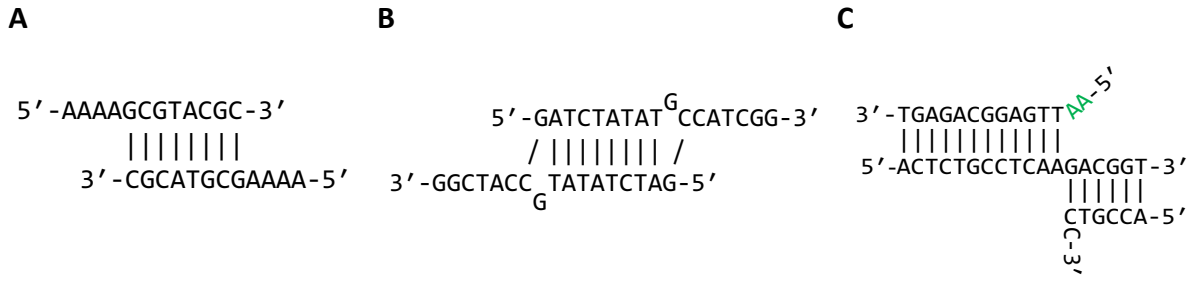


Figure 2-2: Structures of cystallography substrates

(A) 5OV4, (B) 3OV6 and (C) JT2+2 as an example of the double-flap substrate. The length of the 5′ flap determined by the specific JT2 oligonucleotide used

2.11.3. Setting crystal trials

Protein was used to set crystal trials at a concentration of 5 – 20 mg.mL^{-1} after any other additives or DNA was added. Typically, 50 mM $MgCl_2$ or 50 mM $CaCl_2$ was included. For co-crystallization with DNA, the annealed substrate was added to the protein at a 1:1 molar ratio and was incubated on ice for 10 minutes to allow complex formation. Crystal trials were set up using commercial screens (JCSG+, MPD, Morpheus, Natrix, PACT, pH clear and Proplex) in 96-well 2-drop plates. In each reservoir of the crystallization plates, 50 μL of screen condition was added and the Mosquito robot was used for dispensing protein and precipitant in the sitting drop technique. Typically, 200 nl of both protein and precipitant was dispensed, giving a 1:1 ratio. The plates were sealed and stored at either 8°C or 18°C. Plates were checked after 2 days and subsequently every week for initial hits of crystal growth or other interesting occurrences which could be optimized. If screening was not very successful (no crystal growth, mostly clear/ precipitated wells) after several weeks of checking plates, screening conditions

were altered, which included protein to precipitant ratios, protein concentration, protein to DNA ratio and varying salt additives or concentrations (KCl, $MgCl_2$, $CaCl_2$).

2.11.4. Optimization of crystal trials

Conditions selected for optimization generally produced tiny crystals which were too small for collecting diffraction data. The aim of optimization was to produce larger and/ or better-quality crystals by slightly altering crystallization conditions. If crystals had formed in conditions that contained phosphates or sulphates they were first screened with Izit dye (Hampton research) to try and distinguish between protein or salt crystals before attempting to optimize. The dye was added directly to the well and after 10 minutes, the wells were checked if the crystals had turned blue. Typically, protein crystals turn blue whilst salts do not, due to tighter packing of salt crystals they are not able to take up the dye. Generally, optimization involved altering the pH, precipitant and any salt of the initial hit conditions. This included KCl, $CaCl_2$ and $MgCl_2$ added to the protein prior to setting trials and ratios of protein: DNA for co-crystallization. Optimization was carried out with the same plates used for screening.

2.11.5. Data collection

Cryogenic solutions were made by the addition of 20% glycerol or 20% DMSO to the crystallisation condition. Crystals were looped using LithoLoops (Molecular Dimensons), dipped in cryogenic solution for a few seconds and then flash frozen and stored in liquid nitrogen. Crystals were shipped to the Diamond Light Source, UK, for remote data collection using various beamlines to collect diffraction data. During data collection, a partial data set with only a few images was first collected to enable the diffraction to be analysed. If this indicated a protein was diffracting to a reasonable resolution, a full data set was collected. Alternatively, some crystals were analysed using automatic data collection. Data was automatically processed using various programs provided within the Diamond data collection software. This software calculated the resolution, space group and other statistical parameters. Based on these calculated parameters, the most suitable dataset was chosen for structure determination.

2.11.6. Structure determination

The protein structure was determined by molecular replacement (MR) with a model of a previously determined FEN structure, using the MOLREP or PHASER programs (McCoy *et al.*, 2007; Vagin and Teplyakov, 2010). The BUCANNER automated protein building pipeline was used after MR to improve the model (Cowtan, 2006). Rounds of manual building and refinement were performed using COOT and Refmac5 programs, respectively (Emsley *et al.*, 2010; Murshudov *et al.*, 2011). Using COOT, any DNA, metal ions, waters and other solvents were added. Metal ions were validated using the CheckMyMetal server (Zheng *et al.*, 2017). Final model validation was performed in COOT and with the Molprobity and PDB validation servers (Williams *et al.*, 2018).

Chapter 3 - Biochemical analysis of WT T7 gp6

3.1. Introduction

The aim of this chapter is to expand on previous work surrounding the characterization of T7 gp6 (see section 1.8; Kerr and Sadowski, 1972; Mitsunobu *et al.*, 2014), a key component of the Binx Health IO device. Detailed analysis of the performance of this enzyme within the device has not been extensively studied (Binx Health, Personal communication, David Pearce). T7 gp6 was overexpressed in *E. coli* and purified using ion-exchange, affinity and size-exclusion chromatographic techniques. The nuclease activity and DNA binding affinity for endonuclease and exonuclease substrates was analysed. The effect of reaction conditions including pH, monovalent salts (KCl, NaCl) and divalent salts ($MgCl_2$, $MnCl_2$, $CoCl_2$, $CaCl_2$, $NiCl_2$, $Zn(CH_3CO_2)_2$ was also investigated, to identify optimum reaction conditions which enhance the activity of the enzyme.

3.2. Results

3.2.1. Over-expression and purification of T7 gp6

3.2.1.1. Optimization of T7 gp6 expression

The T7 gp6 gene was provided in the pJONEX4 plasmid. The gene was sub-cloned into the bacteriophage T5 promoter vector, pTTQ18-T5, using the restriction sites *Eco*RI and *Hin*dIII and the insert was confirmed by Sanger sequencing. Gene expression from this vector is driven by the lactose-inducible bacteriophage T5 promoter which is recognised by *E. coli* RNA polymerase under control of the $LacI^q$ repressor.

Small-scale expression studies in the pTTQ18-T5 system, were attempted in a range of *E. coli* strains (BL21, W3110, XL1blue) with media utilizing vegetable-derived tryptone supplemented with glucose throughout all plating and culturing stages, to help repress leakage of expression (Wanner *et al.*, 1978; Grossman *et al.*, 1998). For each cell line, protein expression at 28°C was investigated in 5 mL cultures and induced using IPTG (final concentration of 1 mM) in LB media or autoinduction (AI), in SuperBroth AIM or 2YT AIM (Formedium Ltd, UK). SDS-PAGE analysis

shows that T7 gp6 was expressed in the BL21 strain (Figure 3-1) but not detectably in W3110 and XL1blue (data not shown). Time-course analysis of protein expression reveals that low levels of T7 gp6 expression are seen 2 hours and 3 hours after induction and relative expression decreases after this. For autoinduction, peak expression is seen after 6 hours with relative expression also decreasing after this. Furthermore, accumulation of an unknown ~50 kDa protein occurs after 24 h.

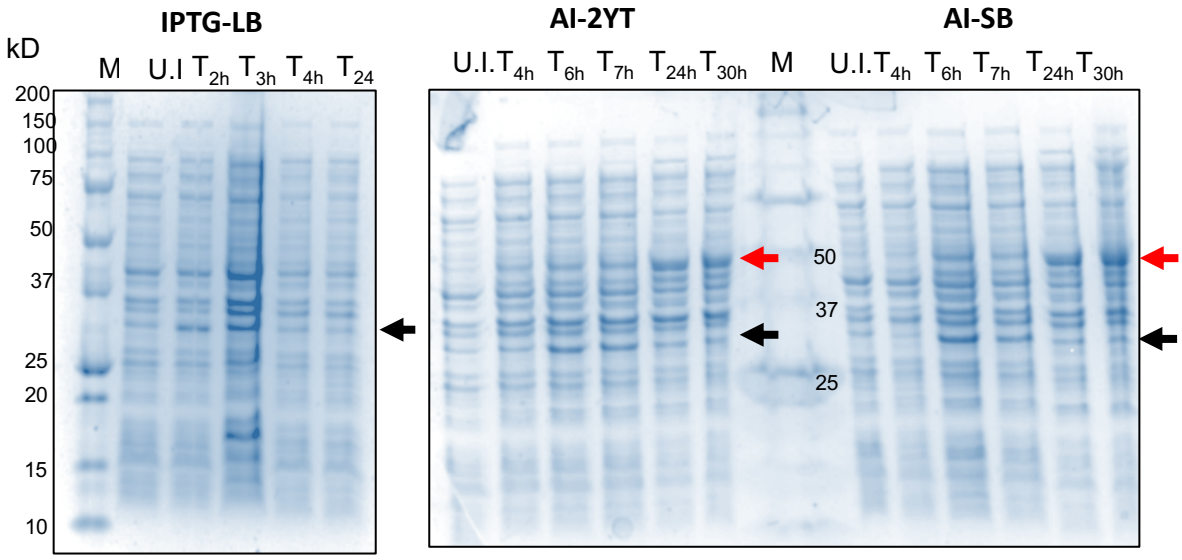


Figure 3-1: Optimization of WT T7 gp6 protein expression from pTTQ18-T5

SDS-PAGE analysis of small-scale WT T7 gp6 protein expression using 1 mM IPTG (IPTG-LB) or autoinduction methods (AI-2YT and Superbroth, AI-SB). Samples were taken before induction (U.I.) and at various timepoints (T_{nh}) after induction (IPTG) or inoculation of autoinduction media. Cultures were grown at 28°C. Samples were run on 10% acrylamide gels with PrecisionPlus marker (lane M). The black arrow indicates the position of target protein expression, whilst red arrow indicates the position of accumulation of an unknown protein.

Although protein expression can clearly be seen in SDS-PAGE analysis, the level of expression is low compared to other FENs studied within the laboratory (Allen *et al.*, 2009; Oates, 2016; Figure 4-4). In an attempt to increase levels of expression, the heat inducible pJONEX4 system was also investigated (Sayers and Eckstein, 1991). This used the *E. coli* strain ΔH1+. Small-scale expression trials were performed in 5 mL 2YT media cultures to assess levels of expression and enable optimization of the expression protocol. As this is a heat-shock induction system driven by λP_L, protein expression was induced when cells were in mid-log growth phase ($A_{600\ nm}$ at around 1) by shifting the temperature to 42°C. In these trials, protein expression was

induced for 2 and 3 hours before an overnight incubation at 20°C, with samples taken at each stage. As seen by SDS-PAGE analysis in Figure 3-2, target protein expression is visible after both 2 and 3 hours of induction, although the relative level of expression decreases following the overnight incubation. Thus, an optimized expression protocol of T7 gp6 from pJONEX4 employed a 2-hour induction, followed by immediate harvesting of cells after induction.

Although the level of expression from pTTQ18-T5 with AI is comparable to pJONEX4, the AI technique requires more monitoring to confirm protein expression has occurred when scaling up, to ensure the optimum relative expression levels are obtained. Furthermore, expression of an unknown protein might also make downstream purification more difficult. Therefore, the pJONEX4 system was chosen for large-scale T7 gp6 expression.

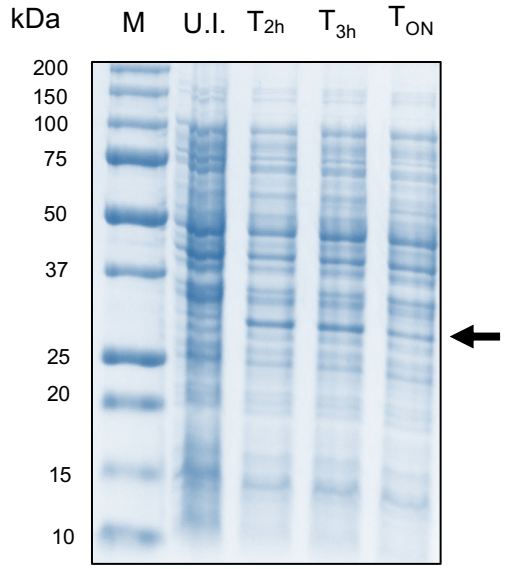


Figure 3-2: Optimization of T7 gp6 expression from pJONEX4

SDS-PAGE analysis of small scale T7 gp6 protein expression. Samples were taken before (lane U.I.), after 2 (lane T_{2h}) or 3 hours (lane T_{3h}) 42°C heat-shock induction and after an overnight incubation at 20°C after 3-hour induction (lane T_{ON}). Samples were run on 10% acrylamide gels with PrecisionPlus marker (lane M). The black arrow indicates the position of target protein expression.

3.2.1.2. Large-scale overexpression and purification of T7 gp6

Expression of T7 gp6 from pJONEX4 was scaled up to a 5 L fermenter using rich 5YT media (Figure 3-3A). After a 2-hour induction at 42°C, cells were harvested immediately to yield a 38 g cell pellet. The entire cell pellet was lysed and ~50 % of the protein remained in the soluble fraction. To inform development of a purification protocol, the amino acid sequence was analysed with Expasy ProtParam, an online software that computes various physical and chemical properties of a protein from the amino acid sequence (Gasteiger *et al.*, 2005). This software computed the theoretical isoelectric point (pI) as 4.93, the extinction coefficient,

$\varepsilon_{0.1\%}$ as 2.258 $(mg/mL)^{-1}cm^{-1}$ and the molecular weight (MW) as 34502.20 Daltons (Da). Soluble T7 gp6 and was purified by affinity (Heparin) and ion-exchange liquid chromatography with Q Sepharose, an anion exchanger and SP Sepharose, a cation exchanger. (Figure 3-3B-D). The Q Sepharose column was run at pH 8 and SP Sepharose column at pH 7. Based on the predicted pI, it was anticipated that at the pH used, the protein would interact with Q but not SP Sepharose. However, the protein interacted with and was eluted from both anion and cation exchange columns. In total from the 38 g cell pellet approximately 50 mg protein was obtained. The final stage of purification was size-exclusion chromatography (SEC) and the protein was eluted with one peak on the chromatogram. SDS-PAGE analysis of the final concentrated protein reveals that purity achieved was >95% (Figure 3-4A). The protein was stored at -20°C in 50% glycerol.

3.2.1.3. Zymogram and mass spectrometry analysis

The final purified T7 gp6 was analysed by zymography (Figure 3-4B). This was used to assess whether the main nuclease activity co-purifies with WT T7 gp6 and also to identify whether any contaminating nuclease activity is present which may interfere with biochemical analysis. The zymogram reveals that even with 16 μg total protein loaded onto the gel, the only visible band of nuclease activity is seen at a molecular weight corresponding to T7 gp6.

Furthermore, T7 gp6 was subject to mass spectrometry analysis *via* MALDI-TOF along with a commercially sourced protein (provided by industrial partner) (Figure 3-5). The molecular weight of the main peak on the spectra of both of these varies by less than 0.5 Da, where the laboratory protein is 34371.54 Da and the commercially sourced protein is 34371.21 Da. This molecular weight corresponds to the post-translational cleavage of the initiating *N*-formylmethionine.

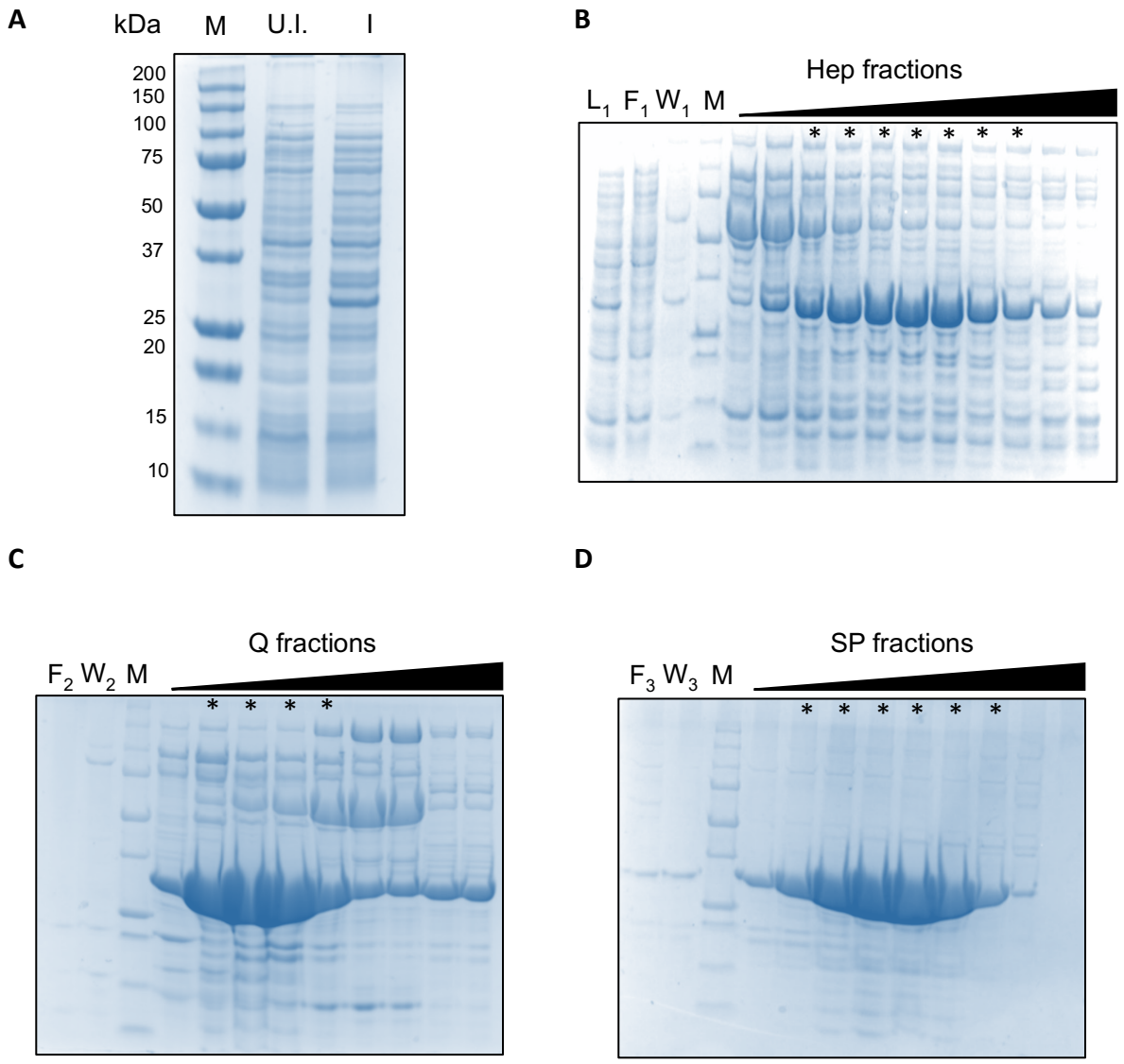


Figure 3-3: Overexpression and purification of WT T7 gp6

SDS-PAGE analysis of WT T7 gp6 overexpression and purification. (A) WT T7 gp6 overexpression before (lane U.I.) and after (lane I) 2 hours heat shock induction. (B) Heparin HP column (20 mL) chromatography in KP8 buffer. Lane L_1, total protein loaded on to the column; lane F_1, flow-through; lane W_1, wash with buffer containing no NaCl; Lanes Hep fractions, 2 mL peak fractions collected over a 150 mL gradient ranging from 0 M - 1 M NaCl. The purest fractions are were pooled and dialysed for further purification. (C) Q HP column (10 mL) chromatography in Tris8 buffer. Lane F_2, flow through; lane W_2, wash with buffer containing 100 mM NaCl; lanes Q fractions, 2 mL peak fractions collected over a 60 mL gradient ranging from 0.1 M – 1 M NaCl. The purest fractions were pooled and dialysed for further purification. (D) SP Sepharose HP column (5 mL) chromatography in KP7 buffer. Lane F_3, flow through; lane W_3, wash with buffer containing no NaCl; lanes SP fractions, 2 mL peak fractions collected over a 60 mL gradient ranging from 0 M – 1 M NaCl. The fractions pooled at each stage are indicated (*). Analysis was performed on 10% acrylamide gels with PrecisionPlus marker (BioRad) (lane M) used.

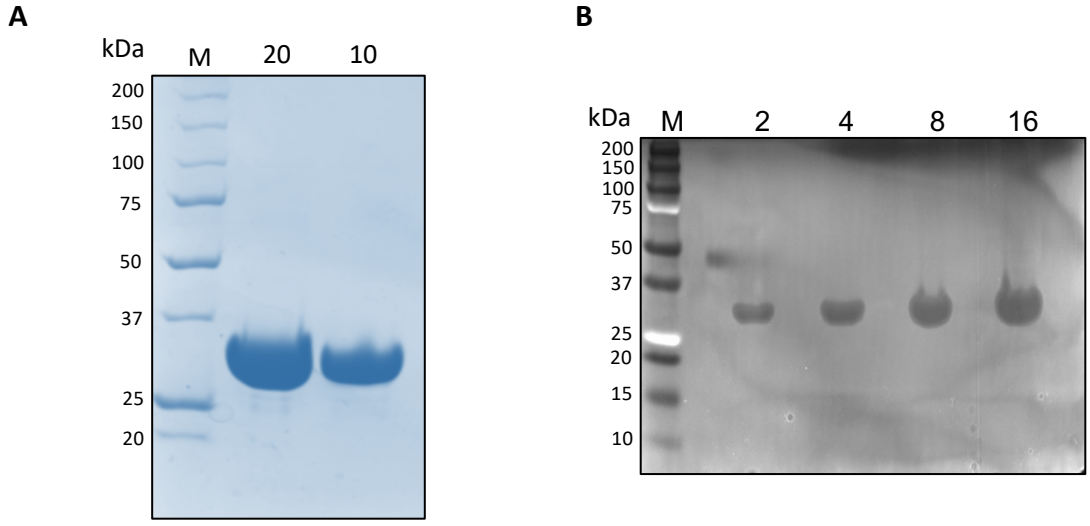


Figure 3-4: SDS-PAGE and zymogram analysis of T7 gp6

Purified T7 gp6 was analysed by (A) SDS-PAGE and (B) zymogram. The zymogram gel was incubated in reaction buffer for 15 minutes. Lane numbers denote amount of protein loaded in μg. Analysis was performed on 10% acrylamide gels with PrecisionPlus marker (BioRad) (lane M) used.

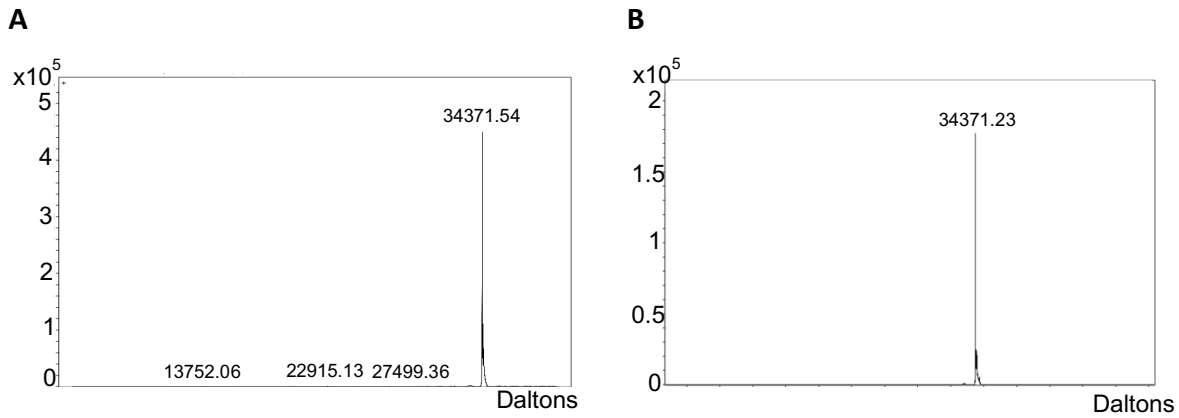


Figure 3-5: Mass spectrometry analysis of T7 gp6

Purified WT T7 gp6 was analysed by MALDI-TOF mass spectrometry. (A) Laboratory purified protein. (B) commercially sourced protein.

3.2.2. Analysis of T7 gp6 nuclease activity

3.2.2.1. Activity of T7 gp6 on plasmid DNA

T7 gp6 was tested on plasmid DNA in the presence of Mg^{2+} as the co-factor. The plasmid, specifically pTTQ18-T5.TaqPol, was prepared using the Monarch Miniprep kit (New England Biolabs) and contains different conformations of the plasmid DNA: covalently closed circular

(ccc), nicked and linear. This is visualised as three bands on an agarose gel (Figure 3-6, Lane C). T7 gp6 is able to remove nicked and linear DNA within 20 minutes, however the cccDNA appears resistant to degradation up to 200 minutes. This indicates that T7 gp6 is free of double stranded endonuclease activity under the conditions tested.

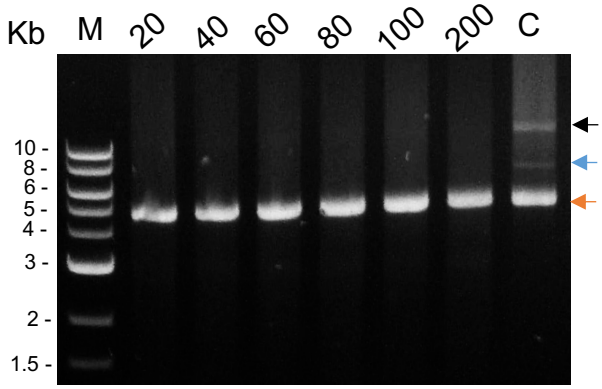


Figure 3-6: Activity of T7 gp6 on plasmid DNA

T7 gp6 (1 μg) was incubated with plasmid DNA (2 μg, specifically pTTQ18-T5.TaqPol) at 37°C in NEB EcoRI buffer (100 mM Tris pH 7.5, 50 mM NaCl, 10 mM $MgCl_2$, 0.025% Triton X-100). Aliquots of 10 μL were removed at time points indicated by lane numbers (minutes) and analysed by a 1% agarose gel stained with Midori Green Advance DNA stain. Lane M, NEB 1 kb DNA ladder; lane C, no enzyme control. The positions of nicked, linear and covalently closed circular plasmid DNA are shown by the black, blue and orange arrow, respectively.

3.2.2.2. UV assay analysis of T7 gp6 exonuclease activity

The exonuclease activity of T7 gp6 was assessed using a UV spectrophotometric assay and compared to a commercially sourced T7 gp6 and also the well-characterized T5FEN. This assay, first described by Fraser (1980) and later modified by Sayers and Eckstein (1990), measures the release of acid soluble nucleotides (A_{260} nm) from a high-molecular-weight DNA. From nucleotide release, the specific activity ($U.\mu g^{-1}$) of the enzyme was calculated assuming that an A_{260} nm of 1.2 corresponded to 100 nmol of acid-soluble nucleotides. In this assay, one unit (U) was defined as the amount of enzyme required to release 1 nmol of nucleotides in 30 minutes at 37°C.

The time course of the reaction shows that the release of acid-soluble nucleotides over time between the laboratory and commercially sourced T7 gp6 are indistinguishable within the errors of the experiment and the reaction reaches a clear endpoint at around 40 nmol (Figure 3-7A). In contrast, an identical amount of T5FEN results in a greater release of acid-soluble

nucleotides with no clear endpoint reached after 60 minutes. However, the initial rate period of both T7 gp6 and T5FEN are indistinguishable, resulting in no statistical differences between the specific activity (Figure 3-7B). Furthermore, addition of T7 gp6 to a T7 gp6 reaction after 30 minutes, at the initial plateau, does not result in additional nucleotide release whilst addition of T5FEN does (data not shown).

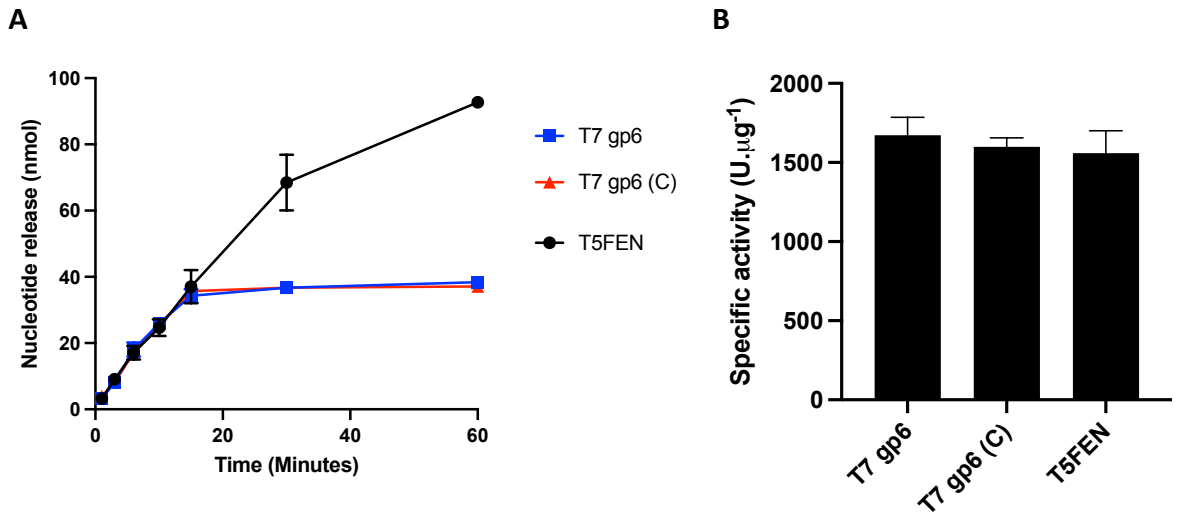


Figure 3-7: Comparison of exonuclease activity between laboratory and commercially sourced WT T7 gp6

The UV spectrophotometric assay was used to determine the exonuclease activity of laboratory purified T7 gp6, commercially sourced T7 gp6 (C) and T5FEN. The release of acid-soluble nucleotides from DNA was measured after the addition of 0.5 μg of protein. (A) Time course of reaction. (B) Specific activity of proteins. The reaction mixture contained 667 $\mu g.mL^{-1}$ type XIV DNA, 50 mM potassium glycinate pH 9.3, 100 mM KCl, 10 mM $MgCl_2$, 1 mM DTT. The mean of three repeats is shown and error bars represent SEM. The rate of release was determined graphically by fitting the initial rate section with a linear regression in GraphPad Prism. Statistical significance was determined by a one-way ANOVA with Tukey's multiple comparisons, where no statistical difference (ns, $p>0.05$) was observed between any group.

3.2.2.3. Determination of T7 gp6 catalytic parameters with the FRET assay

To determine catalytic parameters (K_M, k_{cat}) of T7 gp6, the cleavage of various DNA substrates was investigated in a FRET assay. These substrates are dual-labelled with fluorophores and designed to test either the endonuclease (single-flap, Figure 3-8A; double-flap Figure 3-9B; overhang, Figure 3-10) or exonuclease (nicked, Figure 3-11A; dsDNA, Figure 3-12A) activity of FEN enzymes. Due to the overlapping spectra and close proximity of the two fluorophores in an intact substrate, one is a donor whilst the other acts as a quencher. As T7 gp6 cuts the DNA

at the cleavage site, the fluorophores become separated from each other and the change in energy released from the donor is measured as a change in fluorescence. The wavelength and fluorophore measured in each substrate is shown in Table 2-9.

Progress curves at various substrate concentrations were obtained (Panel B, Figure 3-8 to Figure 3-12). The reactions were carried out at pH 7.4 (HEPES-NaOH) with 10 mM $MgCl_2$ providing the divalent metal ion co-factor. The concentration of T7 gp6 used, as indicated in the figure legends, varied for each substrate and this was optimized to give reaction rates which were not so fast that they could not be accurately measured. A surprising result was observed for the single-flap and double-flap substrates in which fluorescence decreased over time, rather than increase, indicating the fluorophores become closer together. Suggested explanations for this are discussed in section 3.3.4. However, as the progress curves still exhibited a concentration-dependent rate change and a clear endpoint, they were still used for further kinetics analysis. The concentrations of released products were calculated from fluorescent emission intensity. The initial rate of reaction was plotted against substrate concentration and fitted by non-linear regression of the Michaelis-Menten equation in GraphPad prism to give K_M and k_{cat} (Panel C, Figure 3-8 to Figure 3-12; Table 3-1). Data was also transformed using the Hanes-Woolf equation to enable linear visualisation (Panel D, Figure 3-8 to Figure 3-12). The catalytic parameters determined from all substrates tested are summarised in Table 3-1.

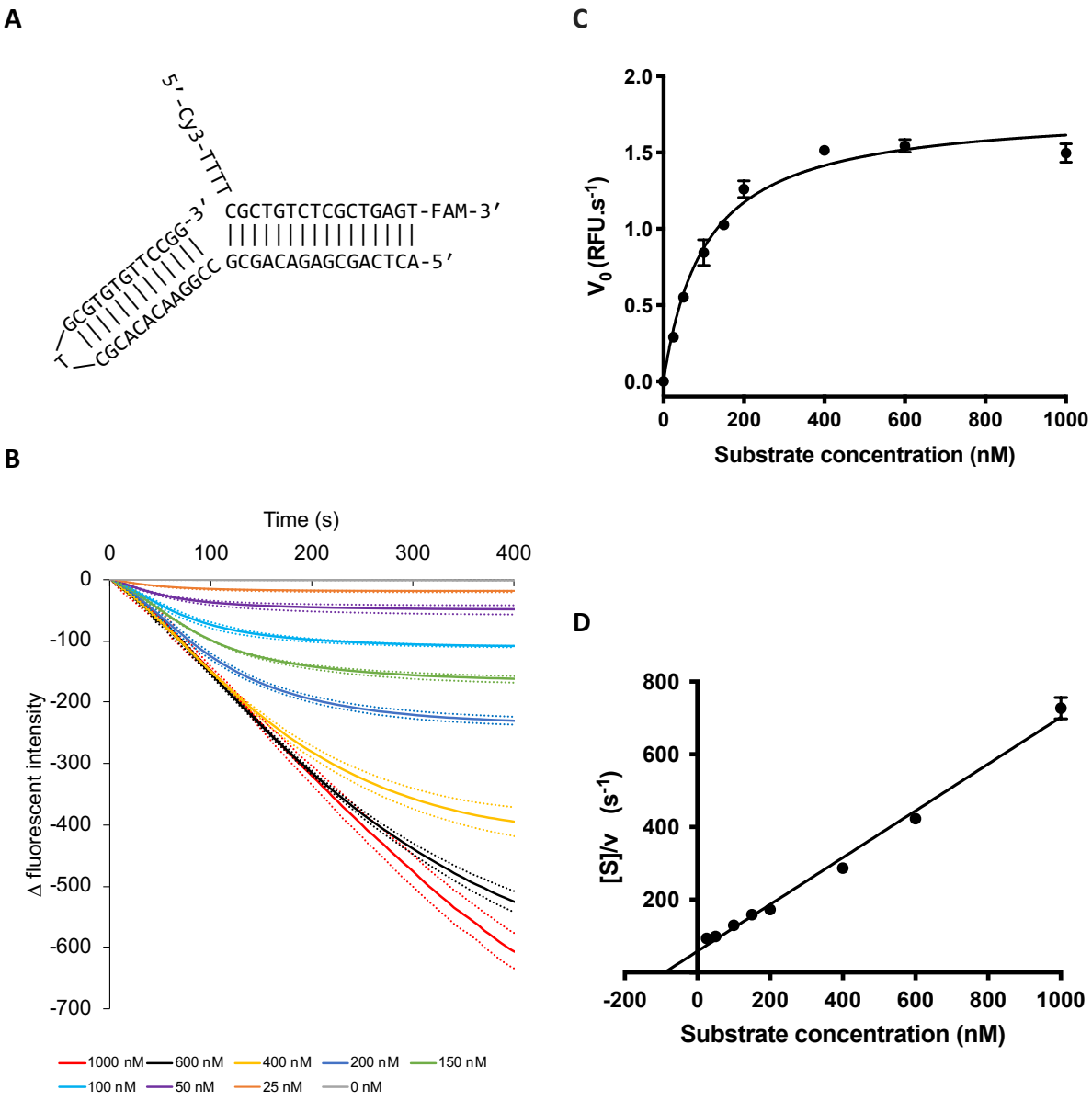


Figure 3-8: Michaelis-Menten analysis of T7 gp6 with single-flap substrate

Catalytic parameters were determined using the FRET assay. (A) Structure of single-flap substrate used; Cy3, cyanine 3; FAM, 6-carboxyfluorescein. (B) Reaction progress curves monitored by change in the fluorescence over time at various substrate concentrations by addition of 14.5 nM T7 gp6. Progress curves at each substrate concentration were normalised against a no-enzyme control. (C) The initial velocity (V_0) was calculated from the initial rate portion of the progress curves and plotted against substrate concentration. This was analysed by non-linear regression of the Michaelis-Menten equation to determine catalytic parameters. (D) Data was transformed using the Hanes-Woolf equation for linear visualisation. The mean of three repeats is shown and dashed lines represent SD and error bars represent SEM.

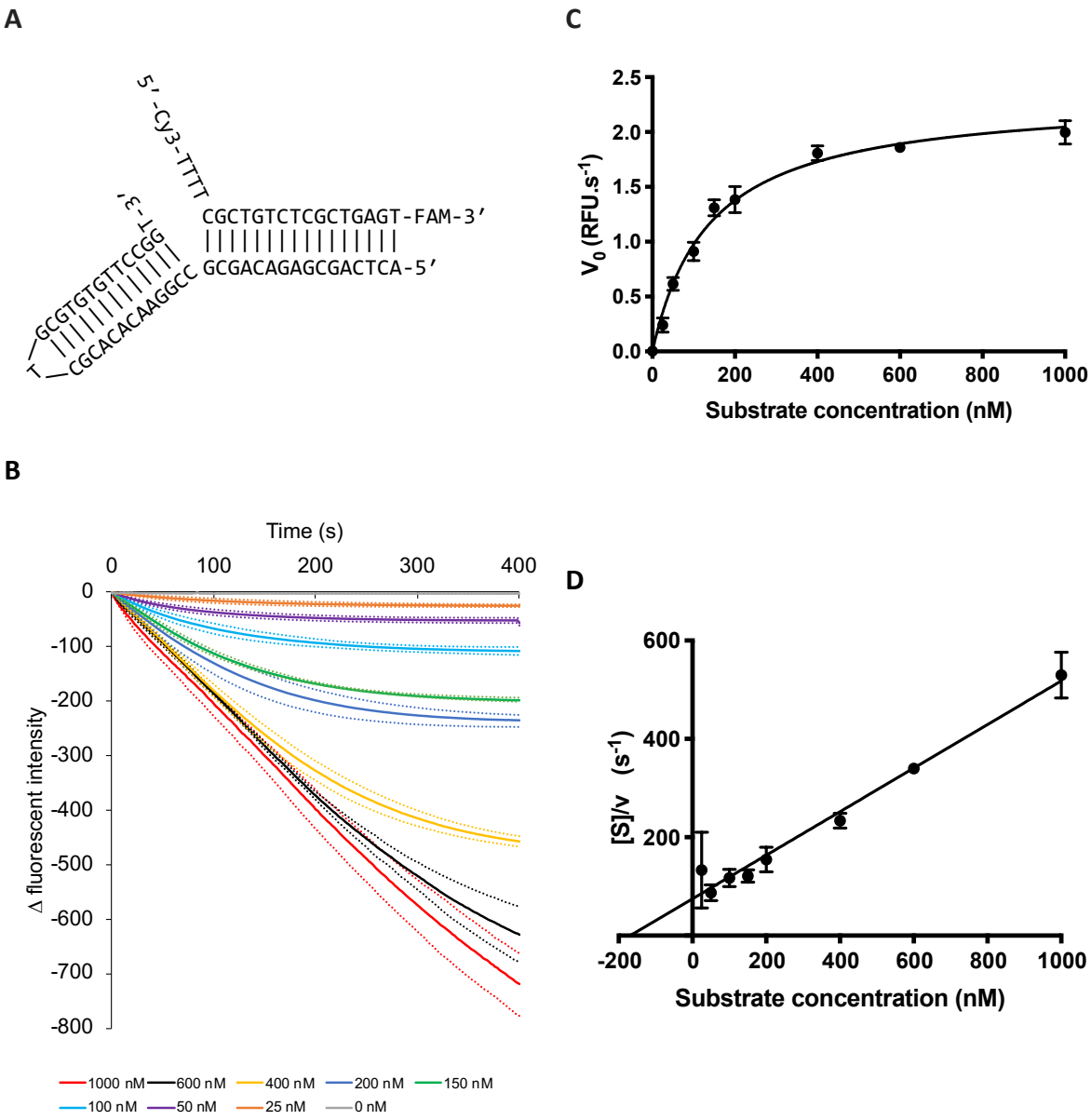


Figure 3-9: Michaelis-Menten analysis of T7 gp6 with double-flap substrate

Catalytic parameters were determined using the FRET assay. (A) Structure of double-flap substrate used; Cy3, cyanine 3; FAM, 6-carboxyfluorescein. (B) Reaction progress curves monitored by change in the fluorescence over time at various substrate concentrations by addition of 29.0 nM T7 gp6. Progress curves at each substrate concentration were normalised against a no-enzyme control (C) The initial velocity (V_0) was calculated from the initial rate portion of the progress curves and plotted against substrate concentration. This was analysed by non-linear regression of the Michaelis-Menten equation to determine catalytic parameters. (D) Data was transformed using the Hanes-Woolf equation for linear visualisation. The mean of three repeats is shown and dashed lines represent SD and error bars represent SEM.

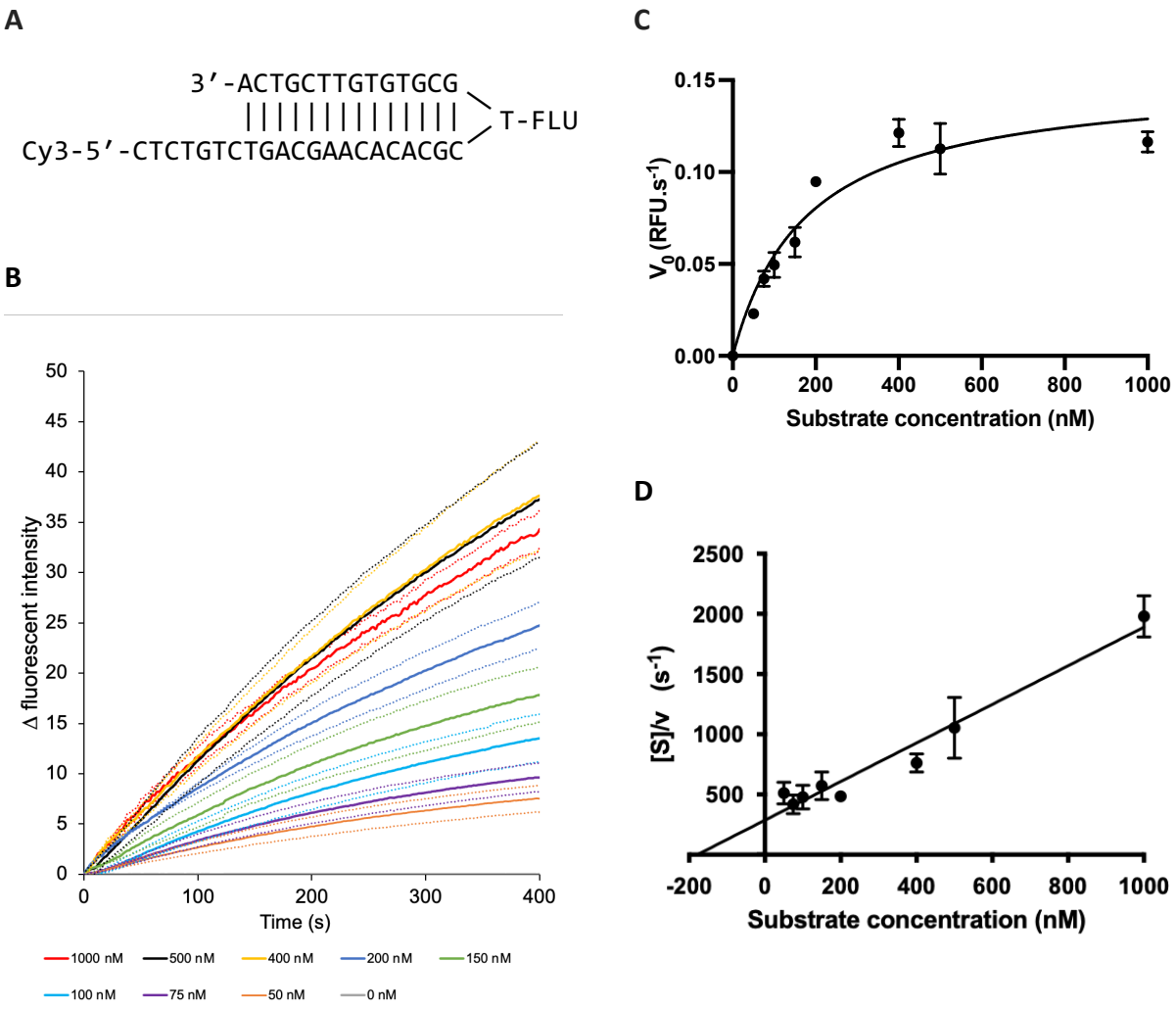


Figure 3-10: Michaelis-Menten analysis of T7 gp6 with overhang substrate

Catalytic parameters were determined using the FRET assay. (A) Cartoon of overhang (OHP4) substrate used; Cy3, cyanine 3; FLU, fluorescein. (B) Reaction progress curves monitored by change in the fluorescence over time at various substrate concentrations by addition of 29.0 nM T7 gp6. Progress curves at each substrate concentration were normalised against a no-enzyme control (C) The initial velocity (V_0) was calculated from the initial rate portion of the progress curves and plotted against substrate concentration. This was analysed by non-linear regression of the Michaelis-Menten equation to determine catalytic parameters. (D) Data was transformed using the Hanes-Woolf equation for linear visualisation. The mean of three repeats is shown and error bars represent SD and error bars represent SEM.

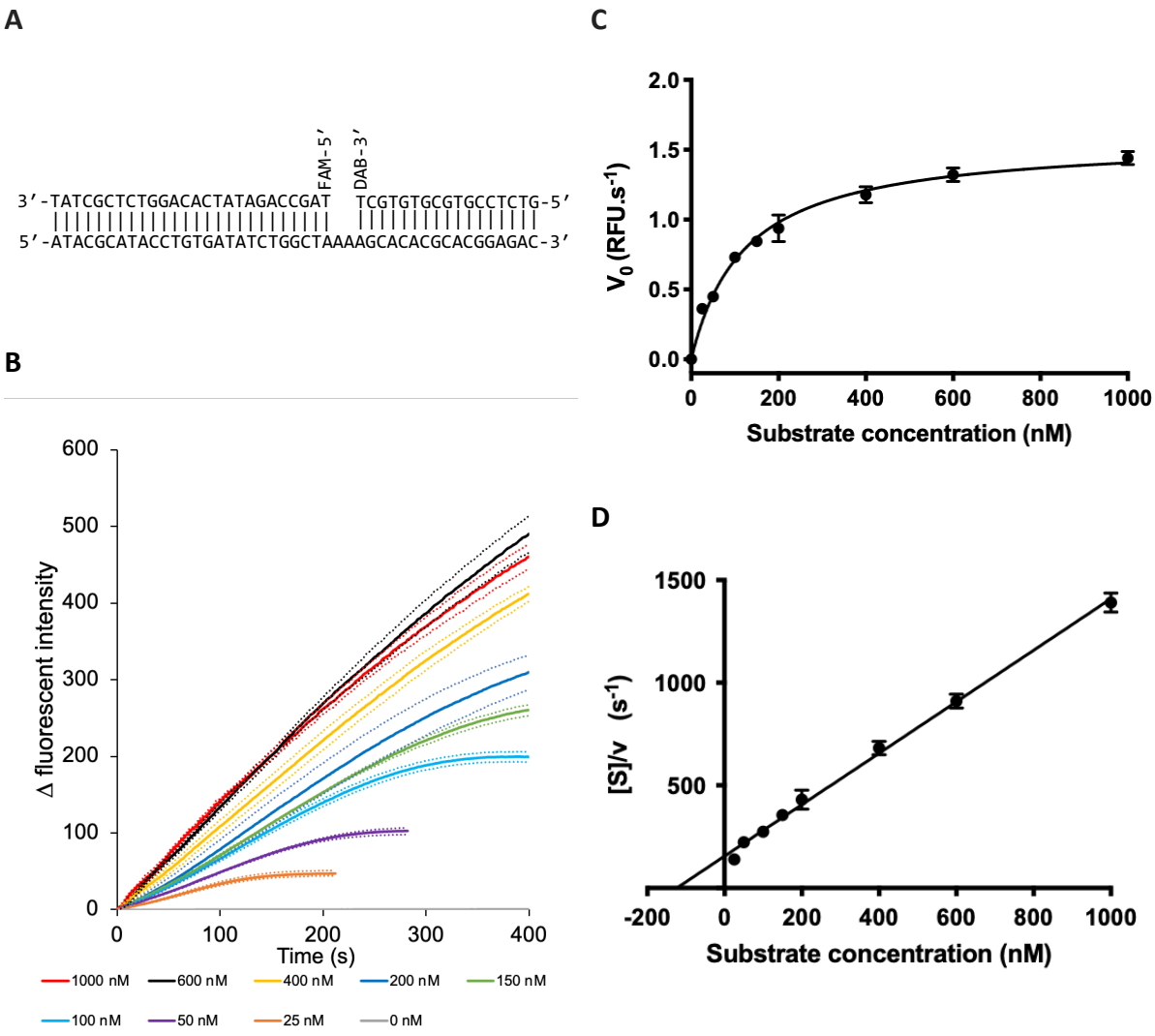


Figure 3-11: Michaelis-Menten analysis of WT T7 gp6 with nicked substrate determined by FRET assay

Catalytic parameters were determined using the FRET assay. (A) Structure of nicked substrate used; FAM, 6-carboxyfluorescein; DAB, DABCYL. (B) Reaction progress curves monitored by change in the fluorescence over time at various substrate concentrations by addition of 2.90 nM T7 gp6. Progress curves at each substrate concentration were normalised against a no-enzyme control. (C) The initial velocity (V_0) was calculated from the initial rate portion of the progress curves and plotted against substrate concentration. This was analysed by non-linear regression of the Michaelis-Menten equation to determine catalytic parameters. (D) Data was transformed using the Hanes-Woolf equation for linear visualisation. The mean of three repeats is shown and dashed lines represent SD and error bars represent SEM.

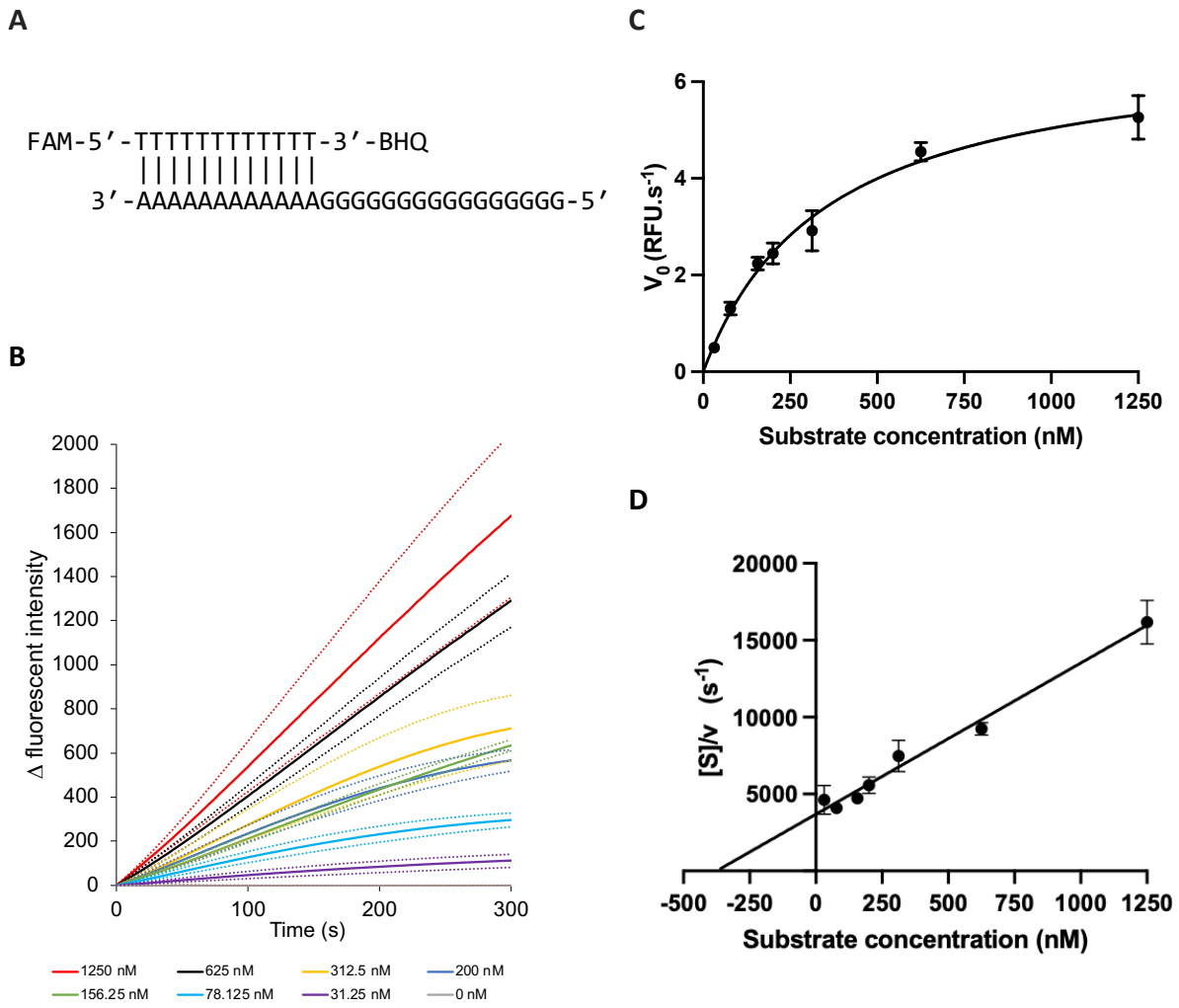


Figure 3-12: Michaelis-Menten analysis of WT T7 gp6 with dsDNA substrate determined by FRET assay

Catalytic parameters were determined using the FRET assay. (A) Structure of dsDNA substrate used; FAM, 6-carboxyfluorescein; BHQ, black hole quencher-1. (B) Reaction progress curves monitored by change in the fluorescence over time at various substrate concentrations by addition of 29.0 nM T7 gp6. Progress curves at each substrate concentration were normalised against a no-enzyme control. (C) The initial velocity (V_0) was calculated from the initial rate portion of the progress curves and plotted against substrate concentration. This was analysed by non-linear regression of the Michaelis-Menten equation to determine catalytic parameters. (D) Data was transformed using the Hanes-Woolf equation for linear visualisation. The mean of three repeats is shown and dashed lines represent SD and error bars represent SEM.

Table 3-1: Catalytic parameters of T7 gp6 with FRET substrates

Substrate	K_M (nM)	k_{cat} (min^{-1})	k_{cat}/K_M (min^{-1} nM^{-1})
Single-flap	104.2 ± 5.7	6.89 ± 0.10	0.0662 ± 0.00374
Double-flap	139.3 ± 8.4	4.59 ± 0.12	0.0329 ± 0.00217
Nicked	83.4 ± 3.2	14.7 ± 0.34	0.176 ± 0.00783
Blunt dsDNA	386.6 ± 75.6	6.30 ± 0.59	0.0163 ± 0.00353
Overhang (OHP4)	180.2 ± 33.1	1.39 ± 0.11	0.00768 ± 0.00154

Data is presented as mean of three repeats ± SEM.

3.2.2.4. Effect of flap length and upstream dsDNA on T7 gp6 endonuclease activity

The overhang substrate (OHP4) used in Figure 3-10 to determine catalytic parameters is composed of 14 nt upstream dsDNA and 7 nt flap. This was designed based on a hairpin single-turnover substrate developed in the Sayers' laboratory to study T5FEN activity in a gel-based assay (Pickering *et al.*, 1999) and later in the FRET assay (Zhang, 2012). The original OHP substrate consists of 10 nt upstream dsDNA and 7 nt flap and was initially used in T7 gp6 reactions. However, the rate of cleavage was extremely slow and 25 nM substrate is not completely cleaved by nearly 50-fold molar excess of T7 gp6 over 400 seconds (Figure 3-13). At lower enzyme concentrations, a requirement for Michaelis-Menten analysis, there was too much noise in the fluorescent signal and therefore in this assay, this substrate was not suitable to determine Michaelis-Menten kinetics, so alternate OHP substrates were investigated.

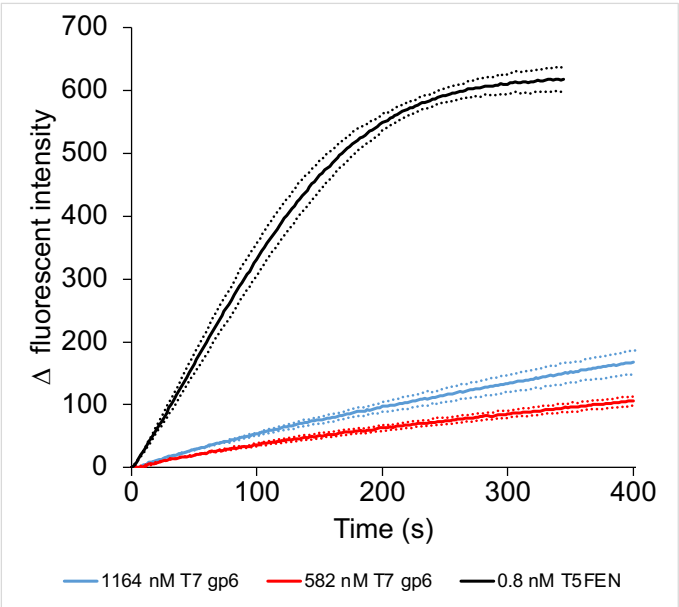


Figure 3-13: Cleavage of overhang (OHP) by T7 gp6

The FRET assay was used to generate progress curves of the cleavage of 25 nM OHP by T7 gp6 (582 nM, 1164 nM) and a T5FEN control (0.8 nM). Progress curves were normalised against no enzyme controls. Data is presented as the mean of three repeats and dashed lines represent SD.

Overhang substrates were designed and tested within this assay to determine the effect of dsDNA and overhang length on T7 gp6 activity. These consisted of varying upstream dsDNA and flap lengths as summarised in Figure 3-14A. Progress curves were collected by measuring the change in fluorescence of 25 nM substrate by the addition of T7 gp6. The initial velocity, $V_{0,}$ was determined and normalised to the amount of enzyme added (Figure 3-14B). The results show that the length of downstream dsDNA considerably impacts the rate of T7 gp6-catalysed cleavage. For example, OHP10, consisting of 20 nt dsDNA, is cleaved over 800-fold faster than the original OHP. The impact of a shorter overhang (OHPs) is not significant compared to OHP. Despite OHP10 being most efficiently cleaved by T7 gp6, it is not the most suitable for FRET analysis as cleavage generates small changes in fluorescent signal with a large amount of noise due to increased FRET distances.

A

Substrate	dsDNA length (nt)	5' overhang length (nt)
OHPs	10	4
OHP	10	7
OHP1	11	7
OHP2	12	7
OHP4	14	7
OHP10	20	7

B

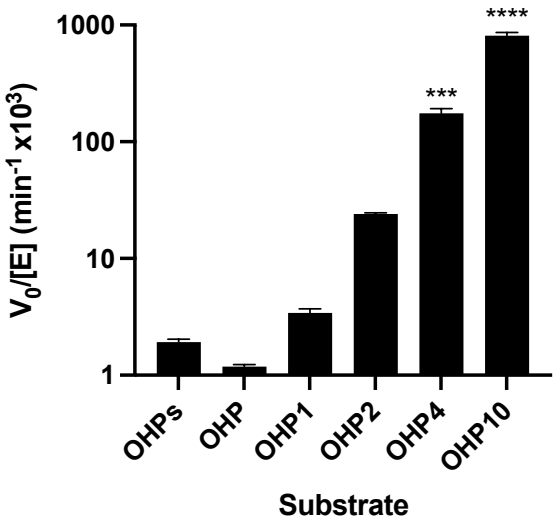


Figure 3-14: Effect of double stranded region and flap length on T7 gp6 cleavage of overhang substrate

The FRET assay was used to measure T7 gp6 cleavage of overhang substrates with varying lengths of dsDNA and flaps. Progress curves for 25 nM substrate with T7 gp6 were generated, normalised against no protein control and initial velocity, V_0, calculated. The concentration of T7 gp6 was varied according to substrate to give a measurable rate of cleavage. (A) Double stranded region and flap length of the overhang substrates investigated. (B) The V_0 was normalised against concentration of enzyme added, [E], to give the turnover rate. Data is presented as the mean of three repeats and error bars represent SEM. Statistical significance was determined by a one-way ANOVA with Dunnett's multiple comparisons against the OHP control; ns $p>0.05$, ***$p<0.001$, ****$p<0.0001$.

3.2.3. Determination of optimal reaction conditions for T7 gp6 nuclease activity

3.2.3.1. T7 gp6 exonuclease activity displays monovalent metal cation sensitivity

In order to determine optimal reaction conditions for recombinant T7 gp6 exonuclease activity, the UV exonuclease assay was used to investigate the sensitivity of the enzyme on the monovalent metal cations K^+ and Na^+. The reaction was carried out at pH 9.3 in a glycine buffer and the pH of the buffer was adjusted with NaOH (for K^+ dependency experiments) or KOH (for Na^+ dependency experiments). The specific activity at varying KCl and NaCl concentrations was calculated (Figure 3-15A-B). These results show that T7 gp6 exhibits monovalent metal cation sensitivity and specificity. Concentrations of KCl up to 250 mM do not significantly affect the specific activity, but 300 mM KCl is inhibitory. The activity of T7 gp6 is more sensitive to inhibition by NaCl, and concentrations at 50 mM and above significantly inhibit the specific activity of the enzyme.

To examine the effect of the monovalent ion present within the hydroxide used to set the pH of the buffer, the experiment was repeated at pH 8 using Tris-HCl, which contained no metal monovalent cation (Figure 3-15C-D). In this buffer, the KCl dependency of T7 gp6 is more significant, with a bell-shaped dependency profile. The optimum KCl concentration is 200 mM and this results in approximately a 2-fold increase in specific activity compared to when no KCl is present. In this buffer, T7 gp6 is more tolerant of NaCl, and no significant difference is observed for any NaCl concentration tested compared to when no NaCl is present.

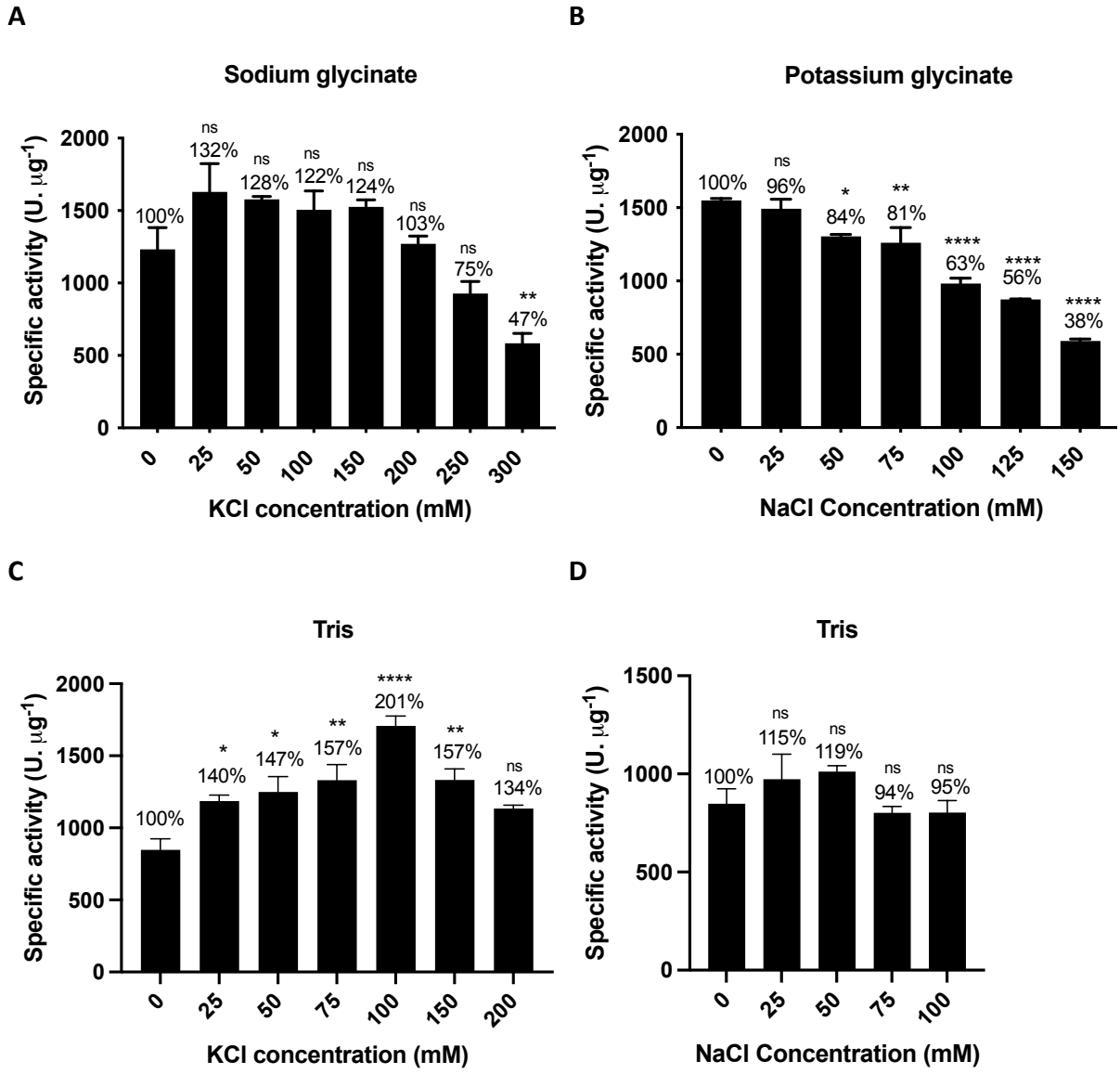


Figure 3-15: Monovalent salt dependency of T7 gp6 exonuclease activity

The UV spectrophotometric assay was used to determine the specific activity of T7 gp6 in the presence of NaCl and KCl. The release of acid-soluble nucleotides from high-molecular-weight DNA was measured. The reactions were carried out in 667 µg.mL^{-1} type XIV DNA, 25 mM buffer, 10 mM $MgCl_2$ and 1 mM DTT. To test the effect of cation identity and ionic strength on this dependency, two buffers were compared: Glycine (A, KCl; B, NaCl) and Tris-HCl (C, KCl; D, NaCl). The pH of the glycine buffer adjusted with KOH (for NaCl dependency) or NaOH (for KCl dependency). Specific activity is depicted as the mean of three repeats and error bars represent SEM. The % activity determined at 0 mM monovalent salt was set as 100% for each separate experiment. Statistical significance was determined by a one-way ANOVA with Dunnett's multiple comparisons against the 0 mM control for each experiment, ns $p>0.05$, *$p<0.05$ **$p<0.01$, ****$p<0.0001$.

3.2.3.2. T7 gp6 exonuclease activity displays divalent metal cation sensitivity

The literature shows that FENs exhibit varying levels of divalent metal ion specificity and sensitivity (Harrington and Lieber, 1994; Garforth *et al.*, 2001; Zheng *et al.*, 2002). This was investigated for T7 gp6 and several divalent metal cations were examined for their ability to support T7 gp6 catalysed DNA cleavage in the UV assay (Table 3-2). Salts were used to provide the divalent ions: Mg^{2+} ($MgCl_2$), Mn^{2+} ($MnCl_2$), Ca^{2+} ($CaCl_2$), Zn^{2+} ($Zn(CH_3CO_2)_2$), Co^{2+} ($CoCl_2$), Ni^{2+} ($NiCl_2$). The concentration of salt was varied in a standard reaction buffer of 25 mM potassium glycinate pH 9.3, 667 $\mu g.mL^{-1}$ type XIV DNA, 50 mM KCl and 1 mM DTT. It was necessary to omit the DTT for Zn^{2+}, Co^{2+} and Ni^{2+} due to formation of precipitation. However, before carrying out these experiments, it was determined that omission of DTT does not result in a significant change in T7 gp6 catalysed hydrolysis of DNA within this assay (data not shown).

In this assay, the divalent ions Mg^{2+}, Mn^{2+} and Co^{2+} are able to support T7 gp6 nuclease activity, whilst no DNA hydrolysis is detected with Ca^{2+}, Ni^{2+} and Zn^{2+} under the conditions investigated. The ions Mg^{2+} and Mn^{2+} facilitate the highest levels of nuclease activity, although, the concentration dependency profiles vary quite considerably between these. Concentrations of 0.5 mM – 10 mM $MgCl_2$ elicit the highest activity and T7 gp6 is rather insensitive to concentrations up to 50 mM. However, T7 gp6 displays slightly lower activity and a much greater concentration sensitivity with $MnCl_2$ as a cofactor, with concentrations 0.05 mM – 0.5 mM producing the highest levels of catalysis.

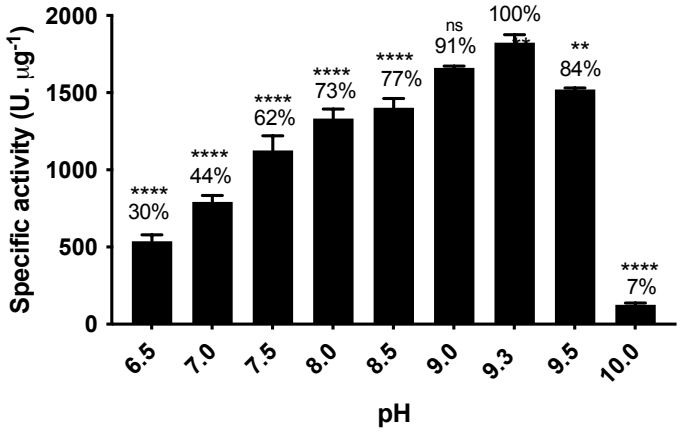
3.2.3.3. T7 gp6 exonuclease activity displays pH sensitivity

FEN nuclease activity is highly dependent on pH (Pickering *et al.*, 1999). The UV assay was used to determine optimum pH of T7 gp6 activity (Figure 3-16). For this, a range of buffers, selected based on their useful pH range were utilized. The results show that the specific activity of T7 gp6 increases with increasing pH, up to an optimum of pH 9.3. There is a sharp decrease to pH 10, which shows only 7% of the specific activity compared to pH 9.3.

Table 3-2: Divalent salt dependency of T7 gp6 exonuclease reaction

[Cofactor] (mM)	Mg^{2+}		Mn^{2+}		Co^{2+}		Ca^{2+}	Ni^{2+}	Zn^{2+}
	Specific activity (U. μg^{-1})	Relative activity (%)	Specific activity (U. μg^{-1})	Relative activity (%)	Specific activity (U. μg^{-1})	Relative activity (%)	Specific activity (U. μg^{-1})	Specific activity (U. μg^{-1})	Specific activity (U. μg^{-1})
50	1350 ± 417	73%	438 ± 41	24%	-	-	-	-	-
20	1440 ± 104	77%	325 ± 16	17%	-	-	-	-	-
10	1860 ± 196	100%	337 ± 18	18%	196 ± 5	11%	ND	ND	ND
8	1690 ± 145	91%	313 ± 18	17%	-	-	-	-	-
5	1730 ± 196	93%	364 ± 45	20%	-	-	-	-	-
2	1970 ± 109	106%	515 ± 84	28%	-	-	-	-	-
1	1950 ± 91	105%	701 ± 54	38%	ND	ND	ND	ND	ND
0.5	1670 ± 39	90%	1020 ± 28	55%	-	-	-	-	-
0.1	521 ± 25	28%	1230 ± 28	66%	-	-	-	-	-
0.05	292 ± 45	16%	1460 ± 134	78%	-	-	-	-	-
0.01	-	-	809 ± 178	43%	-	-	-	-	-

The UV spectrophotometric assay was used to determine the specific and relative activity of T7 gp6 in the presence of divalent cations. The release of acid-soluble nucleotides from high-molecular-weight DNA substrate was measured. The relative activity was normalized to 10 mM $MgCl_2$ (100%). Various concentrations of the salts $MgCl_2$, $MnCl_2$, $CoCl_2$, $CaCl_2$, $NiCl_2$, $Zn(CH_3CO_2)_2$ were tested. The reactions were carried out in 667 $\mu g.mL^{-1}$ type XIV DNA, 25 mM potassium glycine pH 9.3, 50 mM KCl and 1 mM DTT. Specific activity is depicted as the mean of three repeats ±SEM. ND, activity not detected with 5 μg T7 gp6; -, condition not assessed.



pH	Buffer
6.5	MES-NaOH
7.0	HEPES-KOH
7.5	HEPES-KOH
8.0	Tris-HCl
8.5	Tris-HCl
9.0	Potassium glycinate
9.3	Potassium glycinate
9.5	Potassium glycinate
10	CHES-KOH

Figure 3-16: The pH sensitivity of WT T7 gp6 exonuclease activity

The UV spectrophotometric assay was used to determine the specific activity of WT T7 gp6 at different pH values. The release of acid-soluble nucleotides from high-molecular-weight DNA substrate was measured. The reactions were carried out in 667 μg.mL^{-1} type XIV DNA, 25 mM buffer, 50 mM KCl, 10 mM $MgCl_2$ and 1 mM DTT. Specific activity is depicted as the mean of three repeats ±SEM. The % activity determined at pH 9.3 was set as 100%. Statistical significance was determined by a one-way ANOVA with Dunnett's multiple comparisons against the pH 9.3 control for each experiment, ns $p>0.05$, *$p<0.05$ **$p<0.01$, ****$p<0.0001$. The table shows the buffers used for each pH investigated.

3.2.3.4. T7 gp6 endonuclease activity displays monovalent cation sensitivity

The FRET assay was used to determine KCl and NaCl sensitivity of the endonuclease reaction (Figure 3-17). For this, the initial rate of cleavage of 200 nM single-flap FRET substrate by 14.5 nM T7 gp6 was determined. The reactions were carried out in HEPES pH 7.4 which had the pH adjusted with NaOH (for KCl assessment) or KOH (for NaCl assessment). The results show that KCl is stimulatory to T7 gp6 activity, and the optimum concentration of 30 mM KCl increases the initial rate of reaction by over 8-fold compared to when no KCl is present. In contrast, NaCl is inhibitory to the cleavage of the single-flap substrate, resulting in a significant decrease in the initial rate of reaction, at 50 mM NaCl and above, compared to when no NaCl is present. This monovalent cation dependency was also investigated in Tris-HCl pH 8 (Figure 3-17C, D). The results show that the KCl dependency exhibits a similar bell-shaped pattern, with 25 mM KCl resulting the highest increase in the initial rate of reaction, by nearly 14-fold, compared to when no KCl is present. No significant effect is seen by addition of NaCl, although the relative errors in this experiment are high.

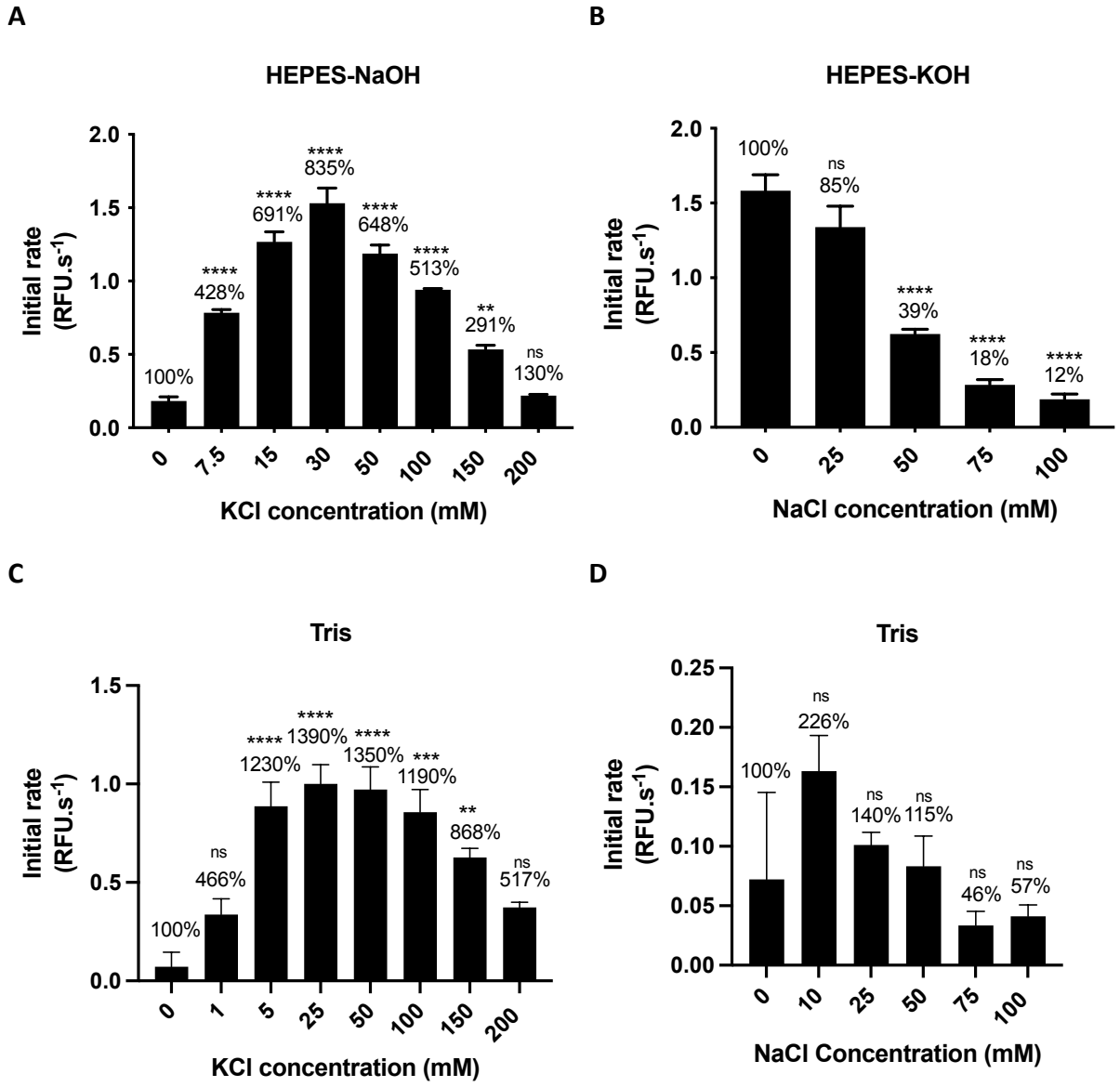


Figure 3-17: Monovalent salt dependency of T7 gp6 endonuclease activity

The FRET assay was used to measure the cleavage of 200 nM single-flap substrate by 14.5 nM T7 gp6 and the initial rate of reaction was calculated. The reactions were carried out in 667 $\mu g.mL^{-1}$ type XIV DNA , 25 mM buffer, 10 mM $MgCl_2$, 2 mM DTT, 0.5 mM EDTA, 0.1 $mg.mL^{-1}$ acetylated BSA and varying concentrations of KCl or NaCl. To test the effect of ionic strength on this dependency, two buffers were compared: HEPES (A, KCl; B, NaCl) and Tris-HCl (C, KCl; D, NaCl). The pH of the HEPES buffer adjusted with KOH (for NaCl dependency) or NaOH (for KCl dependency). Results are presented as the mean of three repeats and error bars represent SEM. The % activity determined at 0 mM monovalent salt was set as 100% for each separate experiment. Statistical significance was determined by a one-way ANOVA with Dunnett's multiple comparisons against the 0 mM control for each experiment, ns $p>0.05$, ** $p<0.01$, *** $p<0.001$, **** $p<0.0001$.

3.2.4. Analysis of T7 gp6 binding to DNA

The ability of T7 gp6 to bind endonuclease and exonuclease DNA substrates was investigated using an electrophoretic mobility shift assay (EMSA; Figure 3-18). The substrates investigated had the same DNA sequences as those used in the FRET assay but were 5'-labelled with ^{32}P to facilitate gel-based visualisation at low concentrations. For the position of ^{32}P labelling, see section 2.8.1.

Various concentrations of T7 gp6 was incubated with 500 pM substrates on ice for 1 hour, in the presence of 5 mM $CaCl_2$ to inhibit DNA cleavage. This incubation time was optimized to enable binding equilibrium to be reached. The enzyme: substrate complex was separated from free substrate on a native PAGE gel ran at 4°C and quantified by phosphorimaging. Gel images show that at high enzyme concentrations, the majority of substrate is bound to T7 gp6, whilst at low enzyme concentrations the majority of substrate is unbound. Furthermore, affinity towards the ssDNA present in the single and double-flap substrates is not detected in this assay (Figure 3-18). Data was plotted in GraphPad prism and fitted by a non-linear regression to the one-site specific binding model to calculate K_D (Figure 3-18D). The binding affinity for single-flap, double-flap and nicked substrates are similar whilst affinity for the overhang is much poorer (Table 3-3).

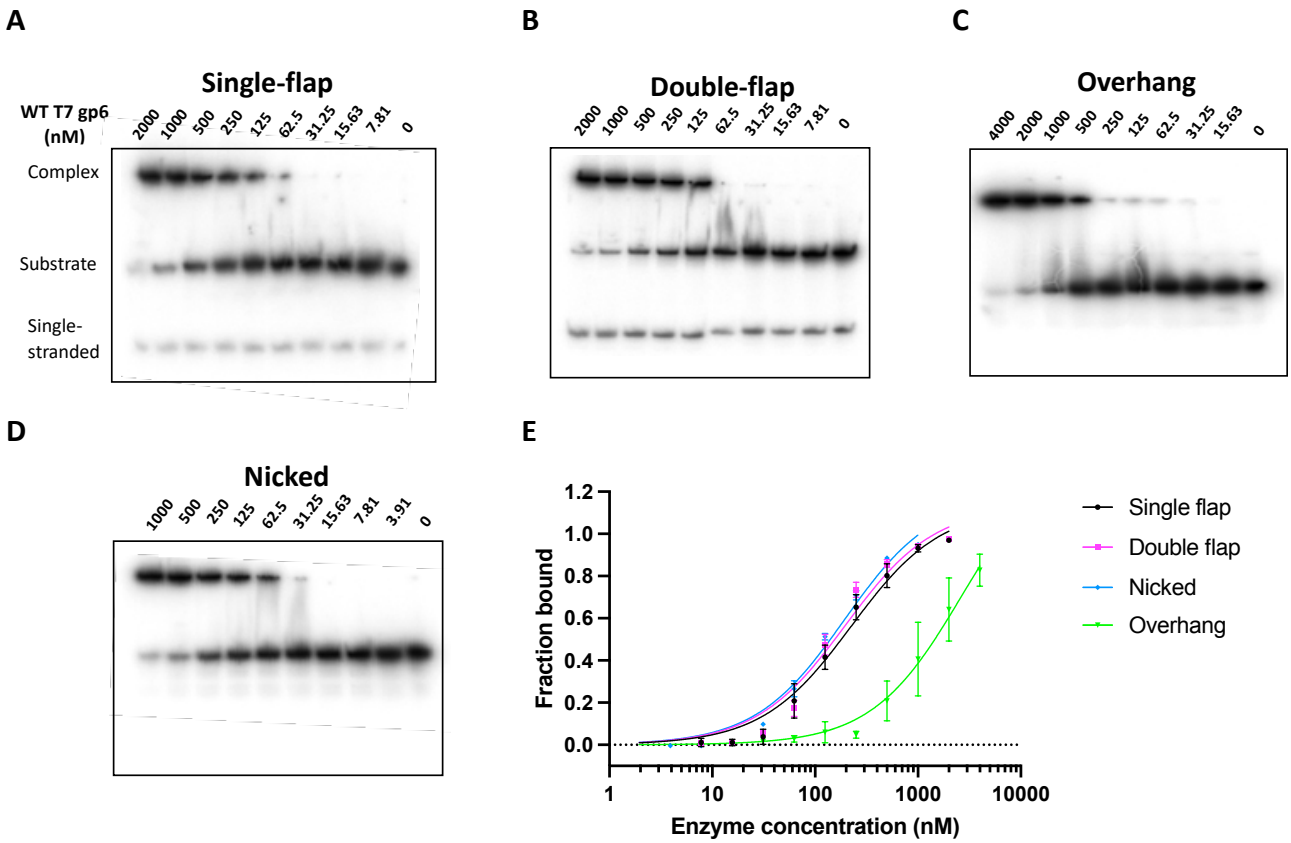


Figure 3-18: Substrate binding affinity of T7 gp6

Binding of substrates to T7 gp6 was examined using the electrophoretic mobility shift assay (EMSA). Various concentrations of T7 gp6 (indicated by lane numbers, nM) was incubated with 500 pM of (A) single-flap, (B) double-flap, (C) overhang or (D) nicked ^{32}P-labelled substrates on ice for 1 hour before resolving enzyme-substrate complex from unbound substrate on a nondenaturing 17% acrylamide gel. Representative images are shown from three replicates. (E) The gels were quantified for substrate bound at each enzyme concentration and fitted to the one-site -specific binding model by non-linear regression on GraphPad Prism. Data is presented as the mean of three repeats and error bars represent SEM. Representative gel image given for each EMSA.

Table 3-3: Summary of calculated K_D for T7 gp6 with various substrates

Substrate	K_D (nM)
Single-flap	243 ± 53
Double-flap	201 ± 22
Nicked	194 ± 11
Overhang (OHP4)	1510 ± 499

3.3. Discussion

3.3.1. Over-expression and purification of T7 gp6

Small-scale expression trials of T7 gp6 were undertaken to develop an expression protocol that was able to provide sufficient purified protein for biochemical characterisation. The plasmids pTTQ18-T5 and pJONEX4 and suitable *E. coli* strains were explored for this (Figure 3-1 to Figure 3-2). The pJONEX4/ΔH1+(λcI857) system was chosen for large-scale protein expression. In the pJONEX4 vector, gene expression is under the control of the promoter, λP_L. The *E. coli* expression strain ΔH1+, which is also known as M72(λcI857), carries a defective lambda prophage carrying a *cI*857 gene, a thermolabile repressor of the λP_L promoter (Remaut *et al.*, 1981). When cells are grown at temperatures below 30°C, the cI857 forms dimers which bind co-operatively to operators for the promoter and inhibit gene transcription. When gene expression is induced at 42°C, the repressor is inactivated due to production of heat-shock response proteins which prevent dimer formation (O'Connor and Timmis, 1987; Valdez-Cruz *et al.*, 2010) and therefore, transcription is permitted. Furthermore, in this plasmid there is a convergent downstream lactose promoter (P_{lac}), which transcribes antisense messenger RNA (mRNA) that block the transcription of leaky mRNA from the λP_L promoter before induction, to give a tight control of protein expression (O'Connor and Timmis, 1987). This is important as T7 gp6 is predicted to be toxic to the cells as it is a nuclease. Therefore, throughout all growth stages, the media supplemented with 0.1 mM IPTG.

Levels of T7 gp6 expression in both expression systems explored is low in comparison to various bacterial FENs expressed within the laboratory from the same plasmids (Figure 4-4; Allen *et al.*, 2009; Oates, 2016). A low level of protein expression could result from residual protein expression prior to induction. Leaky expression can pose an issue to cells, especially when the function of the protein is detrimental to function of the host cells and can also lead to plasmid instability (Rosano and Ceccarelli, 2014). As T7 gp6 is a nuclease, it would not be surprising if it is toxic to *E. coli*, due to digestion of DNA within the cells. In both systems investigated, steps were taken to try to prevent leaky expression. In expression from pJONEX4, all media was supplemented with 0.1 mM IPTG to induce the antisense P_{lac} to transcribe mRNA to help counteract leakage from λP_L. Whilst, pTTQ18-T5 containing cells were grown only in

vegetable-derived tryptone to prevent any basal expression from the T5 promoter due to low levels of lactose present in casein-derived tryptone. Alternative methods which could be used to combat leaky expression include the use of specialised expression strains, alternative expression systems and co-expression of a repressor (Rosano and Ceccarelli, 2014; Giacalone *et al.*, 2006).

Prolonged incubation after induction of protein expression can be used to increase cell biomass. However, in the case of T7 gp6, the relative levels of expression drop after an overnight incubation at a low temperature (Figure 3-2) and therefore, cells were harvested immediately after induction. Expression of a toxic protein can cause death of the expressing cells and relative expression can decrease if there is overgrowth of non-expressing mutant cells. Alternatively, the level of relative expression could decrease if proteolysis is occurring.

The theoretical pI of T7 gp6 is 4.93 (Gasteiger *et al.*, 2005). This suggests that during ion-exchange chromatography, the protein will interact with the anion exchanger (Q) at pH 8 and not with the cation exchanger (SP) at pH 7. However, during purification, the protein interacted with and was eluted from both anion and cation exchange columns. This could be due to how the secondary and tertiary structure affect the charge distribution within the protein, such as through exposed or hidden side chains and salt bridges and therefore, the actual pI of the protein differs from the theoretical pI. Running the SP column at a higher pH may help overcome the interaction, although, SDS-PAGE analysis indicates that the purification protocol used is adequate and final protein yields of 50 mg protein from the 38 g cell pellet are >95% pure (Figure 3-4).

Zymography was used to detect nuclease activity of the final purified protein (Rosenthal and Lacks, 1977). This gel is made in the same way as a normal SDS-PAGE gel, except a high-molecular-weight DNA substrate is incorporated pre-polymerisation. This DNA is double-stranded and is composed of a heterogeneous mixture of nicks, overhangs and blunt ends and therefore, in the case of FENs mainly represents exonuclease substrates. The zymogram indicates that only T7 gp6 possesses nuclease activity with no contaminating bands (Figure 3-4). However, the absence of a band does not completely correlate with an absence of nuclease activity. Zymogram gels contain the detergent SDS and therefore rely on the ability

of proteins to be renatured *in situ* after electrophoresis. Furthermore, the presence of a band does not give definitive evidence that a protein possesses nuclease activity, as tight binding to DNA may prevent dye interactions with DNA and also appear as a dark band. Whilst the zymogram only provides a crude assessment of nuclease activity of proteins, it is useful for determining if a protein sample is obviously contaminated with endogenous *E. coli* nucleases.

3.3.2. T7 gp6 activity on plasmid DNA

T7 gp6 is able to degrade linear and nicked plasmid DNA but not cccDNA in the presence of 10 mM $MgCl_2$ within the timeframe investigated (Figure 3-6). This indicates that T7 gp6 possess double-stranded exonuclease activity but not double-stranded endonuclease activity. This finding is supported in early literature, which showed that T7 gp6 is unable to cleave duplex phage DNA (Kerr and Sadowski, 1972). However, this should not be ruled as other FENs have been shown to cleave cccDNA under specific reaction conditions. For example, T5FEN cleaves cccDNA through the generation of nicked and linearized intermediates at low monovalent salt concentrations in the presence of several divalent metal ion co-factors, with Mn^{2+} the most efficient (Garforth *et al.*, 2001). As only Mg^{2+} was tested in this experiment, it remains to be seen if experimental conditions impact upon this substrate specificity.

The ability of T7 gp6 to cleave certain structures in plasmid DNA preparations may prove useful in the preparation of plasmid DNA for transformation. Selective removal of linear and nicked DNA could be used to improve cloning efficiency. The use of T5FEN to clean-up plasmid DNA preparations has been demonstrated and has been shown to enhance the transformation efficiency of DNA minipreps obtained from plasmid cDNA libraries (Kiss-Toth *et al.*, 2001).

3.3.3. UV assay analysis of WT T7 gp6 exonuclease activity

The UV assay was used to investigate exonuclease activity of T7 gp6. As a comparison, the activity of the commercially-sourced T7 gp6 (provided by the industrial partner) and T5FEN was also determined in the assay under the same reaction conditions. The amount of enzyme added was optimized to 0.5 μg in an 800 μL reaction. This resulted in the first 4 data points (initial 10 minutes) of the T7 gp6 reactions and first 5 data points (initial 15 minutes) of the

T5FEN reaction being linear, giving sufficient data points for determination of specific activity by a linear regression. Protein stocks were freshly diluted to 0.125 $mg.mL^{-1}$ in assay buffer just before starting the experiment, so that 0.5 μg represented 4 μL. This volume was sufficient enough to promote accurate pipetting but also represented only 0.5% of the reaction volume. Alternatively, 4 μL buffer was analysed and used as a negative control. To calculate the specific activity, the $A_{260\ nm}$ of the negative control was first subtracted from each time point, since any nucleotides released with buffer only must be due to experimental conditions and not enzyme cleavage.

The progress curves and specific activity of the lab purified compared to the commercial T7 gp6 are the same within the errors of the experiment (lab, 1670 ± 113 $U.\mu g^{-1}$; commercial, 1600 ± 56.6 $U.\mu g^{-1}$), as determined by a one-way ANOVA. The comparison between lab purified and commercially-sourced T7 gp6 is important as it provides quality control to the protein, which could be crucial for findings to be replicated within the IO device in the future. Whilst there is no statistical difference in the specific activity of T5FEN (1560 ± 142 $U.\mu g^{-1}$) compared to the T7 gp6 proteins, the progress curves between these are visibly different, with a much greater endpoint and therefore release of acid-soluble nucleotides from the T5FEN reaction. This may result from a smaller minimal substrate requirement of T5FEN compared to T7 gp6. Indeed, variations in the smallest substrate requirements of FENs has been noted in the literature. For example, the smallest substrate cleaved by T5FEN is 6 nt (Sayers and Eckstein, 1990), whilst for T4 RNase H, this is 8 -11 nt (Bhagwat *et al.*, 1997) and hFEN-1 is 10 nt (Liu *et al.*, 2006). This minimal substrate requirement was investigated further and when 0.5 μg T5FEN was added to a T7 gp6 reaction which had reached a plateau at 30 minutes, a further release of nucleotides was detected. However, addition of 0.5 μg T7 gp6 to a T7 gp6 reaction or a T5FEN reaction at endpoint did not illicit further release of acid-soluble nucleotides (results not shown).

The specific activity of T5FEN within the same UV assay, albeit with slightly different experimental conditions, (25 mM potassium glycinate pH 9.3, 10 mM $MgCl_2$) has been previously determined as 1300 $U.\mu g^{-1}$ (Sayers and Eckstein, 1990). This is in good agreement with the results obtained here. Several other FENs have been investigated in this assay and

demonstrate that eukaryotic and bacterial FENs possess much lower exonuclease activity. For example, the eukaryotic protist *Trypanosoma Brucei* FEN, had a specific activity of 0.42 $U.\mu g^{-1}$ (Oates, 2016), whilst the FEN domain of the bacterial *Streptococcus pneumoniae* DNA polymerase I had a specific activity of 23 $U.\mu g^{-1}$ (Lau, 2017).

3.3.4. Determination of T7 gp6 catalytic parameters with the FRET assay

To explore the endonuclease and exonuclease activity of T7 gp6 in depth, a FRET assay, utilizing fluorescent dual-labelled substrates was used. This assay was carried out in a temperature controlled single cuvette spectrophotometer to provide real-time analysis of substrate cleavage. Endonuclease (single-flap, double-flap and overhang) and exonuclease (nicked, dsDNA) substrates were investigated and catalytic parameters determined for each, under identical reaction conditions, using the Michaelis-Menten model of enzyme kinetics. In an intact substrate, the two fluorophores are in close proximity and due to overlapping spectra, the signal released after excitation of one fluorophore is quenched by the other due to FRET. However, when the substrate is cleaved, the two fluorophores become separated and FRET no longer occurs. As a result, the emission detected from the excited fluorophore is detected. Conventional nuclease assays predominantly use radioactive labelling to visualise DNA substrates on gels and these prove highly specific and sensitive results with a benefit that cleavage products can be visualised. However, they are often time consuming and are discrete so do not provide real-time visualisation of catalytic activity, whereas the FRET assay overcomes these issues.

To ensure stability of the cyanine-3 (Cy3) fluorophore present in some substrates, this assay was carried out at pH 7.4. Before collecting progress curves at varying substrate concentrations, the amount of enzyme added was optimized for each substrate to give an initial velocity (V_0) that could be measured by linear regression of the first few data points. The concentration of enzyme was also kept low enough, such that $[E] << [S]+K_M$ (Choi *et al.*, 2017), so the steady-state assumption of the Michaelis-Menten equation was met, but also so that visible progress could be seen over the time course of the reaction.

The single-flap substrate is composed of a 4 nt 5' flap with 16 nt downstream and 13 nt upstream dsDNA. The double-flap substrate is the same as the single-flap with the addition of

a 1 nt 3' flap. These flap substrates were annealed with a 25% excess of labelled oligonucleotide in order to favour formation of the complex. Annealing at this ratio prevented sigmoidal progress curves which show an initial lag at the beginning of the reaction present when the oligonucleotides were annealed at 1:1, possibly due to the formation of unlabelled substrates which are preferentially cleaved at the beginning of the reaction (results not shown).

The overhang substrate, OHP4, is composed of 14 nt dsDNA with a 7 nt overhang. It was designed based on the original overhang substrate, OHP, composed of 10 nt dsDNA with a 7 nt overhang, used in earlier studies that determined the kinetic parameters of T5FEN in a gel-based cleavage assay. This study confirmed OHP as a single-turnover substrate for T5FEN and the cleavage site 1 nt into the dsDNA region (Pickering *et al.*, 1999). Initial attempts at kinetic characterisation of T7 gp6 with an overhang substrate in this study utilized OHP. However, initial optimizations indicated that the K_M was small and due to an extremely slow rate of reaction, a high excess of enzyme was required to detect cleavage. Furthermore, whilst photobleaching was corrected for by the use of a non-enzyme control, with this substrate and with experimental conditions which satisfied the Michaelis-Menten requirements, it was a major contributor to the change in fluorescence detected. Therefore, a variety of OHP variants were designed and analysed in more detail (section 3.3.4), which had altered dsDNA and overhang lengths, with OHP4 selected for detailed kinetic characterisation.

The nicked substrate is composed of a 2 nt nick with 17 nt upstream dsDNA and 26 nt downstream dsDNA. This has been used in a 96-well plate format to determine the cation dependence of *E. coli* DNA polymerase I FEN domain (Zhao *et al.*, 2014). Due to the size of the substrate and free 5' ends, it is unlikely that this is single turnover and multiple cleavages could take place after separation of the fluorophore and quencher.

The dsDNA substrate is composed of 12 nt dsDNA with a 16 nt 5' overhang (Figure 3-12A). The overhang is composed of only guanine nucleotides and is expected to form a secondary structure called a G-quadruplex. A G-quadruplex is resistant to cleavage by many nucleases (Bishop *et al.*, 1996). Specifically, T5FEN is unable to cleave only G extensions (Sayers and

Eckstein, 1990). Therefore, this substrate is designated as dsDNA, with the G-rich extension directing cleavage towards one side of the substrate.

The progress curves for the single-flap and double-flap substrates gave a surprising result as the fluorescence decreased over the course of the reaction (Figure 3-8; Figure 3-9) This was also observed for the commercially-sourced T7 gp6 but not for T5FEN and various bacterial FENs assayed within the laboratory (results not shown). A decrease in fluorescence indicates that the fluorophores become closer together during the T7 gp6 reaction and may result from substrate or product interaction with T7 gp6. However, this is unlikely as the substrate is in a high excess over the enzyme and therefore one enzyme molecule would need to be interacting with many substrate or product molecules. An alternative explanation could be due to an energy transfer mechanism, distinct from the quenching involved in FRET, known as static quenching. Compared to FRET, the distances involved in static quenching are much smaller, causing the two fluorophores to physically associate and whilst the energy from excitation is transferred to the contact molecule, it is lost, mainly through heat rather than fluorescent emission and therefore the fluorescent signal decreases (Marras *et al.*, 2002). In the case of this assay, T7 gp6 could bind the substrate in such a conformation or release the product in such an orientation that the two fluorophores physically associate.

An explanation which must also be considered is that the substrate is being cleaved exonucleolytically from the 5' end of the flap or invader oligonucleotide. This could result in an increased flexibility of the labelled A3 oligonucleotide in the substrate and subsequently the two fluorophores are able to move close together. Ideally the cleavage products should be determined. This could be achieved by running cleavage products on a denaturing sequencing gel and imaging for Cy3 or FAM detection. Whilst this will not detect cleavage products of the unlabelled oligonucleotide, it can be used to confirm whether the 5' flap is cleaved. Alternatively, reaction products could be analysed by mass spectrometry to yield the molecular weight of cleavage products. This technique has been used successfully to determine the cleavage products of T5FEN (Pickering *et al.*, 1999). However, samples require de-salting prior to mass spectrometry analysis and desalting resins do not possess the sensitivity to separate single nucleotides from salts.

The nicked substrate possessed the highest specificity constant (k_{cat}/K_M) in this assay (0.176 ± 0.00783 min^{-1} nM^{-1}). This was over 3-fold greater than that of the next most catalytically efficiently cleaved substrate, the single-flap (0.0662 ± 0.00374 min^{-1} nM^{-1}). The results from this indicate that T7 gp6 cleave exonuclease substrates more optimally than endonuclease substrates. However, the length of interacting dsDNA is likely also plays a crucial role as shown by cleavage of various OHP substrates in section 3.3.4.

The single-flap substrate possessed a lower K_M, higher k_{cat} and consequently nearly a 2-fold increase of specificity constrant when compared to the double-flap substrate (Table 3-1). Many FENs have been shown to have preference for a double-flap substrate. This is due to a 3' binding pocket present in eukaryotic and archaeal FENs, which functions to provide a high substrate selectivity and cleavage specificity which result in higher catalytic efficiency (Kao *et al.*, 2002; Friedrich-Heineken and Hübscher, 2004; Williams *et al.*, 2007). To leave a nick which can be directly ligated. The 3' binding pocket is not present in bacteriophage FENs (Ceska *et al.*, 1996; Devos *et al.*, 2007) and explains why this substrate specificity is not observed with T7 gp6.

In the literature, the catalytic parameters of T7 gp6 have been determined for a blunt ended substrate and overhang substrates of various lengths. Mitsunobu *et al.* (2014) found that a blunt ended substrate is cleaved most efficiently and has a k_{cat} of 2.84 min^{-1}, whilst overhang substrates are cleaved less efficiently, with the k_{cat} decreasing as the overhang length increases (0.57 min^{-1} for 2 nt overhang compared to 0.023 min^{-1} for 20 nt overhang). These reactions were carried out at pH 8 in 1 mM $MgCl_2$. The k_{cat} obtained for OHP4 here, is nearly 3-fold higher than for the 2 nt overhang substrate in the study, whilst our K_M also several fold higher (Mitsunobu *et al.*, 2014). However, this is still comparable considering the differing reaction conditions in both studies.

The catalytic parameters for T5FEN determined in the literature are also comparable to the results obtained here. For example, the k_{cat} of T5FEN with fluorescently or ^{32}P labelled OHP substrate is 101 - 185 min^{-1} (Dervan *et al.*, 2002; Zhang, 2012), whilst for a blunt dsDNA substrate the k_{cat} is 687 min^{-1} (Williams *et al.*, 2007), determined at pH 9.3. There is a log relationship between k_{cat} and pH in T5FEN (Pickering *et al.*, 1999) and given this, the k_{cat} values

of 1.39 - 14.7 min^{-1} determined here are in good agreement with T5FEN. The relationship between pH and T7 gp6 activity is explored in section 3.2.3.3.

Viral FENs have much greater exonuclease activity compared to eukaryotic FENs. For example, the catalytic efficiency of mouse FEN-1 and human FEN-1 with a nicked substrate is between 10 - 50-fold lower than that for a flap substrate (Zheng *et al.*, 2002; Liu *et al.*, 2006). In the case of T7 gp6, this may be because T7 DNA polymerase possesses 3' to 5' exonuclease activity for proofreading and therefore does not usually strand displace into 5' flaps (Nakai and Richardson, 1988). As flaps are not a main route of Okazaki fragment maturation in T7 gp6 DNA synthesis, FEN activity is not a priority of the enzyme.

3.3.5. Effect of dsDNA and overhang length on T7 gp6 activity

As previously mentioned, the original overhang substrate OHP was not suitable for determination of catalytic parameters of T7 gp6 due to a low K_M and slow rates of cleavage. FRET analysis shows that 25 nM is not completely cleaved by 1164 nM T7 gp6 in 400s whilst 0.8 nM T5FEN is sufficient for complete cleavage within the timeframe investigated (Figure 3-13). In an attempt to increase the catalytic efficiency of T7 gp6 for an overhang substrate to obtain catalytic parameters, several OHP substrate variants were designed, which differed in the length of dsDNA (10 nt - 20 nt) or 5' flap (4 nt or 7 nt). Investigation of substrate turnover at a single enzyme and substrate concentration shows that length of dsDNA significantly effects T7 gp6 endonuclease activity (Figure 3-14). The initial rate of the reaction of cleavage of OHP10, with a 20 nt dsDNA region is over 800-fold higher compared to OHP, which is composed of a 10 nt dsDNA region. Overhang length has less of an effect as the initial rate of cleavage of a substrate with a 7 nt overhang is not significantly different to a 4 nt overhang (Figure 3-14). However, this does not rule out an effect by longer flap lengths and a dependency in the overhang length has been observed in previous studies of T7 gp6 (Mitsunobu *et al.*, 2014) and T5FEN (Williams *et al.*, 2007). In contrast, 5' flap length has less of an effect on eukaryotic FEN activity in flap substrates. In the yeast *S. cerevisiae* Rad27, similar activity was observed with double-flap substrates with 5' flaps of 5, 20 and 40 nt (Singh *et al.*, 2007). A similar pattern was also observed with human FEN-1 where various flap lengths were investigated and only a substrate composed of 54 nt flap gave pronounced inhibition compared to substrates with shorter flaps (Shin *et al.*, 2012).

3.3.6. Optimizing reaction conditions to enhance T7 gp6 activity

Once the biochemical properties of the T7 gp6 nuclease activities were characterized, the next aim was to determine the effect of reaction conditions on activity in order to define optimum conditions for enhanced activity. For this, the UV assay and FRET assay was used to assess both exonuclease and endonuclease activity.

In the UV exonuclease assay, the monovalent cations K^+ and Na^+ were investigated by varying the concentrations of the salts KCl and NaCl in glycinate or Tris buffers (Figure 3-15). The glycine buffer had the pH adjusted with the hydroxide containing the counter ion (i.e. KCl dependency was investigated in sodium glycinate and vice versa). This enabled comparison of the importance of both cation identity and ionic strength. The specific activity of T7 gp6 displayed a bell-shaped dependency on KCl concentration. The optimum concentration of KCl in glycinate buffer was 25 mM, which enhanced activity approximately 1.3-fold compared to the 0 mM control, although this was not significant. Increasing KCl concentrations above this resulted in a downward trend in specific activity, although the only significant effect ($p < 0.05$) was observed with 300 mM KCl, which resulted in approximately 0.5-fold decrease in specific activity ($p<0.01$). In Tris buffer, the effect of KCl was more statistically significant. The optimum concentration was 100 mM which results in approximately a 2-fold increase compared to the 0 mM control ($p<0.0001$). In glycinate buffer, the optimum NaCl concentration is 0 mM and T7 gp6 is tolerant up to 25 mM, with no significant change in activity, however, above this, there is a downward trend in specific activity with around 0.4-fold decrease at 100 mM NaCl ($p<0.0001$). In Tris buffer, there is no significant change in specific activity with concentrations of NaCl up to 100 mM, however, more repeats may increase the statistical power this.

The FRET assay with the single-flap substrate was also used to investigate the monovalent cations K^+ and Na^+ on T7 gp6 activity (Figure 3-17). The dependency in Tris and HEPES buffers was investigated. A bell-shaped dependency on the initial rate of the reaction on KCl concentration was observed in this assay, similar to the effect observed in the UV assay. In HEPES buffer, the optimum KCl concentration is 30 mM which enhances the initial rate of reaction by over 8-fold compared to the 0 mM control ($p<0.0001$). In Tris buffer, the optimum KCl concentration is 25 mM, which elicits nearly a 14-fold increase in the initial rate of reaction

compared to the 0 mM control. Whilst increasing concentrations of NaCl in HEPES buffer results in a decrease in the initial rate of reaction compared to 0 mM, which is statistically significant for concentrations of 50 mM and above. In Tris buffer, 10 mM NaCl does increase the rate of reaction compared to when no NaCl is present, but this is not statistically significant. However, there are high errors in the 0 mM control, so increasing the number of repeats may increase the statistical power of any dependency present.

These results from both the UV and FRET assay indicate that both cation identity and ionic strength impact on T7 gp6 activity. Monovalent cations also appear to result in different dependencies between endonuclease and exonuclease reaction. However, it is important to note that the catalytic conditions between both assays used vary significantly. The UV assay is carried out at an initial velocity (V_0) that is approaching maximum velocity (V_{max}) as using either half or double the concentration of DNA does not significantly change the rate of reaction (data not shown), a feature also observed with T5FEN (Garforth and Sayers, 1997). However, the FRET assay is run at 200 nM, which is approaching K_M (104.2 ± 5.7 nM in the conditions used for K_M determination). This means that different information is drawn from both assays and indicates that monovalent cations impact catalytic parameters by affecting both the k_{cat} and K_M.

Monovalent dependencies are comparable in other FENs (Sayers and Eckstein, 1990; Lyamichev *et al.*, 1993; Harrington and Lieber, 1994; Zhao *et al.*, 2014). Addition of KCl has been shown to increase binding of T5FEN to DNA (Garforth *et al.*, 2001) and in *E. coli* ExoIX, a FEN-family member (Anstey-Gilbert *et al.*, 2013). This could explain why small amounts of KCl increase the rate of T7 gp6 reactions, due to a more stable enzyme DNA complex. Binding of K^+ ion is mediated, at least in part through the H2/3TH motif, which binds to and interacts with the phosphodiester backbone of downstream dsDNA (Hosfield *et al.*, 1998; Tsutakawa *et al.*, 2011). Furthermore, substrate concentration has been shown to shift the KCl optimum of human FEN-1 (Finger *et al.*, 2009). Therefore, it is also likely that at a high concentration, monovalent cations also interact with DNA, explaining why high concentrations of KCl and NaCl have an inhibitory effect.

The pH has a drastic effect on T7 gp6-catalysed reactions (Figure 3-16). In the UV assay, the specific activity approximately doubles with each increasing pH until the optimum pH of 9.3, before sharply decreasing. The optimum pH of several FENs have been determined and are similar to T7 gp6, also favouring slightly alkaline conditions. For example the optimum of mouse FEN-1 is pH 8, Taq polymerase FEN domain is pH 9 and T5FEN is pH 9.3 (Paul and Lehman, 1966; Lyamichev *et al.*, 1993; Harrington and Lieber, 1994). Lower pH favours endonuclease cleavage of T5FEN (Garforth *et al.*, 1999) and it has also been shown to fix the position of the endonuclease site of cleavage (Williams *et al.*, 2007). It would be interesting to examine the pH dependency of T7 gp6 endonuclease reaction. However, this was not possible in the FRET assay used due to lack of stability of the fluorophores used over a wide range of pH values. Alternatively, a gel-based assay could be used to investigate this, which has the benefit of visualisation of cleavage products.

T7 gp6 activity is stimulated by Mg^{2+}, Mn^{2+} and slightly by Co^{2+} (Table 3-1). T7 gp6 is more tolerant of a variety of concentrations of Mg^{2+} compared to Mn^{2+}, with the best activity at 2 mM $MgCl_2$. An early paper which measured T7 gp6 exonuclease cleavage of *E. coli* DNA found that 1 mM $MnCl_2$ enhanced activity by 50% compared to 5 mM $MgCl_2$. (Kerr and Sadowski, 1972). However, the results obtained here show that $MgCl_2$ is more excitatory compared to $MnCl_2$ and 1 mM $MnCl_2$ gives less than half the specific activity of 5 mM $MgCl_2$. These differences may be due to *E. coli* DNA being a larger substrate and could be related to the processivity of the enzyme. Furthermore, the protein purified in this early paper had higher levels of contamination with nucleases, so may represent activity from contamination (Kerr and Sadowski, 1972). The inhibition by divalent metal ions have been noted previously for FENs (Zheng *et al.*, 2002; Tock *et al.*, 2003; Harrington and Lieber, 1994) and other nucleases (Groll *et al.*, 1997) and this has also been observed here. However, the exact mechanism of this inhibition is not known and may result from metal interactions with DNA due to an increased ionic strength, or due to low affinity binding sites which have inhibitory effect. Furthermore, Mn^{2+} may be better at low concentrations and Mg^{2+} at higher concentrations due to the fact that Mn^{2+} binds to the active site of FENs more tightly, a feature observed in T5FEN (Feng *et al.*, 2004). In mouse FEN-1, >20 mM $MgCl_2$ has been shown to induce conformation change of protein structure and also inhibit substrate binding (Zheng *et al.*,

2002). Therefore, high concentrations of divalent metal ions may inhibit T7 gp6 activity by modulation of protein conformation and substrate binding capability.

3.3.7. DNA binding of T7 gp6

The binding of T7 gp6 to DNA was investigated by EMSA with single-flap, double-flap, overhang (OHP4) and nicked substrates (Figure 3-18). The reactions were carried out in the presence of 5 mM $CaCl_2$ which has shown to be non-permissive to nuclease activity to T7 gp6 (Table 3-2) without disrupting binding, and has been shown to increase the binding affinity of T5FEN, suggesting the shielding of negatively charged active site residues (Feng *et al.*, 2004).

This assay required optimization to produce results which could be quantified to yield a K_D. This included gel percentage and buffer components of both the gel and running buffer such as additives (glycerol, KCl), and concentration and nature of the buffer itself. After optimization, gels appear to indicate one binding state, due to one shifted band. The results show that T7 gp6 binds single-flap, double-flap and nicked substrates with a similar affinity, whilst the overhang is bound with up to 7-fold lower affinity in comparison (Figure 3-18). The results from EMSA for OHP4 are in agreement with the catalytic parameters obtained by the FRET assay and indicate that the low catalytic efficiency for this substrate is due to a decreased binding affinity, suggesting that downstream DNA interactions are important.

Chapter 4 – Mutational studies on T7 gp6

4.1. Introduction

Amino acid sequence comparison between FEN homologues reveal a number of conserved residues (Figure 4-1). These contribute to important functional domains within the protein, as well as conserved active site residues. Mutational studies have been used to help elucidate the function of these residues and we are interested in whether analogous mutations in the context of the T7 gp6 would have similar consequences.

4.1.1. Active site

The FEN active site is a conserved motif composed of 7 - 8 carboxylate residues which co-ordinate the catalytic metal ions. Two or three divalent metal ions have been implicated in FEN activity and are bound to the active site through two metal-ion binding sites, site I and site II (Feng *et al.*, 2004; Syson *et al.*, 2008). Mutation of specific FEN active site residues has been used to generate catalytically inert proteins which can bind but not cleave DNA in the presence divalent ion co-factors (Bhagwat *et al.*, 1997; Feng *et al.*, 2004; Zhang, 2012). This approach has been successfully exploited for co-crystallization of FENs with DNA. For example, the T4 RNase H Asp132Asn active site mutant (equivalent to Asp136 in T7 gp6) has been co-crystallized with pseudo-Y DNA (Devos *et al.*, 2007). Furthermore, two T5FEN active site mutants, Asp153Lys and Asp155Lys (equivalent to Asp160 and Asp162 in T7 gp6, respectively) have also been co-crystallized with DNA, the former also bound to catalytically relevant Mg^{2+} ions (AlMalki *et al.*, 2016).

As can be seen in the multiple sequence alignment (Figure 4-1), other residues around the active site are conserved between bacteriophage and bacterial FENs, including residues analogous to Tyr99 and Lys100 in T7 gp6. The role of these residues in homologous FENs has been studied in the literature. In T5FEN, mutations in the equivalent residues Tyr82Phe or Lys83Ala result in defective exonuclease but not endonuclease activity and an impaired DNA binding affinity (Garforth *et al.*, 1999; Patel *et al.*, 2002). Whilst in T4 RNase H, the Tyr86Phe mutant exhibits a small reduction in exonuclease activity and no apparent DNA-binding

defect, whilst Lys87Ala mutant exhibits a loss of nuclease activity and a significant DNA-binding defect compared to the WT enzyme (Bhagwat *et al.*, 1997).

Figure 4-1: Multiple sequence alignment of FEN sequences

Amino acid alignment of representative FENs from viruses, prokaryotes and eukaryotes. Highly conserved residues (identical in at least 7 prokaryotic or 12 total sequences) are highlighted in red; moderately conserved residues (identical in at least 6 prokaryotic or 10 total sequences) are highlighted in green; partially conserved residues (identical in at least 7 sequences) are highlighted in grey. The N-terminal conserved region is highlighted in yellow and the internal conserved region is highlighted in blue. Figure reproduced with permission from (Mueser *et al.*, 1996).

4.1.2. Helix-3-turn-helix motif

The helix-3-turn-helix (H3TH) domain is a K^+ ion binding motif which is involved in interactions with downstream dsDNA (Tsutakawa *et al.*, 2011). In a study of T5FEN, the functional importance of re-introducing and chemically modifying cysteine residues at distinct positions on a cysteine-free version of the protein (T5FEN 2CA) was investigated (Zhang, 2012). The residue Leu202 is found on the H3TH motif and Leu202Cys mutation (T5FEN 2CA Leu202Cys) resulted in a protein which retained comparable levels of nucleolytic activity to the WT protein. However, chemical modification of this residue with the thiol-modifying Ellman's

reagent DTNB (5,5'-dithiobis(2-nitrobenzoic acid)) generated a protein which possessed severely compromised endonuclease activity, but retained exonuclease cleavage capabilities comparable to the unmodified protein (Zhang, 2012).

4.1.3. Mutational studies on T7 gp6

In this work, mutations of residues at and around the active site of T7 gp6 were studied. Two classes of mutants were generated: active-site mutants and H3TH motif mutants. We are interested in whether analogous mutations in the context of the T7 gp6 would have similar consequences to those described above (section 4.1.2). Structural alignment of T7 gp6 with T5FEN is shown in Figure 4-2.

An aim of this work was to generate T7 gp6 active-site mutants that are catalytically inert in the presence of Mg^{2+} but retain DNA binding. The resulting proteins were designed to be used in co-crystallization trials with DNA substrates in the presence of catalytically relevant divalent metal ions (Chapter 6), as there are no published structures of T7 gp6 either with or without DNA. In the T5FEN: DNA structure, a conserved negatively charged aspartic acid was mutated to a positively charged lysine (AlMalki *et al.*, 2016). This approach was utilized and three T7 gp6 mutants were generated: Asp160Lys, Asp162Lys, Asp202Lys. The nuclease activity and DNA-binding ability of each of these was assessed using the UV assay, FRET assay and EMSA.

Another aim of this work was to characterize mutations in the H3TH motif of T7 gp6. This is of interest as studies in T5FEN indicate that modification of residues here can alter the ratio between exonuclease and endonuclease activity, as described in section 4.1.2 (Zhang, 2012). It was hypothesized that enzymes with altered ratios of endonuclease and exonuclease activity could be engineered through mutagenesis of residues in and around the active site. Enzymes with altered selectivity might prove useful in specific applications. This may arise due to a reduction in off-target cleavage events. For example, in Gibson cloning, endonuclease activity may result in degradation of transient overhangs and flaps that develop or cleavage may occur at the wrong sites, so selective loss of endonuclease activity may enhance the efficiency of the technique.

Previous work in the laboratory engineered a cysteine-free T7 gp6 as a scaffold for the re-introduction of chemically modified cysteines (T7 gp6 8CA). During this project, this protein was overexpressed but was completely insoluble, although successfully solubilized and re-folded. However, kinetic analysis indicated that this protein possessed drastically reduced nuclease activity with a temperature sensitive phenotype (data not shown). Therefore, amino acid modification was investigated instead. Two T7 gp6 mutants were generated: Ile200Arg and Thr201Trp. Additionally, to investigate FEN-specific and amino acid vs chemical modifications, the corresponding T5FEN mutants were generated: Leu202Arg and Gly203Trp. The nuclease activity and DNA binding ability of each of these was assessed using the UV assay, FRET assay and EMSA.

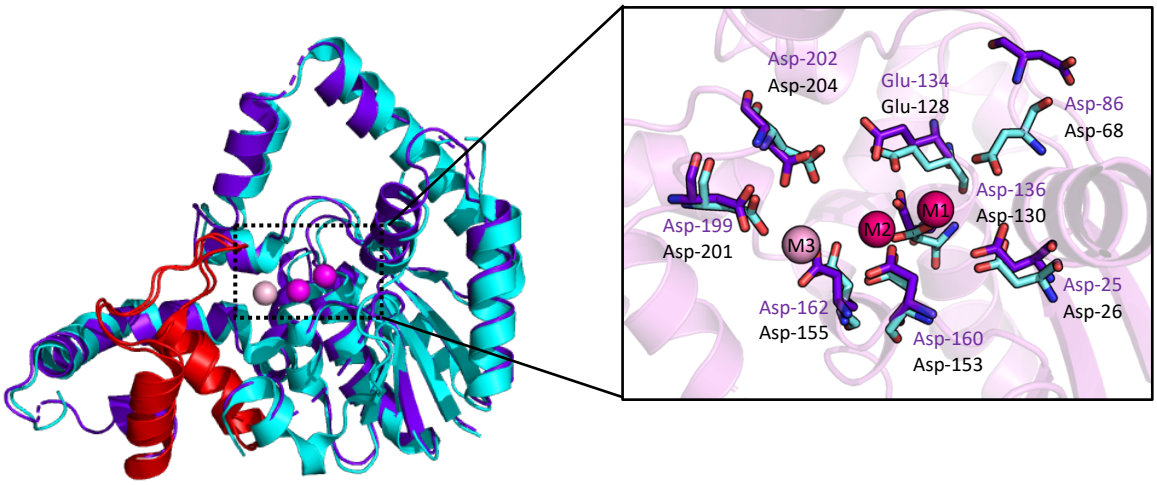


Figure 4-2: Structural alignment of T7 gp6 with T5FEN

The predicted structure of T7 gp6 (purple) was determined using Phyre2 and was aligned to the structure of T5FEN (PDB: 5HMM; cyan) bound to three Mg^{2+} ions (M1, M2; magenta sphere; M3, light pink sphere). The conserved helix-3-turn-helix motif in both structures is coloured in red. Structural alignment of the conserved active site carboxylate residues is shown.

4.2. Results

4.2.1. Cloning, expression and purification of mutant proteins

Primers to introduce point substitutions were designed using NEBaseChanger. The sequences are shown in Table 2-5. The mutations were introduced using the PCR-based Q5 site-directed mutagenesis (New England Biolabs) which utilizes non-overlapping primers and introduces the mutation in the forward primer. The WT pJONEX4:T7 gp6 or pJONEX4:T5FENΔ19 was used

as the template. The T5FEN∆19 lacks the first 19 amino acids and previous studies have shown that this mutation does not affect the structure, binding or cleavage, whilst enhancing the stability and crystallization of the protein (Ceska *et al.*, 1996; Garforth and Sayers, 1997) and therefore, will be referred to as WT for the rest of this work. Example PCR products obtained are shown in Figure 4-3, with the single band indicating the PCR was a success. After treating with the KLD enzyme mix, products were transformed into the *E. coli* cloning strain XL1blue. The presence of the desired mutation in the plasmid was confirmed by extracting and sequencing plasmids from several clones.

Over-expression of target proteins was carried out by heat-shock induction in ∆H1+ *E. coli,* as previously described in section 2.3. After induction of protein expression, T7 gp6 Ile200Arg and T7 gp6 Thr201Trp were harvested immediately whilst for all other proteins, cells were incubated at 18°C overnight to allow cell biomass to increase and protein to accumulate. SDS-PAGE analysis of expression of proteins is shown in Figure 4-4. All proteins were estimated to be over 60% soluble and purified from the soluble fraction using the same methods described for WT T7 gp6, to achieve 90-95% purity, as determined by SDS-PAGE. The approximate yield of protein per gram of cell pellet is summarised in Table 4-1.

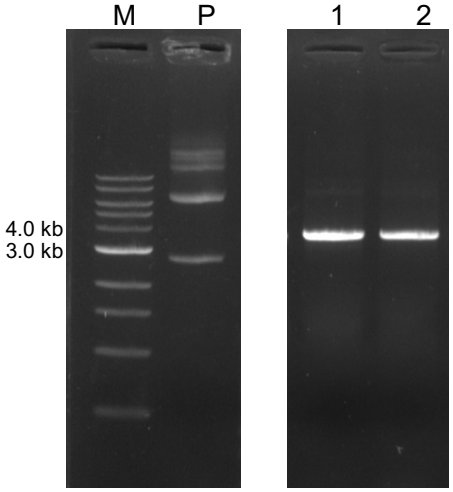


Figure 4-3: Site-directed mutagenesis of T7 gp6

The *pJONEX:T7 gp6* plasmid (lane P, plasmid) was used as a substrate for Q5 site-directed mutagenesis. The PCR products for two mutants; Asp160Lys (lane 1), and Asp162Lys (lane 2) are shown. Lane M, NEB 1 kb DNA ladder (0.5 – 10 kb range). Results were analysed on a 1% agarose gel stained with Midori Green Advance DNA stain.

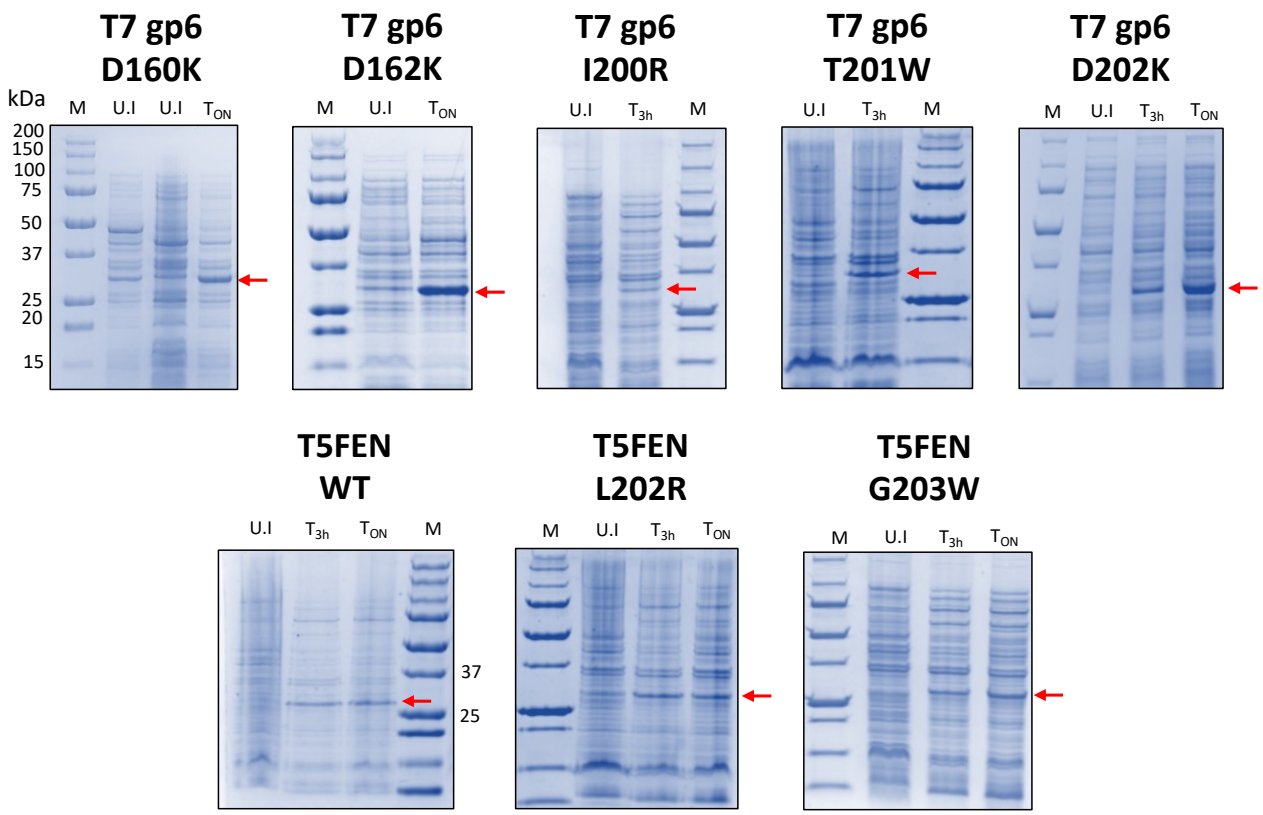


Figure 4-4: Overexpression of T7 gp6 and T5FEN proteins

SDS-PAGE analysis of protein overexpression. Samples were taken before (lane U.I.) and after 3 hours (lane T_{3h}) 42°C heat-shock induction. Some proteins were incubated overnight at 20°C after induction (lane T_{ON}). The red arrow indicates the position of target protein expression. Samples were run on 10% acrylamide gels with PrecisionPlus marker (lane M).

Table 4-1:Yield of FEN proteins obtained from cell pellets

Protein	Approximate yield per g cell pellet (mg)
T7 gp6 wild-type	0.5
T7 gp6 Asp160Lys	8.0
T7 gp6 Asp162Lys	10.0
T7 gp6 Asp202Lys	5.0
T7 gp6 Ile200Arg	0.4
T7 gp6 Thr201Trp	6.3
T5FEN WT	2.0
T5FEN Leu202Arg	1.1
T5FEN Gly203Trp	3.5

4.2.2. Characterisation of active site mutants

4.2.2.1. Activity analysis of active site mutants

Proteins were analysed by zymography after purification and the same amount of each protein was loaded into the lanes (Figure 4-5). No obvious band was present for Asp160Lys and Asp162Lys, indicating that they do not possess nuclease activity whilst Asp202Lys has a faint band, indicating low levels of activity. Furthermore, there are no contaminating bands, suggesting the protein preparations are sufficiently pure enough.

The UV exonuclease assay was used to quantify exonuclease activity of the mutants (Figure 4-6). These results show that no acid-soluble nucleotides are released by 50 μg of either Asp160Lys or Asp162Lys variants. Whilst 50 μg Asp202Lys did result in measurable DNA hydrolysis, its specific activity was 3.05 ± 0.12 $U.\mu g^{-1}$, over 500-fold lower than the wild type protein.

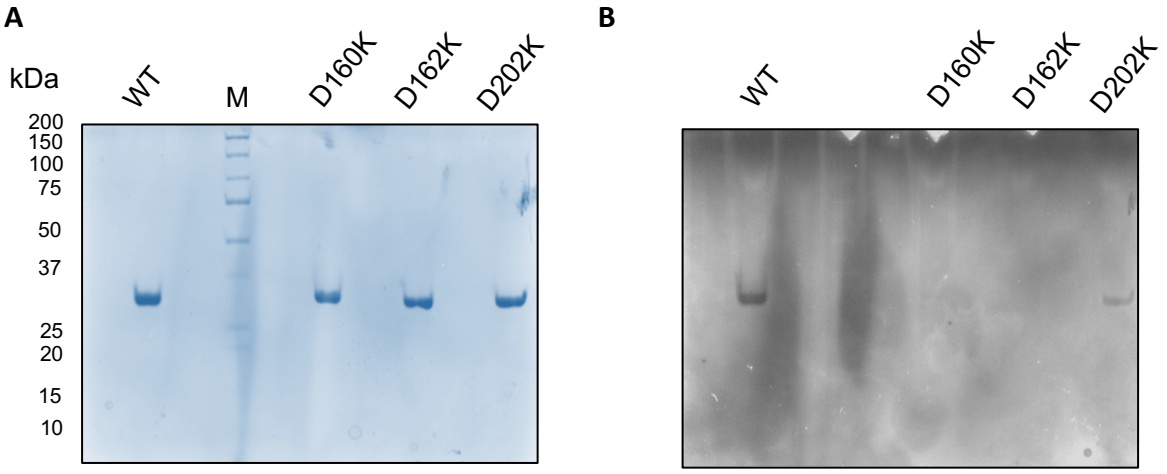


Figure 4-5: SDS-PAGE and zymogram analysis of T7 gp6 active site mutants

Purified T7 gp6 proteins (4 μg) were analysed by (A) SDS-PAGE and (B) zymogram. The zymogram gel was incubated in reaction buffer for 15 minutes and the same gel was subsequently stained with Coomassie dye. Analysis was performed on 10% acrylamide gels with PrecisionPlus marker (BioRad) (lane M) used.

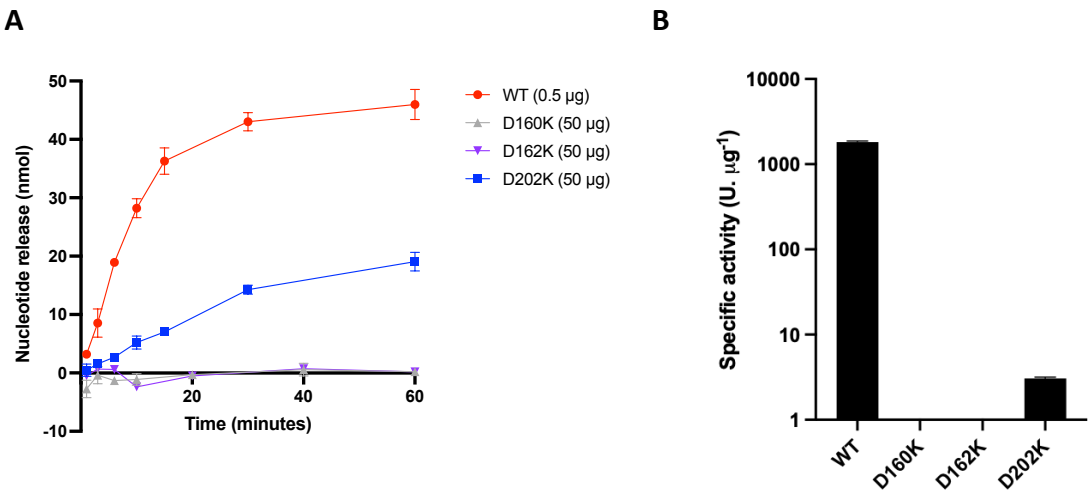


Figure 4-6:UV assay analysis of T7 gp6 active site mutants

The UV spectrophotometric assay was used to determine the exonuclease activity of T7 gp6 active site mutants. The release of acid soluble nucleotides from high-molecular-weight DNA by the addition of 0.5 µg WT T7 gp6 or 50 µg mutant was measured. The reaction mixture contained 667 µg.mL^{-1} type XIV DNA, 25 mM potassium glycinate pH 9.3, 50 mM KCl, 10 mM MgCl2, 1 mM DTT (A) Time-course of the reaction. (B) Specific activity. The rate of release was determined graphically by fitting the initial rate section with a linear regression in GraphPad Prism. The mean of three repeats is shown and error bars represent SEM. Statistical significance was determined by a one-way ANOVA with Tukey's multiple comparison. Compared to WT, all mutants ****<p0.0001.

The FRET assay was also used to monitor activity of the mutants with dual-labelled substrates in real-time (Figure 4-7). Progress curves were collected using 100 nM single-flap, nicked, overhang (OHP4) and dsDNA substrates with various amounts of active site mutants and WT T7 gp6 as a control, as indicated in the graph legends. A slight change in fluorescence is detected for Asp162Lys and Asp202Lys for the nicked, overhang and dsDNA substrates, which could indicate a low level of activity. However, the active site mutants are used in a 20 – 600-fold excess compared to the WT control, with Asp162Lys in 3-fold excess compared to Asp202Lys.

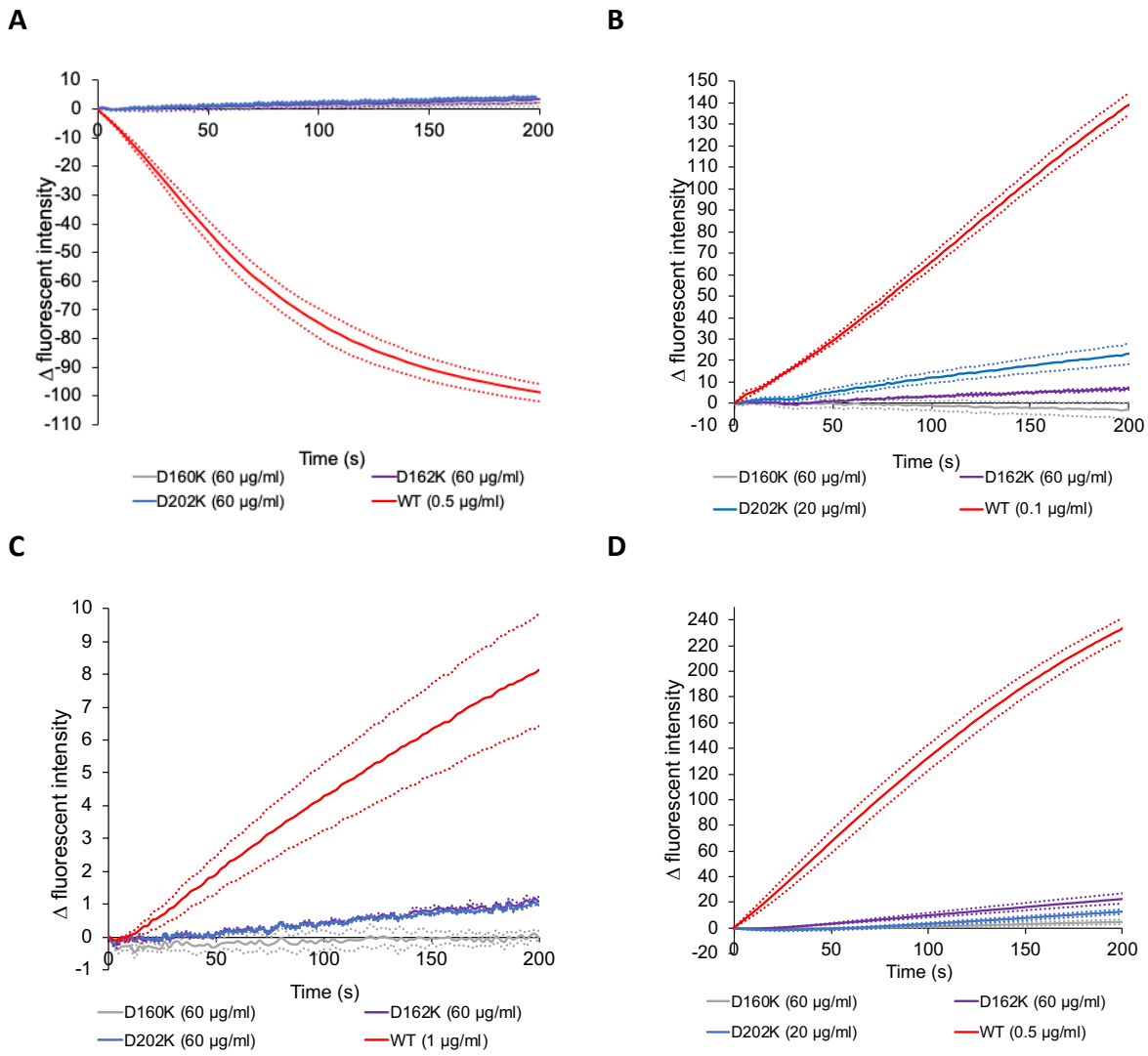


Figure 4-7: FRET assay analysis of T7 gp6 active site mutants

The nuclease activity of T7 gp6 DK mutants (Asp160Lys, Asp162Lys and Asp202Lys) was assessed via the FRET and UV assays compared to WT. Progress curves were collected with 100 nM substrate and the amount of enzyme added is indicated in the individual graph legends. (A) Single-flap, (B) Nicked, (C) 5' overhang and (D) dsDNA substrates were investigated. The results are presented as the mean of three repeats and dashed lines represent SD.

4.2.2.2. Binding affinity of T7 gp6 active-site mutants

The binding affinity of the active site mutants with an endonuclease (single-flap) and exonuclease (nicked) substrates was determined by the EMSA (Figure 4-8). This assay was carried out in the presence of 5 mM $CaCl_2$ to inhibit possible low levels of cleavage and also provide identical binding conditions as the WT, to allow for direct comparison. Quantification of the gel images reveals that all mutants bind single-flap and nicked DNA substrates with comparable affinity to WT and statistical analysis shows no difference between these.

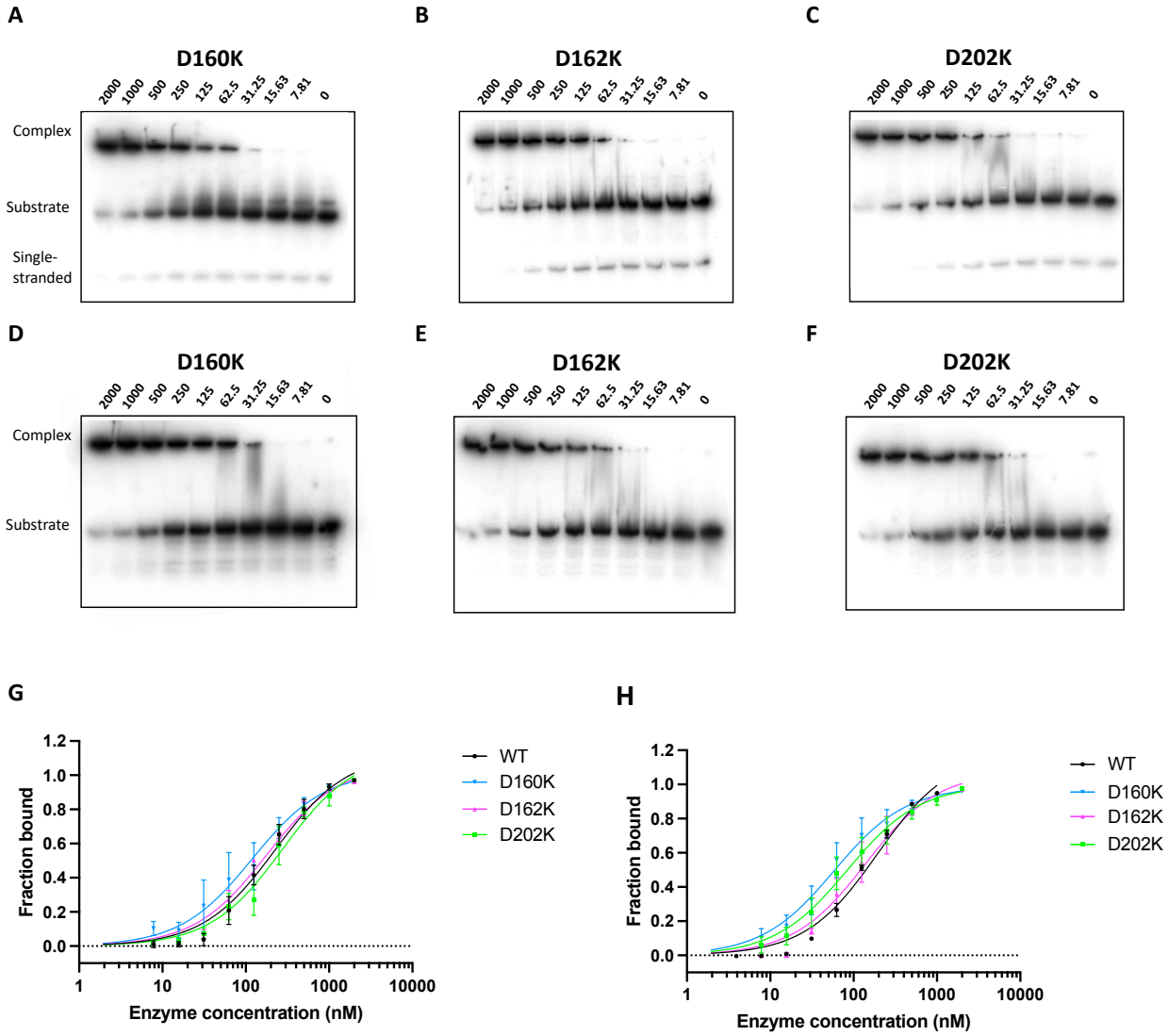


Figure 4-8: DNA binding studies of T7 gp6 active site mutants

Binding of single-flap and nicked substrates to T7 gp6 active site mutants (Asp160Lys, Asp162Lys, Asp202Lys) was examined using the electrophoretic mobility shift assay. Various concentrations of T7 gp6 mutants (indicated by lane numbers, nM) was incubated with 500 pM of ^{32}P-labelled substrates on ice for 1 hour before resolving enzyme-substrate complex from unbound substrate on a nondenaturing 17% acrylamide gel. (A-C) Single-flap substrate. (D-F) Nicked substrate. (G, single-flap; H, nicked) The gels were quantified for substrate bound at each enzyme concentration and fitted to the one-site -specific binding non-linear regression on GraphPad Prism 9. The mean of three repeats is shown and error bars represent SEM. Representative gel image given for each EMSA.

4.2.3. Characterisation of helix-3-turn-helix mutations

4.2.3.1. Activity analysis of helix-3-turn-helix mutants

Purified proteins were analysed by zymography (Figure 4-9). Analysis of the zymogram suggests that there is no contaminating nuclease activity. By qualitatively comparing band intensity there does appear to be a difference in nuclease activity of the mutants compared to WT and this was further investigated in quantitative assays.

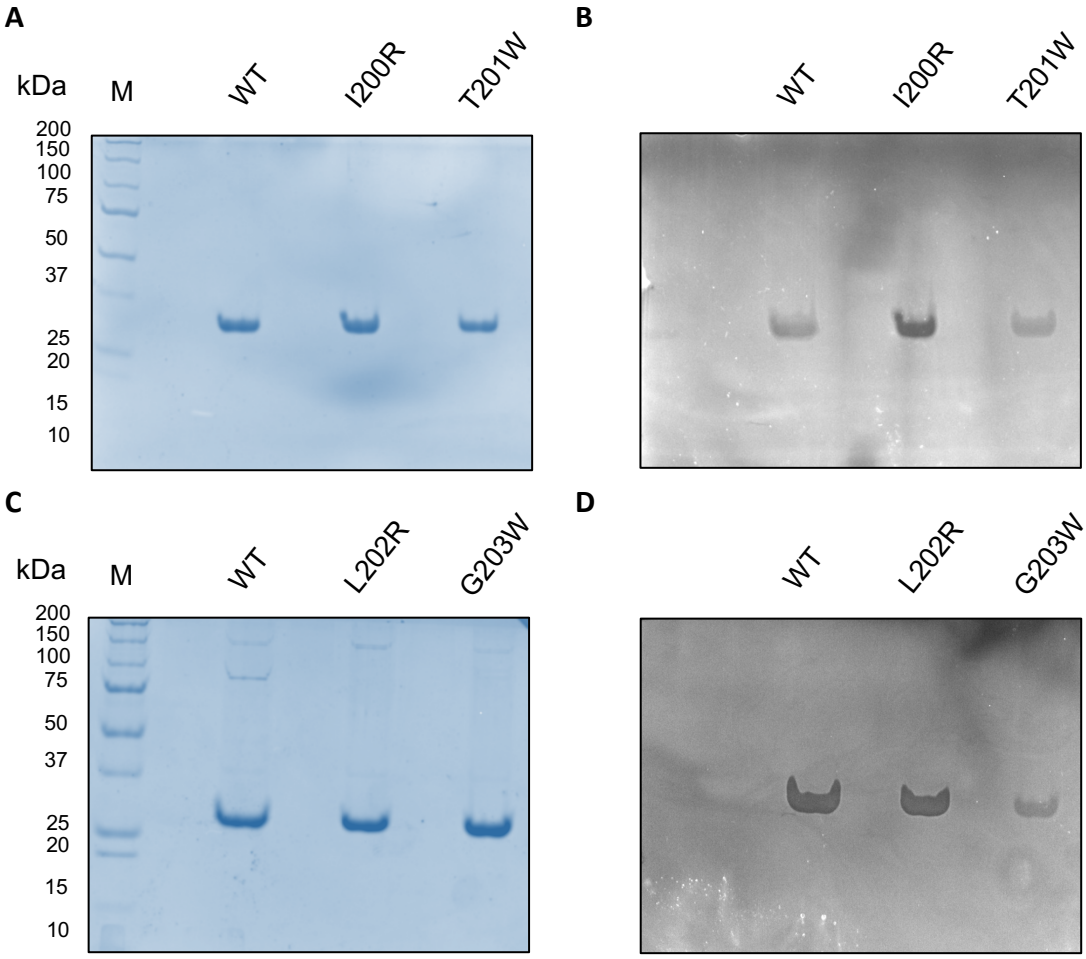


Figure 4-9: SDS-PAGE and zymogram analysis of exonuclease mutant proteins

Purified proteins (4 μg) were analysed by SDS-PAGE (A, T7 gp6 proteins; C, T5FEN proteins) and zymogram (B, T7 gp6 proteins; D, T5FEN proteins) The zymogram gel was incubated in reaction buffer for 15 minutes and the same gel was subsequently stained with Coomassie dye. Analysis was performed on 10% acrylamide gels with PrecisionPlus marker (BioRad) (lane M) used.

The UV assay was used to quantify exonuclease activity and determine the specific activity (Figure 4-10). To determine if the results were statistically significant, a one-way ANOVA with multiple comparisons was used to compare the specific activity to the WT of each protein. The results demonstrate that for T7 gp6, Ile200Arg ($p=0.99$) is statistically different, whilst Thr201Trp ($p=0.0033$) is not statistically different from the WT protein. Additionally, for T5FEN, the results show that and Leu202Arg ($p=0.89$) is statistically different and Gly203Trp ($p<0.0001$) is not statistically different from the WT counterpart.

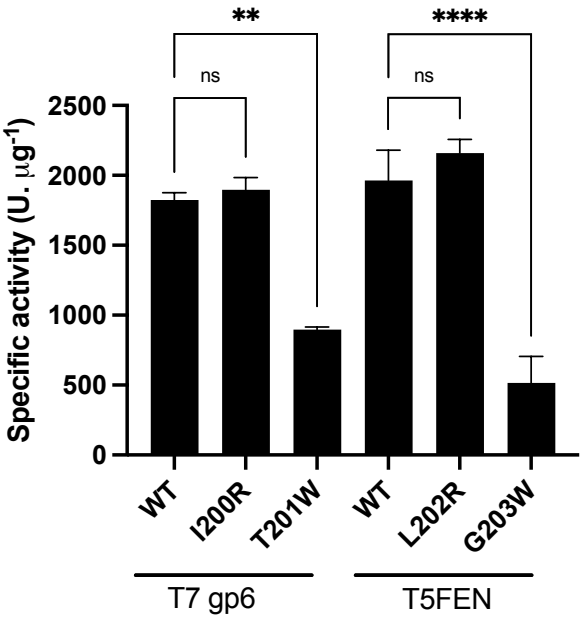


Figure 4-10: Specific activity of exonuclease mutant proteins

The UV spectrophotometric assay was used to determine the exonuclease activity of T7 gp6 and T5FEN proteins. This was calculated by measuring the cleavage of DNA by 0.5 μg of protein. The reaction mixture contained 667 $\mu g.mL^{-1}$ type XIV DNA, 25 mM potassium glycinate pH 9.3, 100 mM KCl, 10 mM $MgCl_2$, 1 mM DTT. The mean of three repeats is shown and error bars represent SEM. Statistical significance was measured by one-way ANOVA with Tukey's multiple comparisons, ns $p>0.05$, ** $p<0.01$, **** $p<0.0001$.

The nuclease activity of the mutants was further analysed and monitored in real-time with the FRET assay with the single-flap, overhang (OHP4), nicked and dsDNA substrates. The cleavage of 100 nM and 200 nM substrate by various concentrations of protein was measured (Figure 4-11 to Figure 4-12). By analysing the progress curves and UV assay data, it is clear that T7 gp6 Thr201Trp and T5FEN Gly203Trp exhibit a large decrease in both endonuclease and exonuclease activity, whilst T7 gp6 Ile200Arg and T5FEN Leu202Arg demonstrate exonuclease activity more comparable to the WT. Therefore, T7 gp6 Ile200Arg and T5FEN Leu202Arg were analysed further and catalytic parameters determined.

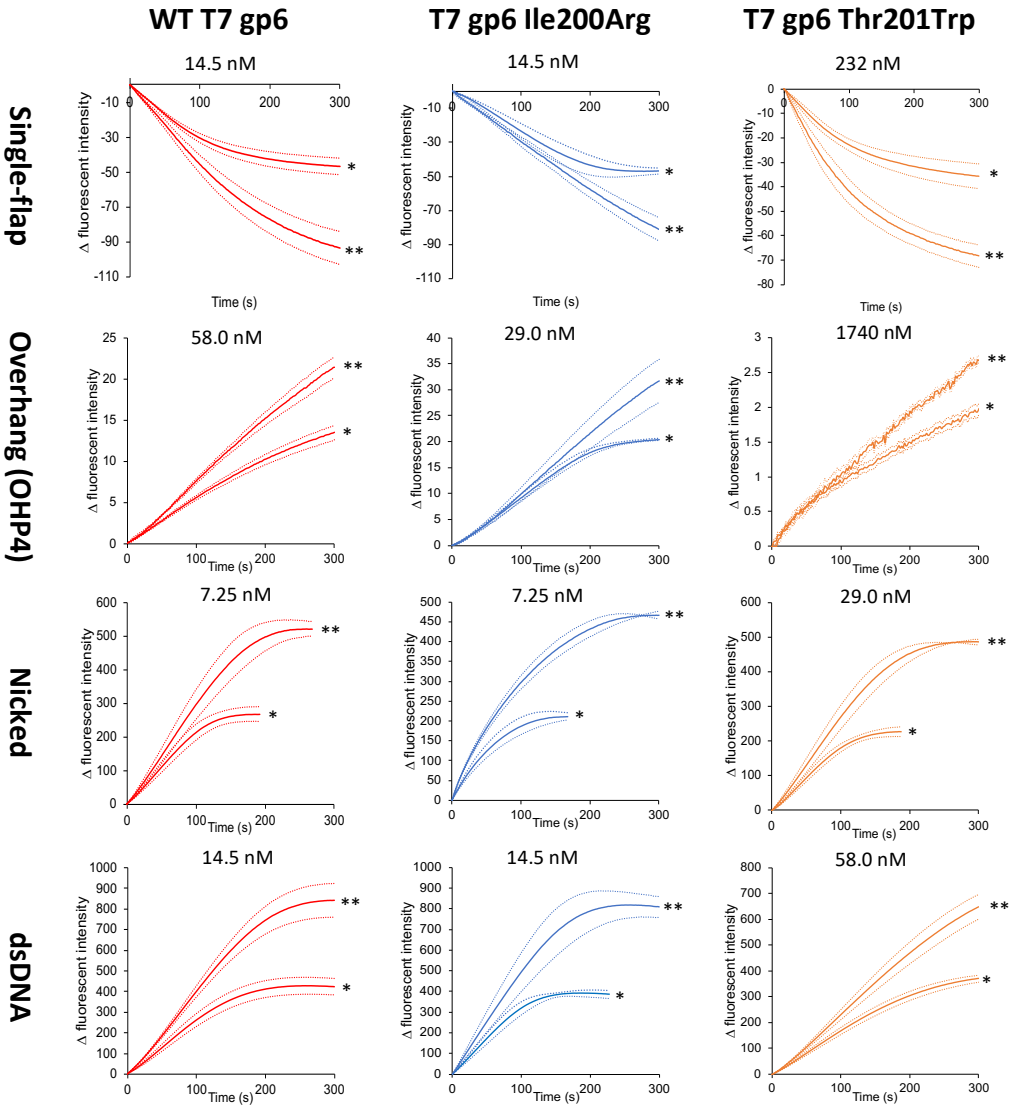


Figure 4-11: Progress curve analysis of T7 gp6 exonuclease mutants with FRET assay

The FRET assay was used to determine the nuclease activity of the T7 gp6 exonuclease mutants (Ile200Arg and Thr201Trp). Progress curves were collected with substrates (single-flap, overhang (OHP4), nicked and dsDNA) at 100 nM (*) and 200 nM (**). The concentration of enzyme added is indicated in the title of each graph. Progress curves were normalised against no enzyme controls. Results are presented as the mean of three repeats and dashed lines represent SD.

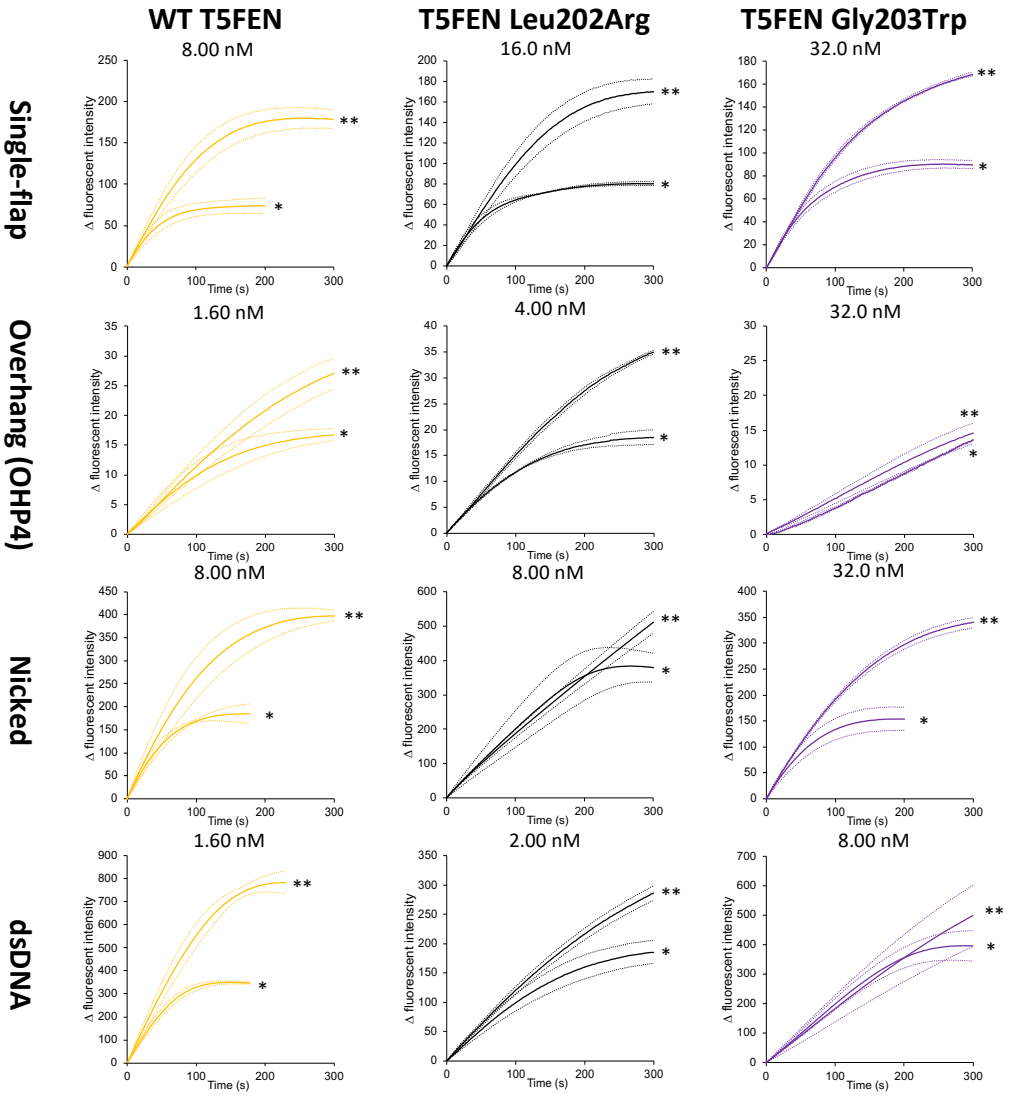


Figure 4-12: Progress curve analysis of T5FEN exonuclease mutants with FRET assay

The FRET assay was used to determine the nuclease activity of the T5FEN exonuclease mutants (Leu202Arg and Gly203Trp). Progress curves were collected with substrates (single-flap, overhang (OHP4), nicked and dsDNA) at 100 nM (*) and 200 nM (**). The concentration of enzyme added is indicated in the title of each graph. Progress curves were normalised against no enzyme controls. Results are presented as the mean of three repeats and dashed lines represent SD.

4.2.3.2. Determination of catalytic parameters of T7 gp6 Ile200Arg and T5FEN Leu202Arg

The FRET assay was carried out with single-flap, overhang, nicked and dsDNA substrates. The catalytic parameters for each substrate was determined as previously described using the Michaelis-Menten model and further visualised by Hanes-Woolf transformation (T7 gp6, Figure 4-13; T5FEN Leu202Arg, Figure 4-14). The quantified catalytic parameters are summarised in Table 4-2. These results show that for all substrates, the specificity constant (k_{cat}/K_M) of T7 gp6 Ile200Arg is increased compared to the WT. Whilst for T5FEN Leu202Arg, the specificity constant is decreased for all substrates compared to the WT.

4.2.3.3. Binding studies on T7 gp6 exonuclease mutants

The binding affinity of the T7 gp6 exonuclease mutants with single-flap and nicked substrates was determined using the EMSA in the presence of 5 mM $CaCl_2$ to prevent substrate cleavage (Figure 4-15). These results show that Ile200Arg has an increased affinity for the single-flap ($p=0.0158$, determined by a one-way ANOVA with Tukey's multiple comparisons) substrate and Thr201Trp has a decreased affinity. Whilst both Ile200Arg and Thr201Trp bind the nicked substrate with comparable affinity to the WT ($p>0.05$).

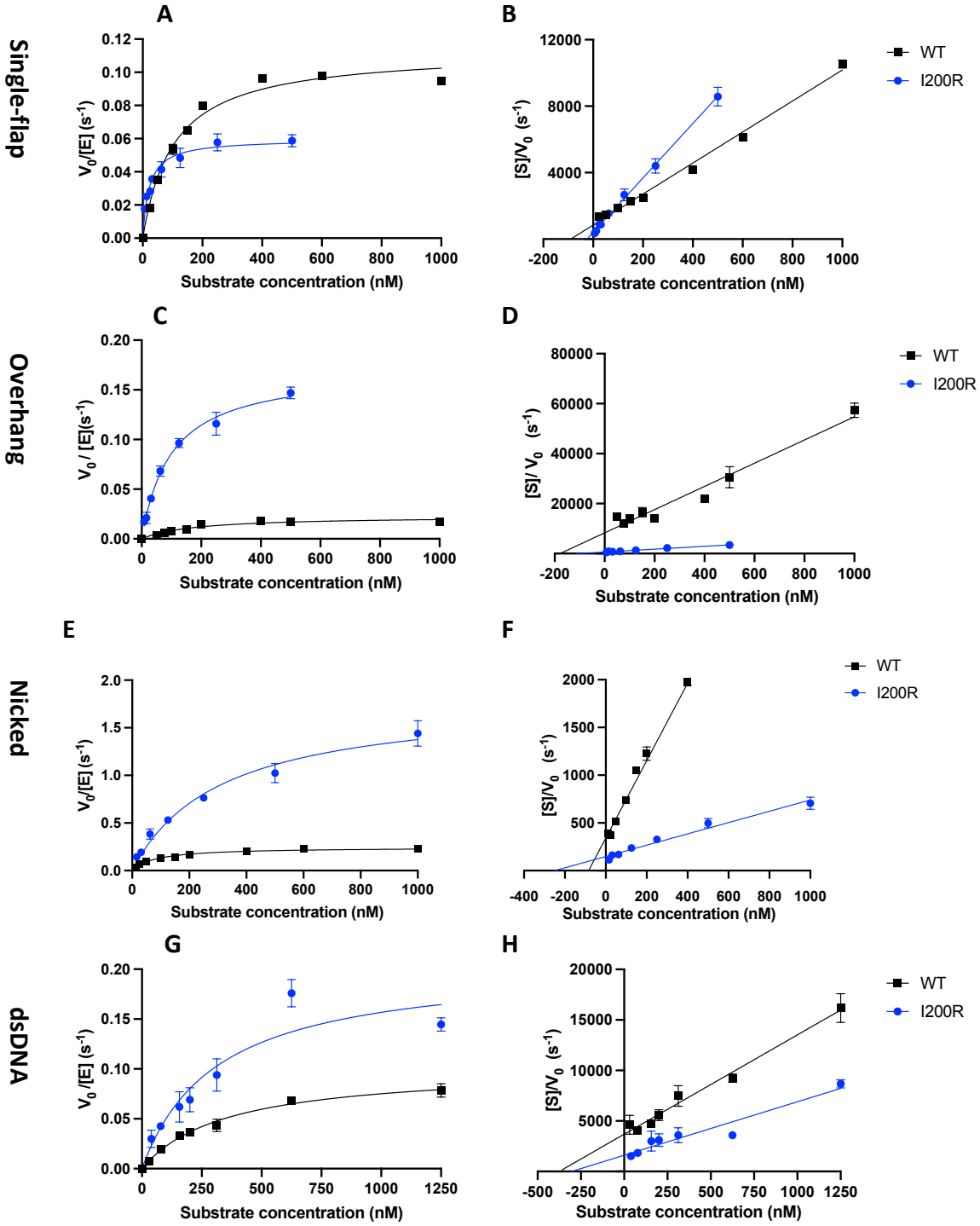


Figure 4-13: Michaelis-Menten analysis of T7 gp6 Ile200Arg

The FRET assay was used to determine catalytic parameters of T7 gp6 Ile200Arg (blue circles, lines) compared to WT T7 gp6 (black squares, lines) for various DNA substrates. Progress curves were obtained at various substrate concentrations with a constant enzyme concentration. The initial velocity (V_0) for each substrate concentration was calculated from the initial rate portion of the progress curves and normalised to the amount of enzyme added to the reaction. The results were plotted and analysed by non-linear regression of the Michaelis-Menten equation on GraphPad Prism (A, Single-flap; C, Overhang; E, Nicked; G, dsDNA). The data was transformed using

the Hanes-Woolf equation for linear visualisation (B, Single-flap; D, Overhang; F, Nicked; H, dsDNA). Data is presented as the mean of three repeats and error bars represent SEM.

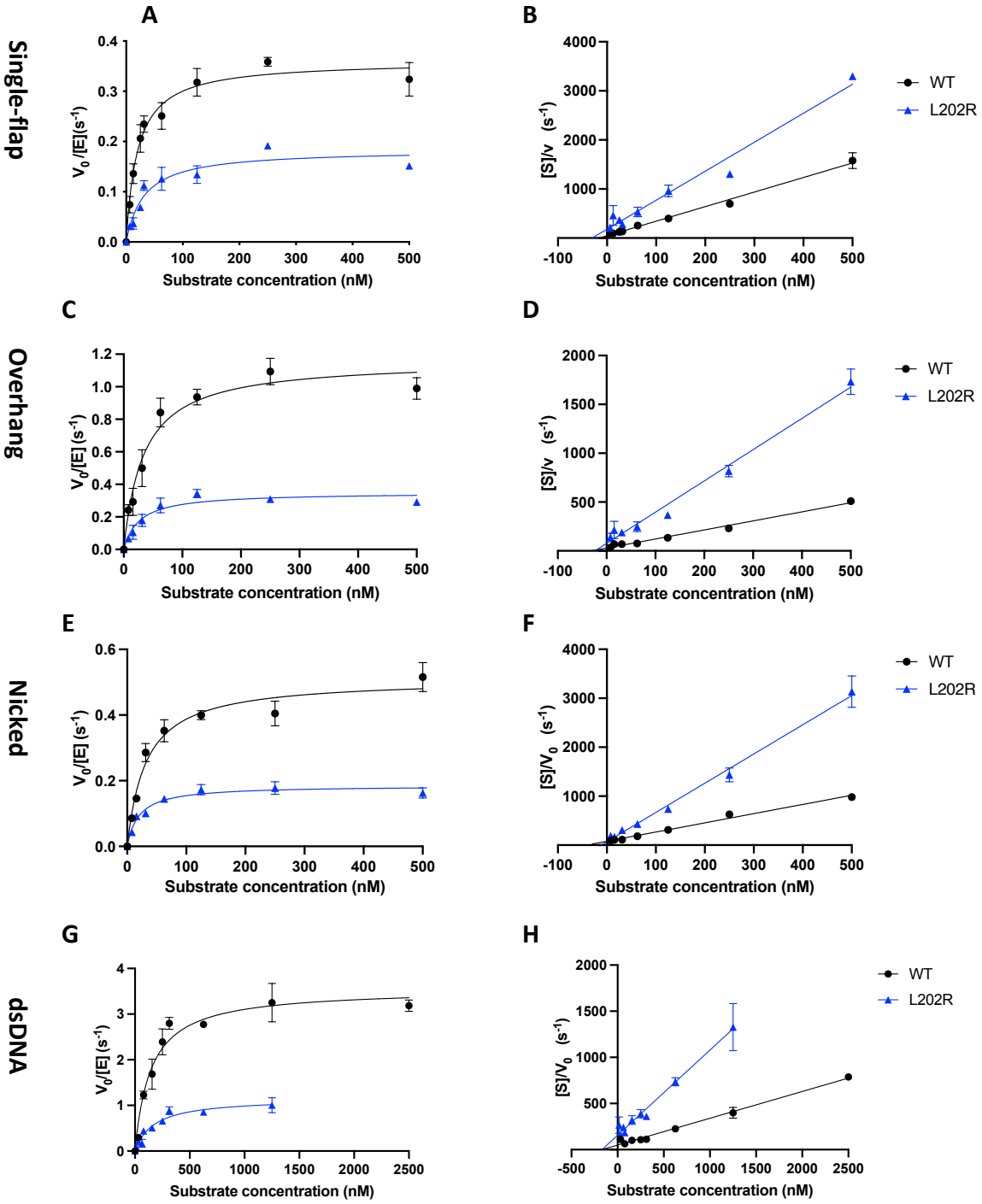


Figure 4-14: Michealis-Menten analysis of T5FEN Leu202Arg

The FRET assay was used to determine catalytic parameters of T5FEN Leu202Arg (blue triangle, lines) compared to WT T5FEN (black circles lines) for various DNA substrates. Progress curves were obtained at various substrate concentrations with a constant enzyme concentration. The initial velocity (V_0) for each substrate concentration was calculated from the initial rate portion of the progress curves and normalised to the amount of enzyme added to the reaction. The results were plotted and analysed by non-linear regression of the Michaelis-Menten equation on GraphPad Prism (A, Single-flap; C, Overhang; E, Nicked; G, dsDNA). The data was transformed using

the Hanes-Woolf equation for linear visualisation (B, Single-flap; D, Overhang; F, Nicked; H, dsDNA). Data is presented as the mean of three repeats and error bars represent SEM.

Table 4-2: Catalytic parameters of T7 gp6 and T5FEN proteins

	K_M (nM)	k_{cat} (min^{-1})	k_{cat}/K_M ($min^{-1}.nM^{-1}$)
Single-flap			
WT T7 gp6	104.2 ± 5.7	6.89 ± 0.10	0.0662 ± 0.00374
T7 gp6 Ile200Arg	22.1 ± 2.6	3.60 ± 0.183	0.162 ± 0.0208
WT T5FEN	21.6 ± 5.8	21.8 ± 0.93	1.01 ± 0.274
T5FEN Leu202Arg	32.9 ± 4.4	11.1 ± 0.16	0.336 ± 0.0450
Overhang			
WT T7 gp6	180.2 ± 33.1	1.39 ± 0.11	0.00768 ± 0.00154
T7 gp6 Ile200Arg	107.4 ± 21.4	10.5 ± 0.72	0.0977 ± 0.0206
WT T5FEN	39.2 ± 13.5	71.0 ± 4.27	1.81 ± 0.632
T5FEN Leu202Arg	32.1 ± 12.2	21.4 ± 1.00	0.667 ± 0.256
Nicked			
WT T7 gp6	83.4 ± 3.2	14.7 ± 0.34	0.176 ± 0.00783
T7 gp6 Ile200Arg	324. 2 ± 54.0	110.4 ± 13.7	0.340 ± 0.0707
WT T5FEN	31.7 ± 2.4	30.7 ± 2.4	0.968 ± 0.106
T5FEN Leu202Arg	20.5 ± 5.7	11.2 ± 1.2	0.545± 0.162
dsDNA			
WT T7 gp6	386.6 ± 75.6	6.30 ± 0.59	0.0163 ± 0.00353
T7 gp6 Ile200Arg	359.7 ± 112.7	12.8 ± 1.13	0.0356 ± 0.0116
WT T5FEN	158.8 ± 39.0	217.0 ± 14.7	1.37 ± 0.348
T5FEN Leu202Arg	182.8 ± 53.1	71.1 ± 11.6	0.389 ± 0.130

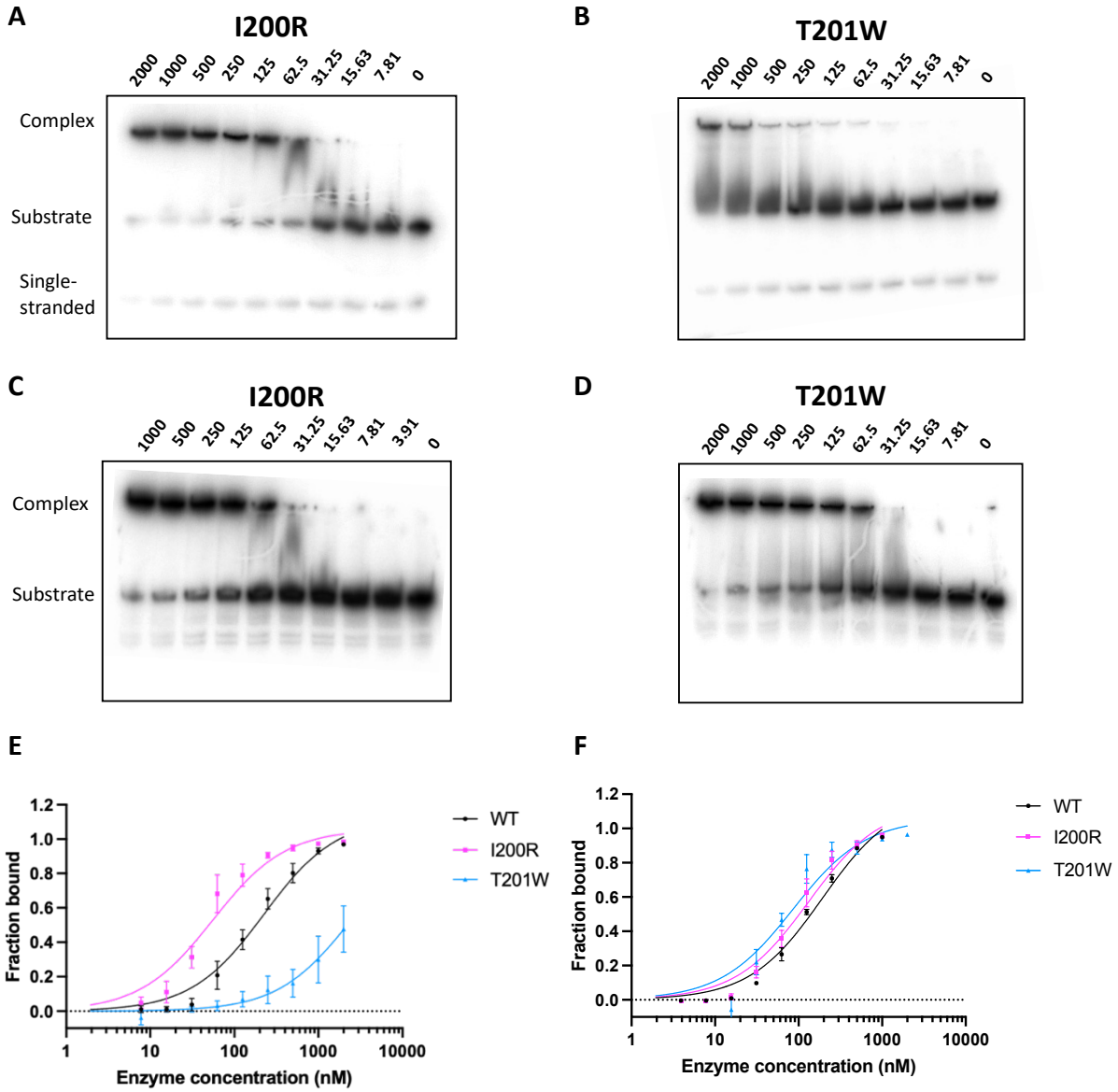


Figure 4-15: DNA binding studies of T7 gp6 exonuclease mutants

Binding of single-flap and nicked substrates to T7 gp6 exonuclease mutants (Ile200Arg, Thr201Trp) was examined using the electrophoretic mobility shift assay. Various concentrations of T7 gp6 protein (indicated by lane numbers, nM) was incubated with 500 pM of ^{32}P-labelled substrates on ice for 1 hour before resolving enzyme-substrate complex from unbound substrate on a nondenaturing 17% acrylamide gel. (A,B) Single-flap substrate. (C,D) Nicked substrate. (E, single-flap; F, nicked) The gels were quantified for substrate bound at each enzyme concentration and fitted to the one-site -specific binding non-linear regression on GraphPad Prism 9. Data is presented as the mean of three repeats and error bars represent SEM. Representative gel image given for each EMSA.

Table 4-3: Dissociation constant, K_D, of T7 gp6 mutants

	K_D (nM)	
Protein	**Single-flap**	**Nicked**
WT T7 gp6	243 ± 53.0	194 ± 11.1
T7 gp6 Asp160Lys	170 ± 77.7	75.1 ± 34.4
T7 gp6 Asp162Lys	177 ± 17.2	167 ± 52.6
T7 gp6 Asp202Lys	328 ± 121	99.5 ± 29.4
T7 gp6 Ile200Arg	60.9 ± 16.9	141 ± 29.1
T7 gp6 Thr201Trp	>2000	87.4 ± 11.9

4.3. Discussion

4.3.1. Cloning, over-expression and purification of mutant proteins

The results presented in this chapter show that the T7 gp6 and T5FEN mutant proteins were successfully cloned, overexpressed and purified. The yield per g of cell pellet varied between proteins and this was probably due to differences in expression levels and solubility. These differences between expression levels may be due to varying levels of toxicity of the mutant proteins to the *E. coli* as discussed in section 3.6.1. Indeed, the inactive mutants displayed higher levels of expression compared to the active proteins (Figure 4-4), however, other factors such a slight variation in growth and induction conditions also effect levels of protein expression.

Proteins were only purified from the soluble fraction despite also being directed into inclusion bodies. The kinetic analysis undertaken in this chapter only required a small amount of protein so low yields were acceptable. Preparations of bio-active proteins from inclusion bodies can be a time-consuming process and various conditions are usually screened, many of which denature the proteins and require subsequent refolding and may result in low yields of fully folded and bioactive proteins (Singh *et al.*, 2015). Furthermore, there is no guarantee that a refolded soluble protein has a fully native conformation and therefore, activity may also be affected.

4.3.2. T7 gp6 Asp160Lys and Asp162Lys are catalytically inert and suitable for co-crystallization trials

An aim of this study was to generate a T7 gp6 mutant that was capable of binding but not cleaving DNA in the presence of catalytically relevant divalent ions and would therefore be suitable for co-crystallization studies. Residues contributing to the conserved active site of wild-type FENs provide binding sites for 2 or 3 catalytic divalent metal ions. These are co-ordinated through a combination of direct and water-mediated interactions (Tsutakawa *et al.*, 2011). In this study, active-site mutations were introduced into T7 gp6 proteins to replace one of the conserved aspartic acids with a positively charged lysine (Asp160Lys, Asp162Lys and Asp202Lys). The aim of this mutation was to repel the co-ordination of one or more of the positively charged catalytic ions but not the negatively charged phosphodiester DNA backbone, allowing it to enter and interact with the active site.

Activity analysis of the mutants reveals that Asp160Lys and likely Asp162Lys are catalytically inert whilst Asp202Lys retains a low level of exonuclease activity. The zymogram reveals a faint band of activity for Asp202Lys, suggesting the presence of exonuclease activity, whilst no bands were present for Asp160Lys and Asp162Lys (Figure 4-5). However, as previously discussed, a zymogram should only be used as a crude assessment of nuclease activity. A band could indicate tight binding to the DNA preventing staining, which could arise if the mutation increases DNA binding affinity. This could be investigated by soaking the gel with 1% SDS to denature the proteins and then re-staining. A nuclease false-positive band would disappear following this.

The enzymatic activity of the active-site mutants was investigated further with quantitative kinetic assays. The UV assay indicates that Asp160Lys and Asp162Lys do not possess exonuclease activity as the release of acid-soluble nucleotides from 50 μg protein (100-fold excess compared to WT) is not detected. Whilst the release of nucleotides from 50 μg Asp202Lys was detectable, the specific activity is over 500-fold lower than the WT (Figure 4-6), under the conditions used. These results are in agreement with the zymogram and suggest that Asp160Lys and Asp162Lys are most suitable for co-crystallization. The FRET assay with specific endonuclease and exonuclease substrates was also used to further characterize the

kinetics of these mutants (Figure 4-7). This showed that Asp160Lys appeared completely inactive with all the substrates tested, suggesting that it is catalytically inert, whereas Asp202Lys and interestingly Asp162Lys appeared to demonstrate low but detectable levels of activity. However, the change in fluorescence in the FRET assay is negligible, especially for Asp162Lys, which was added at a 60 to 600-fold excess compared to the WT positive control. Therefore, higher levels of nuclease contaminant could be present which may contribute to low levels of nuclease activity not detected in the zymogram.

As this protein was to be used for co-crystallization, it was important to confirm that the active-site mutations did not cause a DNA binding defect, as low affinity interactions with DNA could prevent co-crystallization. A DNA binding defect may arise due to the introduced lysine blocking the active site to prevent positioning of the scissile phosphodiester. Indeed certain active site mutations in human FEN-1 have been shown to disrupt DNA binding ability (Shen *et al.*, 1997). However, in this study the EMSA shows that Asp160Lys, Asp162Lys and Asp202Lys have a binding affinity comparable to the WT and a one-way ANOVA with multiple comparisons reveals no statistical difference between these proteins (Figure 4-8).

4.3.3. Roles of active site residues in FENs

FENs co-ordinate divalent metal ions through two binding sites designated site I and site II. Functional analysis of active site residues has led to debate regarding the precise roles of each of these. Whilst site I mutations generally result in complete loss of nuclease activity; site II seems more variable in effect on enzyme activity (Bhagwat *et al.*, 1997; Shen *et al.*, 1997; Feng *et al.*, 2004; Syson *et al.*, 2008).

Site I of the T5FEN active is made up of residues Asp26, Asp68, Asp128, Glu131 and Asp153 whilst site II is composed of Asp153, Asp155, Asp201 and Asp204 (Figure 1-15). A double mutation of the site I residues Glu128Gln/Asp130Asn resulted in complete loss of nuclease activity and a decrease in the dissociation constant with pseudo-Y DNA compared to the WT protein (7.8 nM compared to 2.6 nM) (Feng *et al.*, 2004). Whilst the site II double mutation Asp201Ile/Asp204Ser retained endonuclease activity and a greatly reduced but still detectable exonuclease activity. The authors reasoned that low levels of activity in the site II mutant could be due to other site II residues Asp153 or Asp155 facilitating transient occupation of the site

by a divalent co-factor. When Asp201 and Asp204 were mutated to positively charged arginine to repel the divalent ion, endonuclease activity was retained whilst a complete loss of detectable exonuclease activity was exhibited and DNA binding affinity was also enhanced compared to WT. These results indicate that in T5FEN, both site I and site II is necessary for exonuclease activity whilst only site I is needed for endonuclease activity (Feng *et al.*, 2004). Studies in T4 RNase H demonstrate that site I (Asp19, Asp71, Asp132, Asp155) is essential for activity and mutagenesis of any of these residues to asparagine (Asn) result in a complete loss of nuclease activity but retain DNA binding. The same mutations in site II (Asp157, Asp200), retained endonuclease and exonuclease activity, albeit at a reduced level for Asp157 (Bhagwat *et al.*, 1997). In comparison, all active-site residues in human FEN-1 appear necessary for endonuclease activity. When a single residue is mutated from site I (Asp34, Asp86, Glu158 and Glu160) or site II (Asp179, Asp181 and Asp233) to an alanine residue, a complete loss of cleavage of a flap DNA substrate is observed. Furthermore, in Glu158, Asp179, Asp233, the alanine mutation also abolished DNA-binding ability (Shen *et al.*, 1997).

These differences observed between the functional roles of FEN active-site residues may be due to the evolution of divergent mechanisms of co-factor interaction. In T5FEN, three divalent co-factors are implicated in the active site; M1 and M2 in site I and M3 in site II and isothermal titration calorimetry (ITC) studies demonstrate that site I has a 30-fold higher affinity for co-factors than site II (Feng *et al.*, 2004; Syson *et al.*, 2008). As shown in Figure 4-16, in archaeal and mammalian FENs, the N-terminus of the protein occupies the analogous position of site II from bacteriophage FENs (Syson *et al.*, 2008). Furthermore, ITC of the mouse FEN-1 active site indicates that both metal ions are bound with the same affinity, which is characteristic of the high affinity binding site in T5FEN (site I) (Zheng *et al.*, 2002). Additionally, in *E. coli* ExoIX, a FEN-family member, only site I is conserved and the enzyme possesses no detectable exonuclease activity (Anstey-Gilbert *et al.*, 2013). This therefore indicates that site II is not fully conserved and can explain the functional differences observed between the active sites of FENs.

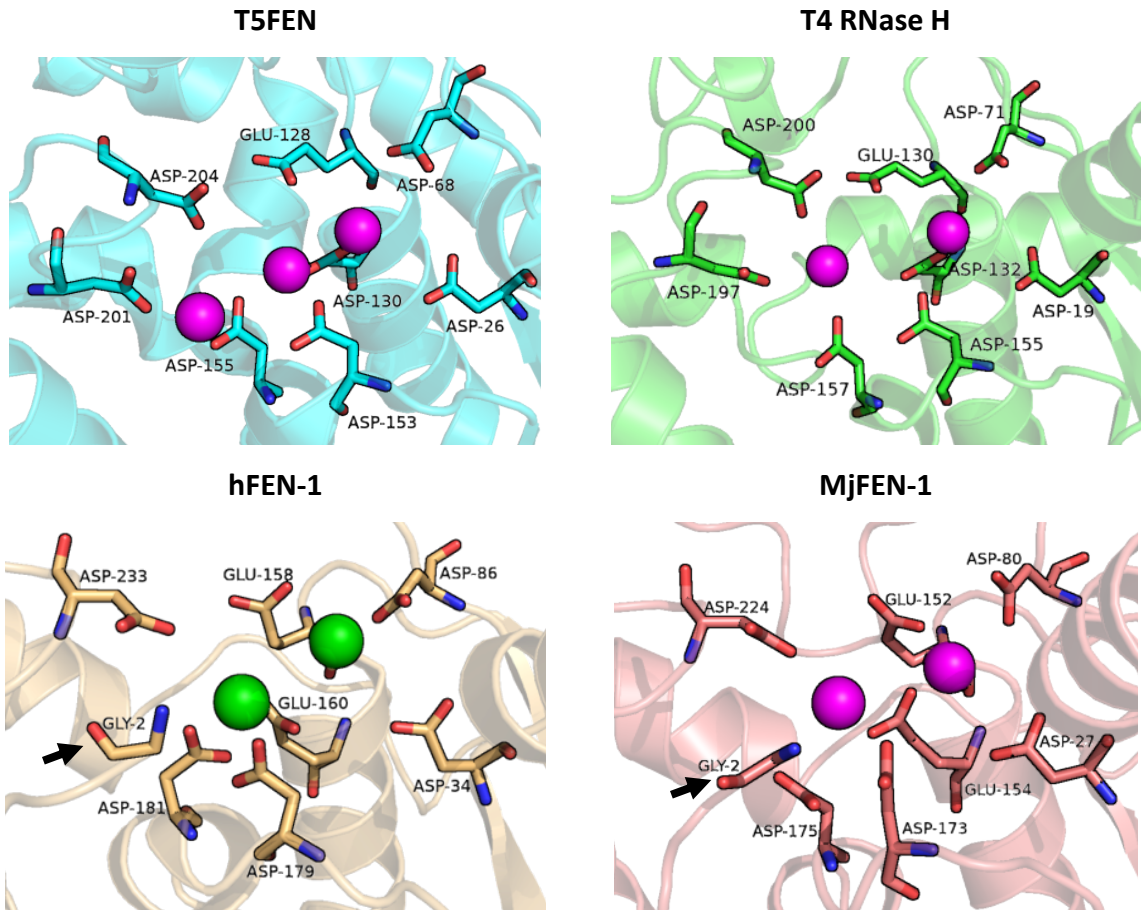


Figure 4-16: Active site from FEN homologoues

The conserved FEN active site is composed of 7 – 8 conserved carboxylate residues, which co-ordinate divalent metal ions required for catalysis. In bacteriophage FENs (T5FEN, PDB: 5HMM; T4 RNase H, PDB 1TFR) metal ions are bound in two binding sites (site I and site II). In archaeal (MjFEN, PDB: 1A77) and mammalian (hFEN-1, PDB: 3Q8L) FENs, two metal ions are bound in site I whilst site II is occupied by the N-terminus of the protein (indicated by black arrow). Metal ions Mg^{2+} are shown by magenta sphere and Sm^{3+} by green sphere.

The active-site mutants characterized here show similarities to the T5FEN and T4 RNase H active-site mutants previously described. The T7 gp6 Asp202Lys (site II) exhibits very low levels of endonuclease and exonuclease activity. This is most similar to the T4 RNase site II Asp157Asn mutant, that retained low levels of nuclease activity (Bhagwat *et al.*, 1997). Whereas in T5FEN, the double site II mutation of positions Asp201Arg/ Asp204Arg resulted in a complete loss of exonuclease activity (Feng *et al.*, 2004). This however, is a double mutation so Asp199 of T7 gp6 (equivalent to Asp201 in T5FEN) and other site II residues may facilitate transient occupation of the active site by Mg^{2+} resulting in low but detectable levels of activity.

The T5FEN site II double mutant also retained higher levels of endonuclease activity compared to T7 gp6 Asp202Lys, which was barely detectable. This result is more similar to the result in *Tb*FEN in which mutation of the equivalent residue, Asp190Lys led to complete abolishment of nucleolytic activity (Oates, 2016). This may suggest evolution of different function of active-site residues between organisms. In this study both Asp160Lys and Asp162Lys led to complete abolishment of nuclease activity. This is in agreement with the findings from T5FEN, in which Asp153Lys and Asp155Lys lost all nucleolytic activity. However, binding affinity also increased, which was not observed for the T7 enzyme (Zhang, 2012; AlMalki *et al.*, 2016).

In T5FEN, two active-site mutants have been successfully crystallized with DNA: Asp153Lys (equivalent to T7 gp6 Asp160Lys) and Asp155Lys (equivalent to T7 gp6 Asp162Lys). Structural analysis reveals that the M1 (T5FEN Asp153Lys) or M2 (T5FEN Asp155Lys) site is occupied by the ε-amino group of the lysine, which prevents the binding of catalytic Mg^{2+} (AlMalki *et al.*, 2016). The predicted effect of the mutations on the active site architecture of T7 gp6 are shown with the active site metals of T5FEN by mutating the residue in PyMol (

Figure 4-17). This suggests that Asp160Lys disrupts binding of M2, whilst Asp162Lys and Asp202Lys disrupt binding of M3. However, lysine is a large, flexible residue and based on the T5FEN data in which Asp153Lys and Asp155Lys result prevent binding of a site I co-factor, it is possible that Asp160Lys and Asp162Lys also result in the same structural effect.

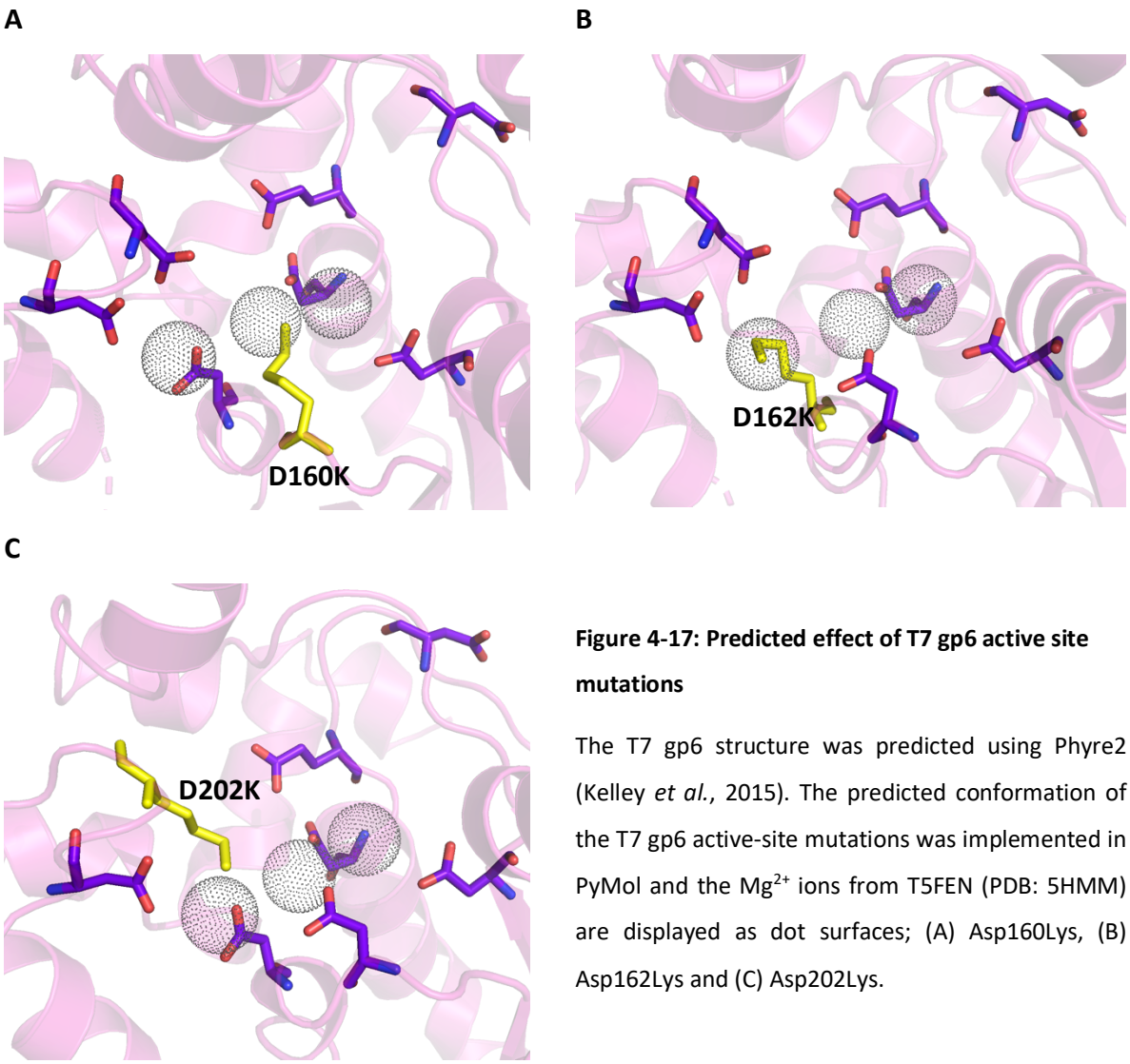


Figure 4-17: Predicted effect of T7 gp6 active site mutations

The T7 gp6 structure was predicted using Phyre2 (Kelley *et al.*, 2015). The predicted conformation of the T7 gp6 active-site mutations was implemented in PyMol and the Mg^{2+} ions from T5FEN (PDB: 5HMM) are displayed as dot surfaces; (A) Asp160Lys, (B) Asp162Lys and (C) Asp202Lys.

4.3.4. Analysis of the exonuclease mutants demonstrates enhanced cleavage capabilities of T7 gp6 Ile200Arg but not T5FEN Leu202Arg

An aim of this work was to investigate whether amino acid modification of residues in the H3TH motif of T7 gp6 will alter the ratios of endonuclease and exonuclease activity. Two exonuclease T7 gp6 mutants were generated (Ile200Arg and Thr201Trp) and analogous mutations introduced into T5FEN (Leu202Arg and Gly203Trp) to enable comparison between these enzymes. These amino acid changes were used to mimic a chemical modification of a cysteine previously introduced in T5FEN (Leu202Cys) which resulted in a selective loss of endonuclease activity when modified with a bulky thiol-reactive agent (Zhang, 2012). The amino acids in this study (arginine and tryptophan) were chosen as they possess more bulky sidechains compared with the original amino acids.

Initial qualitative analysis of the zymogram (Figure 4-9) suggests that compared to the WT of each FEN, Ile200Arg possesses an increased exonuclease activity, Thr201Trp and Gly203Trp exonuclease activity decreases and the exonuclease activity of Leu202Arg is not affected However, zymography analysis has a potential caveat in that the ability of the protein to refold *in situ*, may be affected by the mutation, which can influence band intensity on the gel. The mutants were analysed further with the UV (exonuclease) and FRET (exonuclease and endonuclease) assays. Both Thr201Trp and Gly203Trp have significantly reduced endonuclease and exonuclease activity (Figure 4-10 to Figure 4-12). Progress curve analysis of Thr201Trp indicates that the nuclease defect is more severe with endonuclease substrates with nearly 20 - 50 excess over WT protein used for progress curve determination of Thr201Trp endonuclease. Furthermore, DNA binding studies indicate that Thr201Trp results in a binding defect with endonuclease but not exonuclease substrates, due to a large increase in K_D (Figure 4-15), explaining the more severe defect in endonuclease activity. Due to the severe catalytic effects in Thr201Trp and Gly203Trp, the other two mutants, Ile200Arg and Leu202Arg, were selected for further characterization.

The specificity constant (k_{cat}/K_M) for T7 gp6 Ile200Arg compared to the WT is enhanced for all substrates. Interestingly, the opposite effect is observed in the analogous T5FEN Leu202Arg, in which the k_{cat}/K_M is decreased, demonstrating FEN-specific effects. But why is this the case? Perhaps this is due to differences in the H3TH architecture. For example, residues 199-202 in T7 gp6 are Asp-Ile-Thr-Asp (DITD), where Asp199 and Asp202 are active-site residues. In T5FEN the corresponding residues 200-203 are Asp-Leu-Gly-Asp (DLGD). Isoleucine and threonine both have C-beta branching which means that there is more bulkiness next to the main chain of the protein, which may reduce main chain flexibility and introduction of arginine with a large sidechain may provide flexibility to improve interactions. However, in T5FEN, leucine and glycine are more flexible residues, so arginine does not offer improvement to the interaction with DNA.

Comparing the activities of T5FEN Leu202Arg to the results obtained from chemical modification of Leu202Cys with thiol-modifying agent, indicates that the amino acid substitution gives a similar effect. The exonuclease activity of modified Leu202Cys was investigated with the UV exonuclease assay and the results showed no substantial difference

compared to the unmodified version (Zhang, 2012) and here, the Leu202Arg mutation also shows no statistical difference compared to the WT in the UV assay (Figure 4-10). Modified Leu202Cys had 12% endonuclease activity compared to the unmodified determined with the overhang substrate in the FRET assay (Zhang, 2012), whilst Leu202Arg also exhibited a reduction in endonuclease activity and had over 3-fold reduction in k_{cat} compared to the WT.

The binding affinity of T7 gp6 Ile200Arg for various DNA substrates was determined (Figure 4-15). Compared to the WT, the binding affinity for the endonuclease substrate increases ($p=0.031$) but not for the exonuclease substrate. Tighter binding to the single-flap explains the in overall increase in catalytic efficiency, despite a drop in k_{cat}. The mutation may increase the interacting surface area with the single-flap substrate and thus decrease the K_D. Whereas, the nicked substrate already has a larger interacting surface area so change in K_D is not seen. This interaction may be facilitated by the positive charge of the arginine residue.

The binding affinity for T5FEN and mutants could not be accurately determined in the EMSA in this project. Several attempts were made at optimizing the assay but multiple shifted bands were present which could indicate that the protein aggregated at high concentrations or that multiple binding modes were present. This was disappointing as the EMSA with identical reaction conditions has been used successfully in the past to determine the binding affinity of T5FEN with various substrates (Garforth and Sayers, 1997; Garforth *et al.*, 1999; Dervan *et al.*, 2002). Although the substrate sequences were different and the flap in this study is composed of a hairpin and flap oligonucleotide, whereas previously it was composed of three oligonucleotides. Unfortunately, due to time constraints, the previously successful substrates could not be investigated. However, preliminary investigation of T5FEN with the flap and nicked substrates indicates that Leu202Arg exhibits drastic reduction in K_D compared to WT (results not shown). This increase in binding affinity for the nicked substrate may explain why opposing effects are seen for the catalytic parameters in the T5FEN and T7 gp6 mutation. If the T5FEN Leu202Arg binds DNA much more tightly, a trade-off between rate and affinity is observed. The T5FEN nuclease activity has been shown to be limited by diffusion and if the product remains more tightly bound the turnover rate will be decreased (Sengerová *et al.*, 2010).

Chapter 5 - Structural studies on flap endonucleases

5.1. Introduction

The crystal structure of several FENs have been solved demonstrating that these proteins exhibit a conserved architecture (Figure 1-14). However, some key differences have also been identified such as the helical archway, position of metal ions bound to active site and the presence of a 3' binding pocket (Mueser *et al.*, 1996; Hwang *et al.*, 1998; Chapados *et al.*, 2004; Tsutakawa *et al.*, 2011; AlMalki *et al.*, 2016).

Currently, there is no published structure of T7 gp6. In an attempt to elucidate the structure and uncover DNA-structural interactions of the protein complex, crystal trials of the active site mutants Asp160Lys and Asp162Lys generated in Chapter 4 were attempted. These mutants have been shown to be catalytically inert but retain DNA binding ability and trials were set in the presence and absence of DNA substrates. The T7 gp6 is an important component of the Binx Health IO device, so understanding the molecular interactions might inform assay design or protein engineering approaches.

Whilst crystallization of T7 gp6 was successful, none of the crystals obtained diffracted and so structure determination was not possible. Therefore, structural studies were also carried out on relevant homologues and three novel structures were solved. The N-terminal 5' nuclease domain of *Thermus aquaticus* (Taq) DNA polymerase I (Figure 5-1; TaqFEN) carrying the Asp142Lys mutation (corresponding to Asp160Lys in T7 gp6) was crystallized with a double-flap DNA substrate in the presence of either Mg^{2+} or Ca^{2+}. These represent the first structures of Taq polymerase with DNA bound to the 5' nuclease domain. The structure of T5FEN Asp155Lys was also solved and is a novel structure of this mutant protein. These homologues are relevant to the project due to uses within a number of molecular biological techniques as discussed in section 1.3.

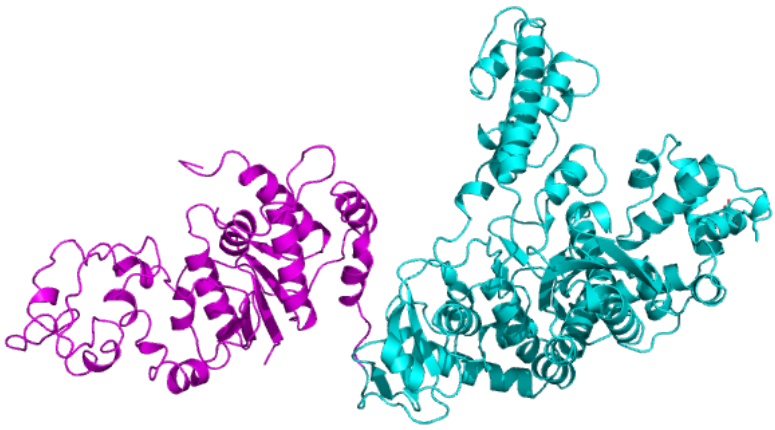


Figure 5-1: Structure of *Thermus aquaticus* DNA polymerase I

Thermus aquaticus DNA polymerase I (PDB: 1TAQ) is composed of two domains: the N-terminal FEN domain (magenta, residues 1 - 293) and C-terminal large (Klenow) fragment (Kim *et al.*, 1995).

5.2. Results

5.2.1. Crystallization of T7 gp6 exonuclease

In an attempt to solve the structure of T7 gp6, crystal trials were set using WT and the catalytically inert active site mutants, Asp160Lys and Asp162Lys. The protein had been purified with size-exclusion chromatography as the last purification stage (section 2.4.4.). The SEC buffer was exchanged to crystallization buffer (25 mM Tris-HCl pH 8, 50 mM KCl, 2 mM DTT), using a centrifugal concentrator. Trials were set with protein alone and with DNA substrates (5OV4, 3OV6, JT2+2; mixed at a 1:1 ratio with active site mutant) in the presence of 50 mM $MgCl_2$ or 50 mM $CaCl_2$. For screening, 5 commercial screens were used and 96-well, 2-drop screening plates were set according to section 2.11.3. This equated to 960 separate conditions, per protein (+DNA) combination screened. Initial screening was performed at a final protein concentration of 10 $mg.mL^{-1}$ although this resulted in an immediate heavy precipitate forming in around 90% of wells and subsequently trials were set up at a protein concentration of 5 $mg.mL^{-1}$.

Overall, this protein proved difficult to crystallize under the conditions explored in the screens. Despite extensive screening at 8°C and 18°C, only 7 promising hits were obtained from all the protein and DNA combinations screened. Examples of some of the crystals obtained are shown in Figure 5-2. Several of these hits were not stained blue by the addition of Izit dye, indicating tight crystal packing which is characteristic of salt crystals. These conditions are

indicated in Figure 5-2 A, B, C, K and L. The conditions which were optimized further are indicated in Figure 5-2 D, E, F, G, H, I and J.

Optimization plates were set, in 96-well format, which slightly varied the conditions of the initial hits (pH, additive and precipitant concentration) and seeding and streak seeding was also performed as seeding can provide nucleation for crystal growth to occur in the metastable zone (Till *et al.*, 2013). The optimization was rather unsuccessful and tended to result in tiny micro-crystals which could not be improved, or in crystals which were large enough to loop but were extremely fragile and many of the crystals broke apart in the looping process.

However, two crystals each from two condition were sent off for data collection but did not diffract to a resolution higher than around 8 Å. Despite extensive screening (>1000s conditions) yielding some apparently crystalline material in various morphologies, only poor datasets were obtained. In order to progress the project and include structural studies, attention was refocussed instead on homologous proteins.

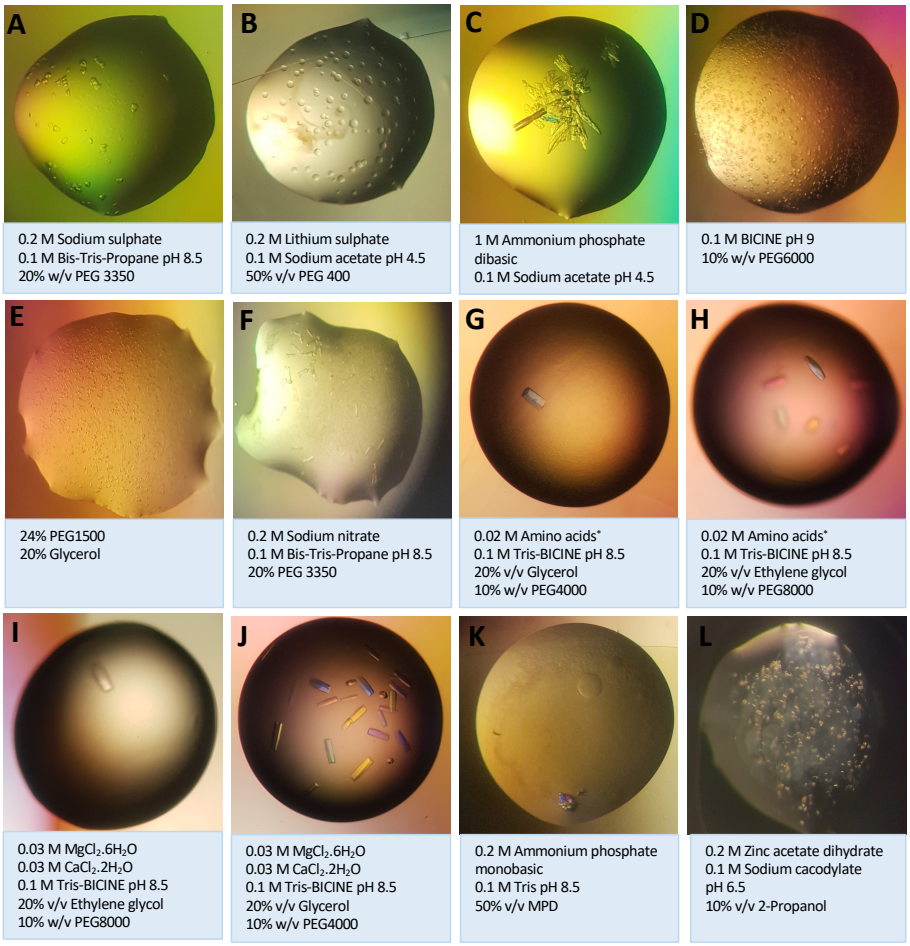


Figure 5-2: Interesting hits of T7 gp6 from screening

Examples of interesting hits from crystallization screening of T7 gp6 proteins (WT, Asp160Lys and Asp162Lys) with and without DNA substrates. Drops viewed under polarised light. *Amino acid additive mix from Morpheus screen is composed of 0.02 M each: L-Na-Glutamate, alanine (racemic), glycine, lysine-HCl (racemic), serine (racemic).

5.2.2. Crystallization of TaqFEN Asp142Lys

5.2.2.1. Purification of TaqFEN Asp142Lys

A 20 g cell pellet of the expressed N-terminal 5' nuclease domain of *Thermus aquaticus* DNA polymerase I (TaqFEN) carrying the Asp142Lys mutation was provided by the Sayers laboratory. This is an active-site mutation of a metal-binding aspartic acid to lysine and corresponds to T7FEN Asp160Lys and T5FEN Asp153Lys, both of which have been shown to be catalytically inert but retain DNA binding ability (Chapter 4; AlMalki *et al.*, 2016). The protein had been overexpressed in the *E. coli* strain BL21 from the bacteriophage T5 promoter plasmid pTTQ18-T5. *Thermus aquaticus* is a thermophilic organism. This property was

exploited during lysis and the cell suspension was incubated at 70°C for 20 minutes, which denatured and precipitated many of the *E. coli* proteins, acting as an initial purification stage. The protein was purified from the soluble fraction by ion exchange (SP and Q Sepharose), affinity (Heparin) and SEC chromatography (Figure 5-3). To aid with purification, the protein sequence was analysed with Expasy ProtParam (Gasteiger *et al.*, 2005). This software computed the pI as 6.28, the extinction coefficient, $\varepsilon_{0.1\%}$, as 1.30 $(mg/mL)^{-1}cm^{-1}$ and the MW as 32619 Da. TaqFEN Asp142Lys does not contain cysteine residues, thus no DTT was included in any purification stage. In total, approximately 500 mg protein was purified after the Q column and was precipitated with 4 M ammonium sulphate for long-term storage at 4°C. Batches of 20 mg were purified via SEC as required for setting crystal trials. Purification from SEC resulted in elution of the protein in a single peak on the chromatogram.

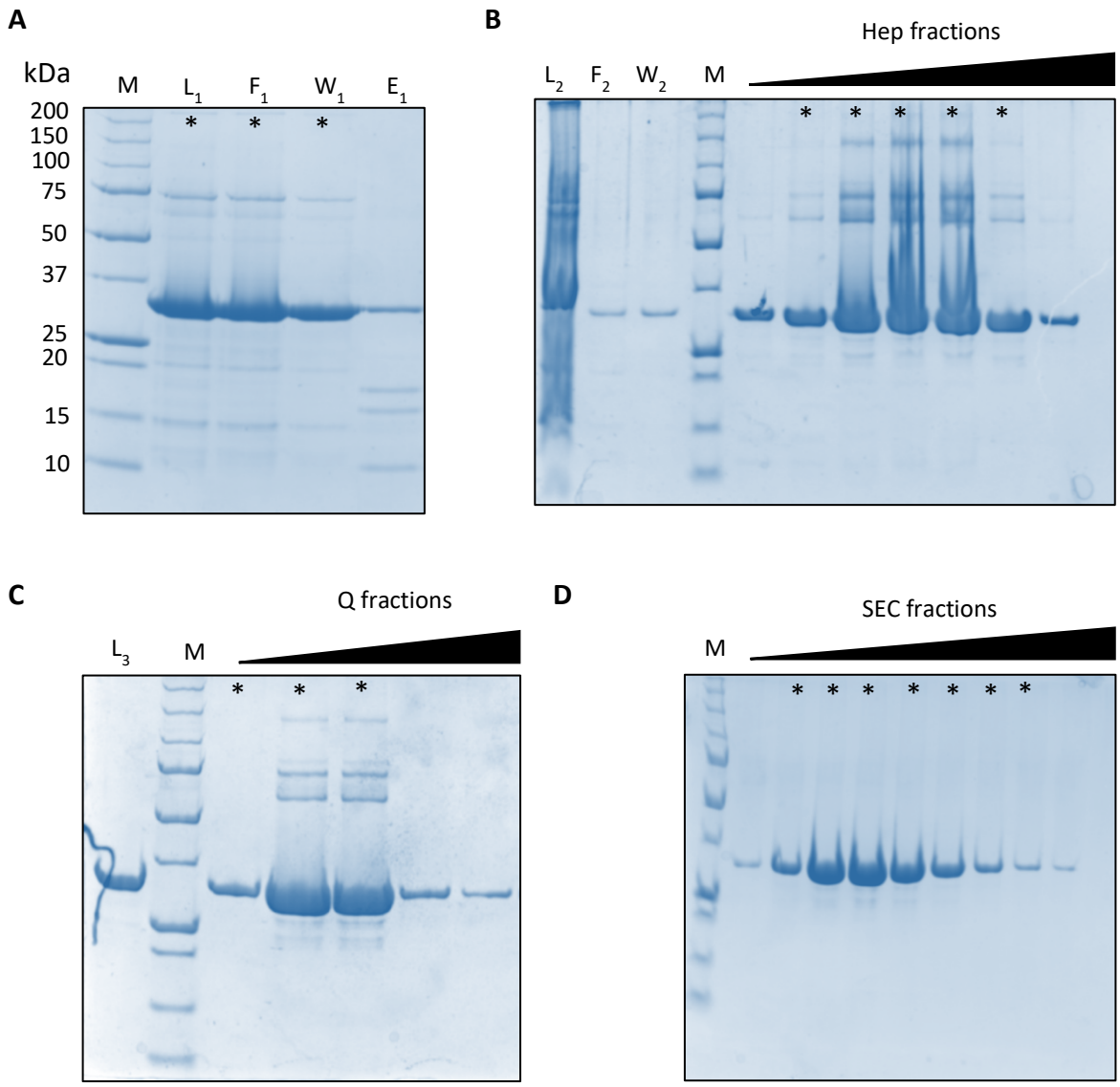


Figure 5-3: Purification of TaqFEN Asp142Lys

SDS-PAGE analysis of TaqFEN Asp142Lys purification. (A) SP Sepharose HP (20 mL) column chromatography in KP8 buffer. Lane L_1, total *E. coli* lysate loaded on to the column; lane F_1, flow-through; lane W_1, wash with buffer containing no NaCl; lane E_1, 1 M NaCl elution. The flow-through and wash were pooled for further purification. (B) Heparin HP (10 mL) chromatography in KP8 buffer. Lane L_2, total protein loaded on to the column; lane F_2, flow-through; lane W_2, wash with buffer containing no NaCl; lanes Hep fractions, 2 mL peak fractions collected over a 50 mL gradient from 0 M – 1 M NaCl. Peak fractions were pooled and dialysed for further purification. (C) Q HP (10 mL) column in Tris8 buffer. Lane L_3, total protein loaded on to the column; lanes Q fractions, 2 mL peak fractions collected over a 50 mL gradient from 0 M – 1 M NaCl. (D) Size exclusion chromatography using a 120 mL column volume. Lanes SEC fractions, 1 mL peak fractions collected and every other fraction was analysed as shown. The fractions pooled at each stage are indicated (*). Analysis was performed on 10% acrylamide gels with PrecisionPlus marker (BioRad) (lane M).

5.2.2.2. Crystallization of TaqFEN Asp142Lys

The solution containing purified protein was exchanged for crystallization buffer (25 mM Tris pH 8, 50 mM KCl) in a centrifugal concentrator. Crystal trials were set at a final protein concentration of 13 $mg.mL^{-1}$ in the presence and absence of 50 mM divalent salts $MgCl_2$ or $CaCl_2$. For co-crystallization with DNA, the protein was mixed with annealed substrates at a 1:1 ratio before setting the trials. The DNA was prepared as described in section 2.11.2 and the substrates investigated were a 5' overhang (5OV4), a 3' overhang pseudo product (3OV6), and a double-flap with 5' flap from 2 - 5 nucleotides in length (JT2+2 - JT2+5). Crystal trials were set with three commercial screens (JCSG+, Proplex and Natrix), using the sitting-drop technique. The plates were incubated at 18°C and crystal growth was observed within 7 days.

A general observation from the crystal trials indicated that protein:DNA mixtures crystallized more readily than protein alone. Crystals were optimized in 96-well plates with optimization screens which were based on initial hit conditions and varied pH, additive, divalent and precipitant concentration. Protein concentration was also varied during optimizations. Approximately 20 of the best crystals (large, single crystals) were looped and flash frozen in liquid nitrogen with 80% mother liquor and 20% glycerol or DMSO used as the cryoprotectant. The crystals were shipped to the Diamond Light Source (Oxford, UK) for data collection.

Most crystals diffracted to between 1.8 Å and 3.5 Å and reasonable data sets were obtained with TaqFEN Asp142Lys with both JT2+2 and JT2+5 substrates. The data was processed using the inbuilt processing pipelines in the *ISPyB* interface. The data was examined by comparing parameters such as resolution and completeness to select which datasets were to be further refined.

5.2.3. TaqFEN Asp142Lys:JT2+2 co-crystal structure

5.2.3.1. Data collection and processing

The first dataset collected was with the double-flap substrate JT2+2, composed of a 2 nt 5' flap and 1 nt 3' flap. The protein was crystallized at a final concentration of 15.5 $mg.mL^{-1}$ in 25 mM Tris pH 8, 50 mM KCl, 50 mM $MgCl_2$. The crystallization condition was 0.1 M MIB buffer

pH 7.5 (sodium malonate, imidazole, boric acid combined at 2:3:3 molar ratio) and 20% (w/v) PEG 1500. The cryoprotectant was 20% DMSO and 80% mother liquor.

A full data set of 3600 images was collected for this crystal with the beamLine I04 (Resolution = 1.80 Å, Exposure = 0.1 s, Beamsize = 60x50 μm, Ω oscillation = 0.10 °, wavelength = 0.9159 Å, Transmission = 100%, Type = SAD) (Figure 5-4). The diffraction data was processed using the Xia2 DIALS automated pipeline to a resolution of 1.82 Å and the space group C 2 2 2_1. The crystal parameters and data processing statistics are shown in Table 5-1.

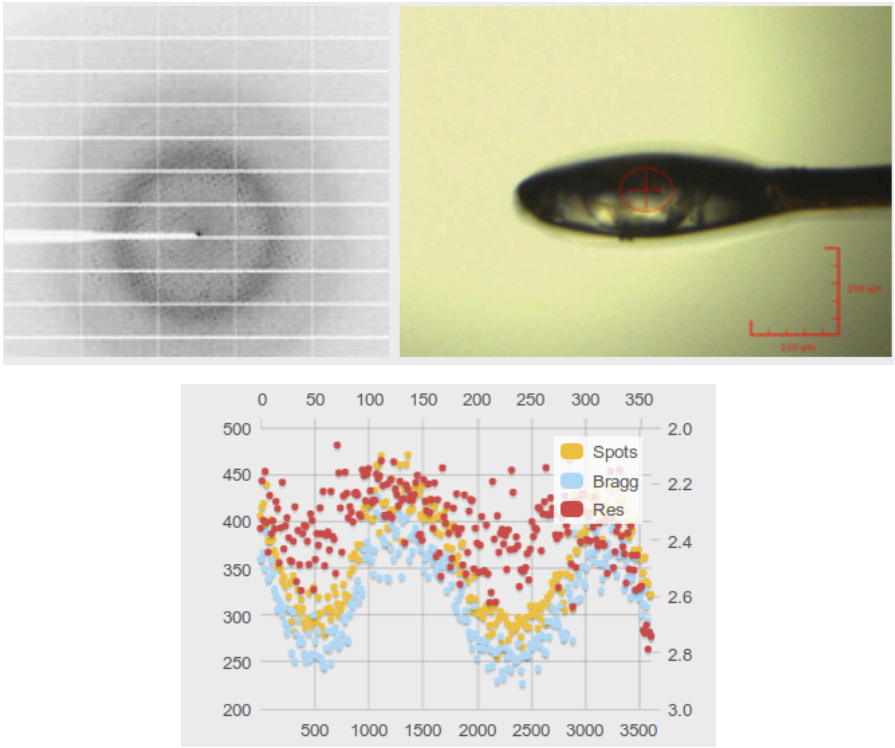


Figure 5-4: Data collection summary of TaqFEN Asp142Lys:JT2+2

Data collection summary showing an example diffraction image, looped crystal snapshot and a *DISTL* plot.

Table 5-1: Data collection statistics for TaqFEN Asp142Lys:JT2+2 dataset

Data collection	TaqFEN Asp142Lys:JT2+2
Wavelength (Å)	0.9159
BeamLine	I04
Resolution range (Å)	90.6 – 1.82 (1.885 – 1.82)
Space group	C 2 2 2_1
Unit cell (a, b, c, α, β, γ)	65.0336, 77.5021, 362.416, 90, 90, 90
Total reflections	1076365 (107308)
Unique reflections	82625 (8142)
Multiplicity	13.0 (13.2)
Completeness (%)	99.86 (99.57)
Mean I/sigma(I)	11.43 (1.04)
Wilson B-factor	34.28
R-merge	0.1063 (2.453)
R-meas	0.1107 (2.552)
R-pim	0.0306 (0.6962)
$CC_{1/2}$	0.999 (0.649)

Diffraction data was automatically processed by the Xia2 dials pipeline. Values in brackets indicate the data in the high resolution (outer) shell.

5.2.3.2. Molecular replacement

The processed data was imported into CCP4i2. The protein and DNA sequences were defined and the task 'define AU contents' was run. This task performed analysis to determine Matthews coefficient, V_M, the crystal volume per unit of molecular weight, which is used to predict the number of copies per asymmetric unit (AU). The output is shown in Table 5-2. As 2 copies is the highest probability by far (98%), this was used for the molecular replacement (MR).

The model for molecular replacement (MR) was prepared in PyMOL and included residues 10-289, corresponding to the FEN domain, of the published structure of Taq polymerase (PDB: 1TAQ; Kim *et al.*, 1995). The programme MOLREP was used to carry out MR (Vagin and Teplyakov, 2010). The best solution was selected by the program and subsequently submitted

to Refmac5 for 20 cycles of refinement (Murshudov *et al.*, 2011). The output model had R_{work} and R_{free} of 0.43 and 0.46, respectively.

Table 5-2: Matthews probability of TaqFEN Asp142Lys:JT2+2

Number of copies	Solvent %	Matthews coefficient	Matthews probability
1	75.93	5.11	0.01
2	51.86	2.55	0.98
3	27.79	1.70	0.01
4	1.28	1.28	0.00

5.2.3.3. Structure building and refinement

The model from the MR output was viewed in COOT (Emsley *et al.*, 2010). Two molecules of TaqFEN were found in the AU. Density fit analysis showed that whilst residues 68-83 were missing (as with the original MR model), residues up to 190 had a good fit to the density. The C-terminus of the model had a much poorer fit to the density. Furthermore, in the mFo-DFc difference map, green un-modelled density indicates the presence of DNA in the model (Figure 5-5).

The BUCANEER automatic protein building pipeline was run to attempt to improve the model before manual building and fitting the DNA (Cowtan, 2006). The pipeline ran 16 iterations and built 663 residues in 13 fragments. The R_{work} and R_{free} dropped to 0.336 and 0.389, respectively. Inspection of the model and density, revealed much clearer density for the DNA, however, some of the protein chain had been built into this region and these were manually deleted before continuing.

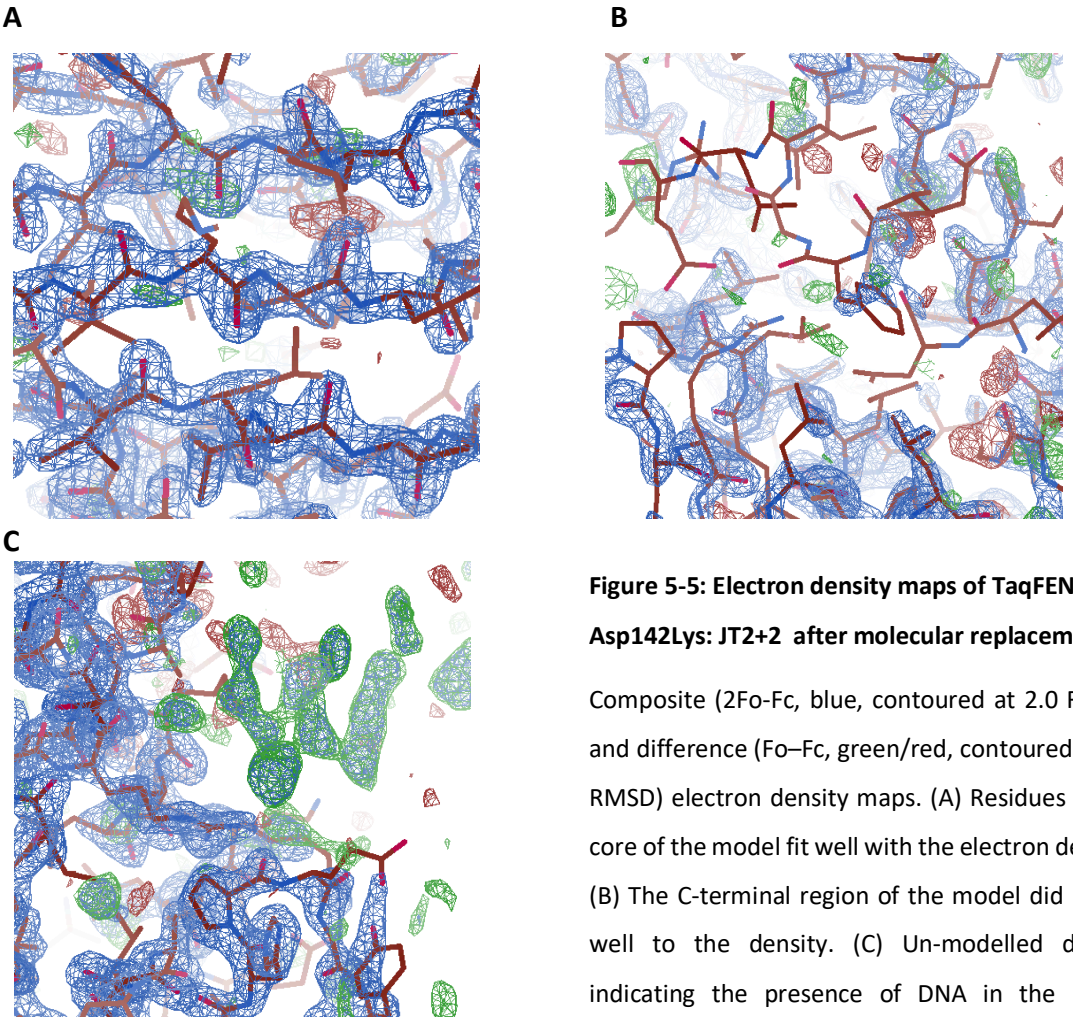


Figure 5-5: Electron density maps of TaqFEN Asp142Lys: JT2+2 after molecular replacement

Composite (2Fo-Fc, blue, contoured at 2.0 RMSD) and difference (Fo–Fc, green/red, contoured at 3.0 RMSD) electron density maps. (A) Residues of the core of the model fit well with the electron density. (B) The C-terminal region of the model did not fit well to the density. (C) Un-modelled density indicating the presence of DNA in the model appears as green mesh.

To build the DNA, the structure of human FEN-1 with flap DNA (PDB: 3Q8L, (Tsutakawa *et al.*, 2011)) was aligned to the protein molecules and the DNA was used as a guide for building and the model was refined for 10 cycles with Refmac5. Figure 5-6A shows an example of how the DNA fits to the density after manual building and refinement.

Several rounds of manual modelling building and refinement were performed to try to improve the structure so that it fits the density better. As seen in Figure 5-6B, the C-terminal region fits the density much better than initially after MR. Once the protein and DNA were built, water molecules and ions were added. The process of model building and refinement was repeated until the R_{work} and R_{free} stopped decreasing. The electron density for the residues corresponding to the archway remained poor and could not be built into the model. This

corresponded to residues 70 - 88 on protein chain A and 77 - 87 on protein chain B of the AU. The final R_{work} and R_{free} was 0.22 and 0.26, respectively.

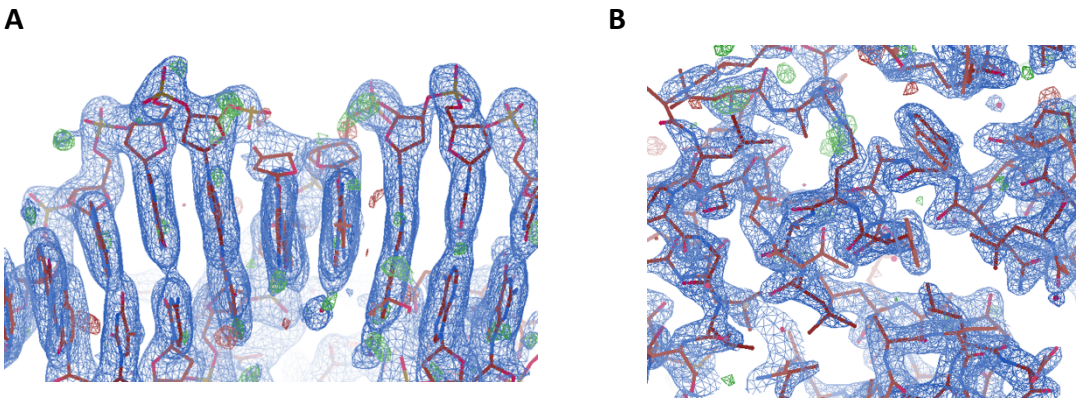


Figure 5-6: Model fitting to the density

(A) DNA. Composite (2Fo-Fc, blue, contoured at 1.0 RMSD) and difference (Fo–Fc, green/red, contoured at 3.0 RMSD) electron density maps. (B) The C-terminal region fits to the density after manual building and refinement. Composite (2Fo-Fc, blue, contoured at 2.0 RMSD) and difference (Fo–Fc, green/red, contoured at 3.0 RMSD) electron density maps.

5.2.3.4. Structure validation

Throughout model building the structure was validated using tools within COOT. This included parameters such as Ramachandran outliers, rotamers and geometry. The final refined structure was analysed by the MolProbity online server (Williams, *et al.*, 2018) The MolProbity report is shown in Figure 5-7.

All-Atom Contacts	Clashscore, all atoms:	3.92		98th percentile* (N=816, 1.82Å ± 0.25Å)
	Clashscore is the number of serious steric overlaps (> 0.4 Å) per 1000 atoms.			
Protein Geometry	Poor rotamers	5	1.11%	Goal: <0.3%
	Favored rotamers	421	93.35%	Goal: >98%
	Ramachandran outliers	1	0.19%	Goal: <0.05%
	Ramachandran favored	514	98.09%	Goal: >98%
	Rama distribution Z-score	0.04 ± 0.34		Goal: abs(Z score) < 2
	MolProbity score^	1.21		99th percentile* (N=11325, 1.82Å ± 0.25Å)
	Cβ deviations >0.25Å	0	0.00%	Goal: 0
	Bad bonds:	1 / 4415	0.02%	Goal: 0%
	Bad angles:	8 / 5964	0.13%	Goal: <0.1%
Peptide Omegas	Cis Prolines:	0 / 24	0.00%	Expected: ≤1 per chain, or ≤5%
Nucleic Acid Geometry	Bad bonds:	1 / 1770	0.06%	Goal: 0%
	Bad angles:	14 / 2720	0.51%	Goal: <0.1%

Figure 5-7: MolProbity analysis of TaqFEN Asp142Lys:JT2+2 structure

5.2.3.5. Final structure of TaqFEN Asp142Lys:JT2+2

The final structure after refinement and validation is shown in Figure 5-8. The residues which correspond to the conserved archway region in homologous FENs (Chain B, residues 77-87) are not traceable in the model but the 5' flap of the DNA appears to interact with a groove in the protein. Charged residues have been highlighted in Figure 5-9 and demonstrate the DNA is positioned and interacting with positively charged residues on the surface of the protein. The structure of the conserved active site is shown in Figure 5-10. There was no obvious electron density to indicate divalent ions are bound to the active site.

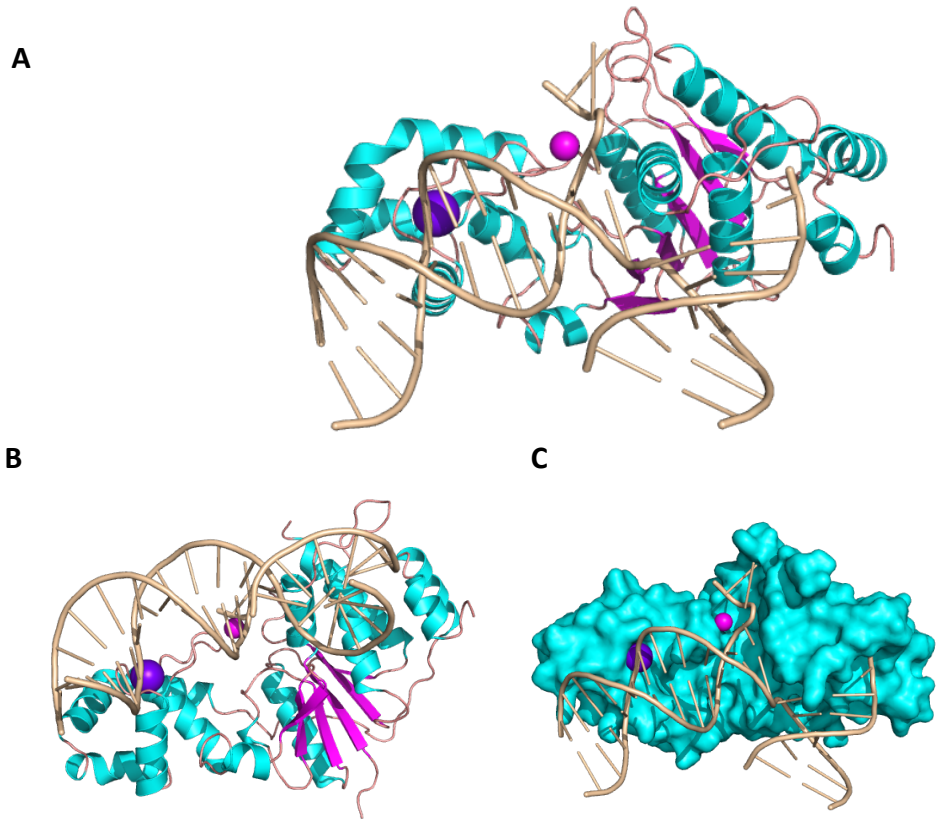


Figure 5-8: Final model of TaqFEN Asp142Lys:JT2+2 co-crystal structure

The final refined structure was viewed in PyMOL. The protein is coloured in cyan (helices), magenta (beta-sheets) and light pink (loop regions). The DNA is light brown and K^+ (purple sphere) and Mg^{2+} (magenta sphere) ions are observed in the structure. (A) Top view. (B) Side view. (C) Top view with surface of protein shown. These figures were collected from protein chain B of the AU. There was no density for the region corresponding to residues 77-87 so they are not modelled in this structure.

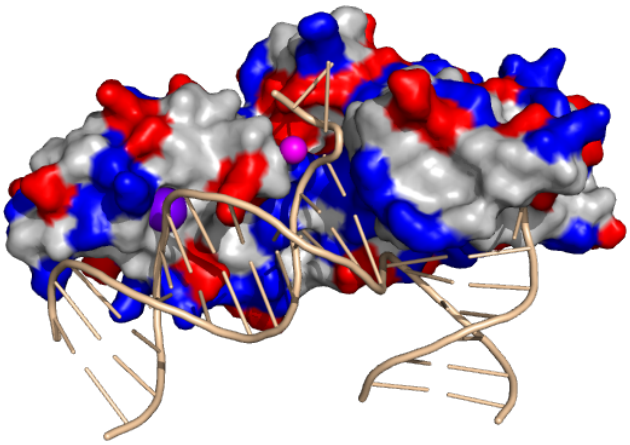


Figure 5-9: Charged residues in TaqFEN Asp142Lys

Chain B of TaqFEN Asp142Lys:JT2+2 structure with bound DNA and ions is shown. Protein is shown as grey surface. Positively charged residues (Arg, Lys, His) are coloured in blue and negatively charged residues (Asp, Glu) are coloured in red. DNA is shown as light brown cartoon. A Mg^{2+} ion (magenta sphere) and K^{+} ion (purple sphere) are bound.

In each molecule of the structure, one Mg^{2+} ion is found, coordinated octahedrally through the oxygen atoms of Ser190, DNA phosphate and four water molecules (Figure 5-11). The interactions between the Mg^{2+} ion enforce a 'looped-up' structure in the 5' flap of the DNA (JT2+2 oligonucleotide) with DNA interactions directly with the phosphate backbone of nucleotide 4 (dT) and mediated through interactions with a water molecule with the phosphate backbone of nucleotide 2 (dA).

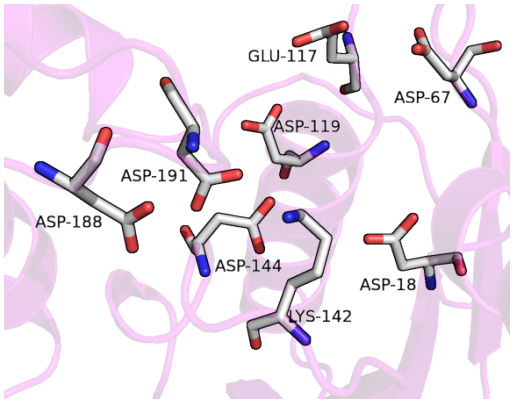


Figure 5-10: Active site of TaqFEN Asp142Lys

Conserved active site residues of TaqFEN Asp142Lys: JT2+2 chain A.

A conserved helix-3-turn helix (H3TH) motif is present in the structure. This is composed of residues 176 – 211. In the model presented here, a K^{+} ion is co-ordinated octahedrally through oxygen atoms of residues Leu185, Thr186, Val196 and Ile199, DNA phosphate of JT3 oligonucleotide (dG-6) and a water molecule (Figure 5-12).

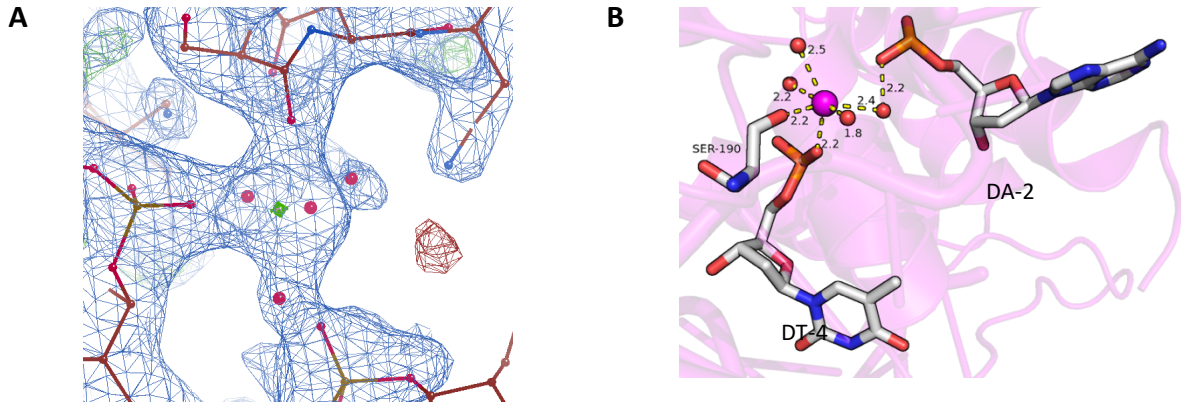


Figure 5-11: Binding of Mg^{2+} ion to TaqFEN Asp142Lys

The Mg^{2+} ion interactions with chain B of TaqFEN Asp142Lys:JT2+2. (A) Composite (2Fo-Fc, blue, contoured at 1.4 RMSD) and difference (Fo–Fc, green/red, contoured at 3.0 RMSD) electron density maps of the Mg^{2+} ion bound to the model (green sphere). (B) Molecules and residues which interact with the Mg^{2+} ion are shown. Distances (Å) of co-ordinating atoms are measured.

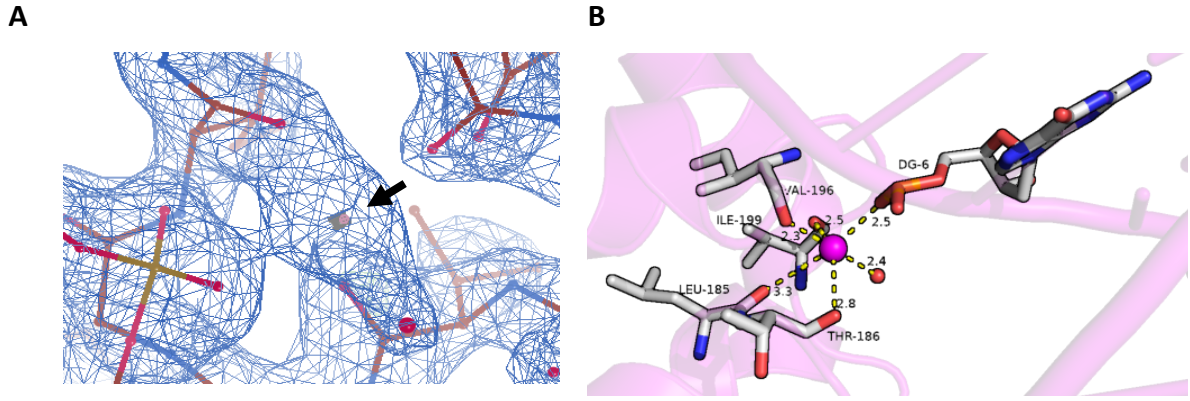


Figure 5-12: Helix-3-turn-helix motif with bound K^+ ion

The helix-3-turn-helix (H3TH) motif of chain B of TaqFEN Asp142Lys:JT2+2. (A) Composite (2Fo-Fc, blue, contoured at 1.7 RMSD) and difference (Fo–Fc, green/red, contoured at 3.0 RMSD) electron density maps of the K+ ion bound (grey sphere, indicated by arrow) to the H3TH motif. (B) The interacting residues and molecules are shown. Distances (Å) of co-ordinating atoms are measured.

5.2.4. TaqFEN Asp142Lys:JT2+5 co-crystal structure

After the TaqFEN Asp142Lys:JT2+2 structure was solved, crystal trials were set with the JT double-flap substrate composed of differing 5' flap lengths up to 5 nucleotides (JT2+5 oligonucleotide), with the aim of capturing a protein: DNA structure, in which all the residues could be modelled.

5.2.4.1. Data collection and processing

The most successful crystallization condition was very similar to those for TaqFEN Asp142Lys:JT2+2 structure described above. The protein was crystallized at a final concentration of 15.5 mg.mL^{-1} in 25 mM Tris pH 8, 50 mM KCl, 50 mM $CaCl_2$. The crystallization condition was 0.1 M MIB buffer pH 5 and 20% (w/v) PEG 1500. The best crystal was obtained from the initial screening plate and although optimizations were set, these subsequent crystals gave worse diffraction data. The cryoprotectant was 20% glycerol and 80% mother liquor.

A full data set of 3600 images was collected for the best crystal on the beamline I03 (Diamond, UK) (Resolution = 1.80 Å, Exposure = 0.01 s, Beamsize = 80x20 μm, Ω oscillation = 0.10 °, wavelength = 0.9789 Å, Transmission = 100%, Type = SAD) (Figure 5-13). The diffraction data was processed using the Xia2 3dii automated pipeline to a resolution of 2.18 Å and a space group of C 2 2 2_1. This is the same space group as the above structure and the cell dimensions were also similar, which indicated that it was the protein: DNA complex which had crystallized. The crystal parameters and data processing statistics are shown in Table 5-3.

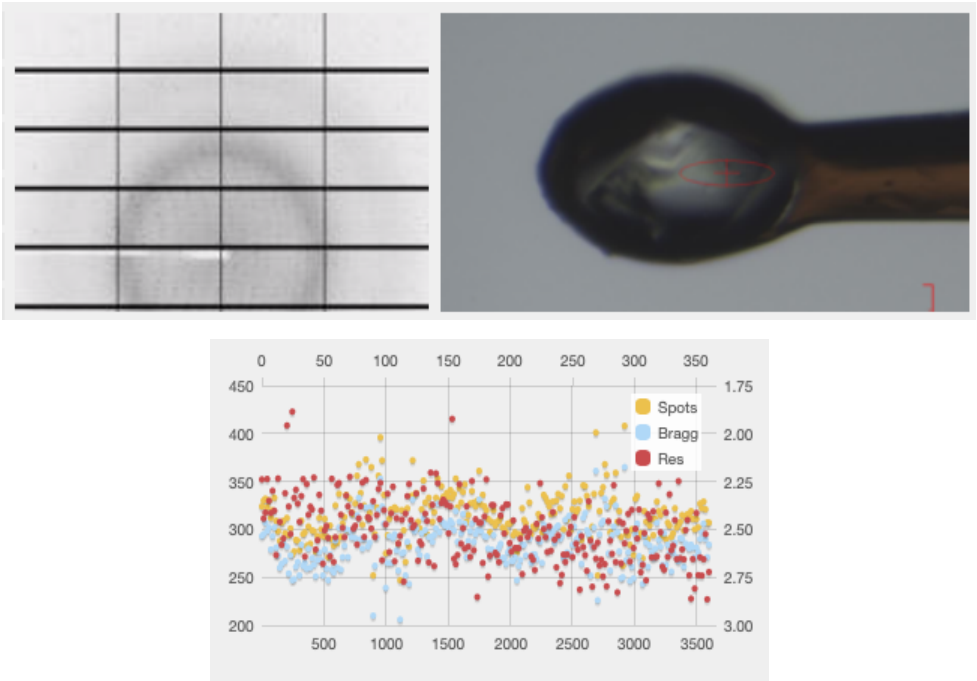


Figure 5-13: Data collection summary of TaqFEN Asp142Lys:JT2+5

Data collection summary showing an example diffraction image, looped crystal snapshot and a *DISTL* plot.

Table 5-3: Data collection statistics for TaqFEN Asp142Lys:JT2+5 dataset

Data collection	TaqFEN Asp142Lys:JT2+5
Wavelength (Å)	0.9789
BeamLine	I03
Resolution range (Å)	46.02 - 2.18 (2.258 - 2.18)
Space group	C 2 2 2_1
Unit cell (a, b, c, α, β, γ)	65.12, 77.16, 362.73, 90, 90, 90
Total reflections	650147 (63077)
Unique reflections	48305 (4717)
Multiplicity	13.5 (13.4)
Completeness (%)	99.71 (99.60)
Mean I/sigma(I)	19.30 (1.52)
Wilson B-factor	63.62
R-merge	0.0557 (1.901)
R-meas	0.05807 (1.977)
R-pim	0.01616 (0.5385)
$CC_{1/2}$	0.999 (0.786)

Diffraction data was automatically processed by the Xia2 3dii pipeline. Values in brackets indicate the data in the high resolution (outer) shell.

5.2.4.2. Molecular replacement

The processed data was imported into CCP4i2. The protein and DNA sequences were defined and the task 'define AU contents' was run to determine Matthews coefficient, V_M, to estimate the number of copies per AU. The output is shown in Table 5-4. Two copies per AU was the highest probability (98%) and this was used for molecular replacement.

The TaqFEN Asp142Lys:JT2+2 structure solved above was used as the search model for MR. This was prepared in PyMOL from one AU consisting of chain B with the corresponding DNA. The programme PHASER was used to carry out the MR and two copies were searched for (McCoy *et al.*, 2007). One solution was found and after submitting to Refmac5 for refinement, the R_{work} and R_{free} were 0.27 and 0.32, respectively.

Table 5-4: Matthews probability of TaqFEN Asp142Lys:JT+5

Number of copies	Solvent %	Matthews coefficient	Matthews probability
1	75.37	4.99	0.01
2	50.74	2.50	0.98
3	26.11	1.66	0.01
4	1.48	1.25	0.00

5.2.4.3. Structure building and refinement

The model from the MR output was viewed in COOT (Emsley *et al.*, 2010). Two molecules of TaqFEN: JT DNA were found in the AU. Density fit analysis showed that most of the model had a good fit to the density, but some areas fitted less well and required more work, such as downstream DNA and some residues corresponding to the archway (Figure 5-14). The residues of the archway region which did not fit the density were removed from the model. After refinement, the density for this region still could not be seen (residues 68 - 89) and there was also no density for the three additional 5' nucleotides from the JT2+5 oligonucleotide (compared to JT2+2 used above), which meant that this model would not yield the 'threaded' structure. However, the model was further refined to see if any different structural interactions were present. Several rounds of model building in COOT and refinement with Refmac5 were undertaken and solvent and ions added. Figure 5-15 shows density maps after model building and refinement. The final refined structure had a R_{work} and R_{free} was 0.244 and 0.287, respectively.

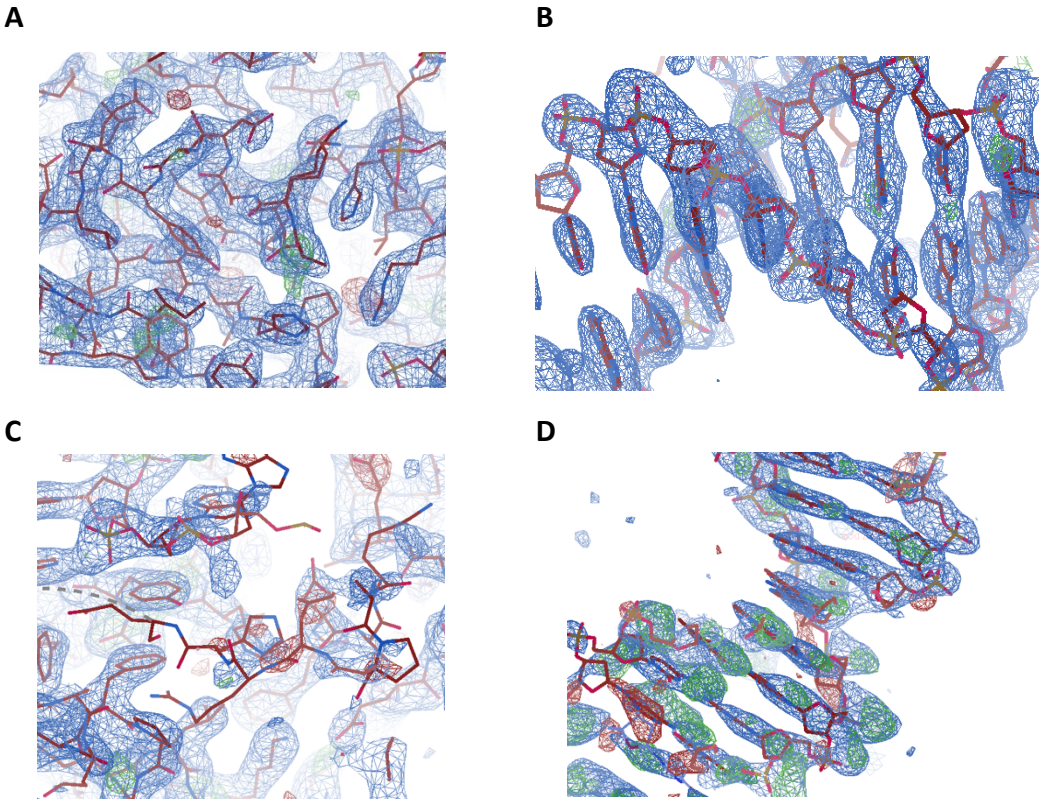


Figure 5-14: Electron density maps of TaqFEN Asp142Lys:JT2+5 after molecular replacement

Composite (2Fo-Fc, blue, contoured at 1.5 RMSD) and difference (Fo–Fc, green/red, contoured at 3.0 RMSD) electron density maps. (A) Most of the model fits well with the electron density. (B) The DNA in the density. (C) The archway region did not fit the density well. (D) Downstream DNA that did not fit the density well.

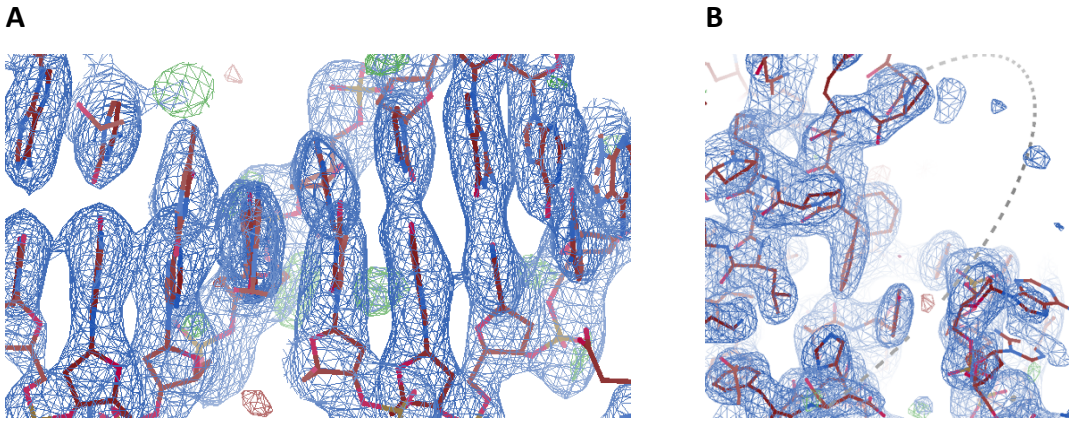


Figure 5-15: Electron density maps of TaqFEN Asp142Lys:JT2+5 after model building and refinement

Composite (2Fo-Fc, blue, contoured at 1.0 RMSD) and difference (Fo–Fc, green/red, contoured at 3.0 RMSD) electron density maps. (A) The DNA which fits well to the density after refinement. (B) Residues 68-89 of the protein and several 5' nucleotides of JT2+5 could not be visualised in the density.

5.2.4.3. Structure validation and final model of TaqFEN Asp142Lys: JT2+5

The final refined structure was analysed by the MolProbity online server (Williams *et al.*, 2018). The MolProbity report is shown in Figure 5-16. These validation statistics are worse than for the TaqFEN Asp142Lys:JT2+2 structure however, the data is of a lower resolution and less time was spent on building and refinement once it was apparent the archway region and additional 5' nts could not be visualised. The final structure is shown in Figure 5-17. As with the previous co-crystal structure, a K^+ ion is bound to the H3TH and interacts with downstream DNA. Additionally, a Ca^{2+} ion is bound (in the place of the Mg^{2+} ion in the previous structure) and interacts with the DNA backbone of the 5' flap.

All-Atom Contacts	Clashscore, all atoms:	3.63		99th percentile* (N=479, 2.18Å ± 0.25Å)
	Clashscore is the number of serious steric overlaps (> 0.4 Å) per 1000 atoms.			
Protein Geometry	Poor rotamers	7	1.59%	Goal: <0.3%
	Favored rotamers	391	89.07%	Goal: >98%
	Ramachandran outliers	1	0.20%	Goal: <0.05%
	Ramachandran favored	494	96.48%	Goal: >98%
	Rama distribution Z-score	-2.00 ± 0.32		Goal: abs(Z score) < 2
	MolProbity score^	1.53		98th percentile* (N=10453, 2.18Å ± 0.25Å)
	Cβ deviations >0.25Å	1	0.21%	Goal: 0
	Bad bonds:	32 / 4323	0.74%	Goal: 0%
	Bad angles:	19 / 5836	0.33%	Goal: <0.1%
Peptide Omegas	Cis Prolines:	0 / 23	0.00%	Expected: ≤1 per chain, or ≤5%
Nucleic Acid Geometry	Bad bonds:	2 / 1757	0.11%	Goal: 0%
	Bad angles:	15 / 2696	0.56%	Goal: <0.1%

Figure 5-16: MolProbility analysis of the TaqFEN Asp142Lys:JT2+5 structure

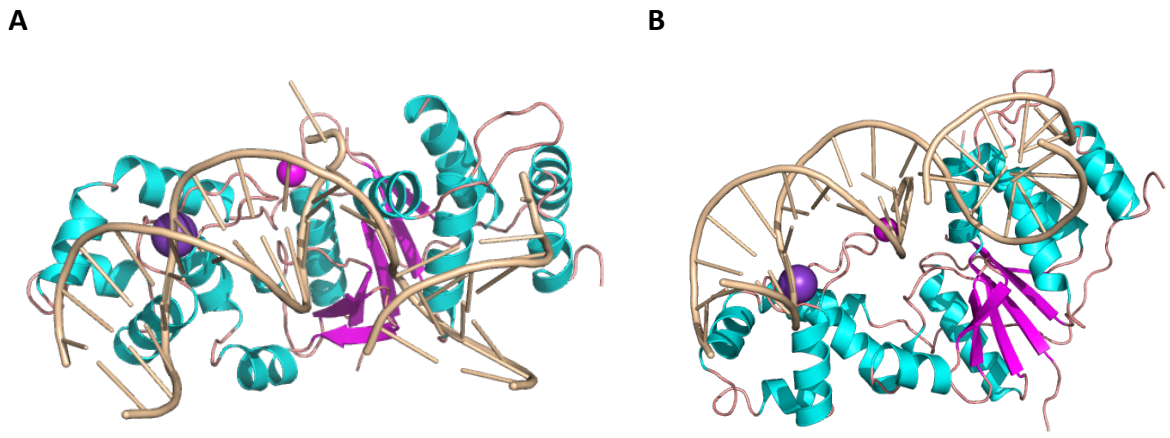


Figure 5-17: Final model of TaqFEN Asp142Lys:JT2+5 co-crystal structure

The final refined structure was viewed in PyMol. The protein is coloured in cyan (helices), magenta (beta-sheets) and light pink (loop regions). The DNA is light brown and K^+ ion (purple sphere) and Ca^{2+} ion (magenta sphere) are observed in the structure. (A) Top view. (B) Side view. These figures represent protein chain B of the AU.

Alignment of both TaqFEN Asp142Lys co-crystal structures was carried out in PyMOL, with both asymmetric units of each structure (Figure 5-18). All molecules aligned very well, with an RMSD of ~0.3 Å, indicating only slight structural variations are present.

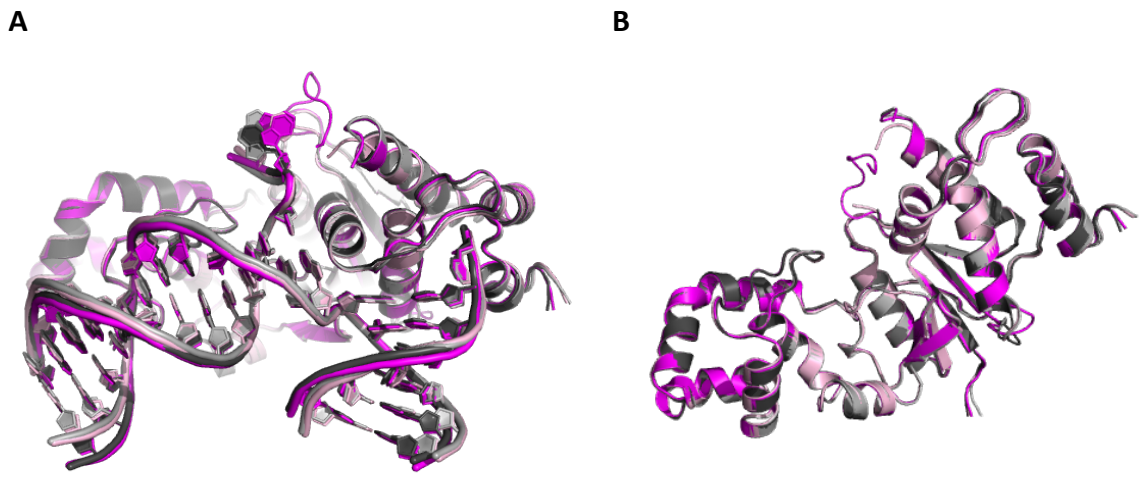


Figure 5-18: Alignment of TaqFEN:DNA structures

Alignment of all TaqFEN Asp142Lys: DNA molecules was carried out using PyMOL. Two molecules in each asymmetric unit were present; TaqFEN Asp142Lys:JT2+2, light and dark pink; TaqFEN Asp142Lys: JT2+5, light grey and dark grey. (A) Alignment including DNA. (B) Alignment without DNA present. All structure aligned with an RMSD of ~0.3 Å.

5.2.5. Crystallization of T5FEN

5.2.5.1. Overexpression and purification of T5FEN Asp155Lys

The T5FEN active-site mutant Asp1155Lys was overexpressed using the pJONEX4 heat inducible system. Figure 5-19A shows induction of protein expression after 3-hour heat shock induction and overnight incubation at 20°C. The protein was purified from the soluble fraction using affinity (heparin HP), ion exchange (Q HP) and size exclusion chromatography columns (Figure 5-19). The SEC was run in the crystallization buffer (25 mM Tris pH 8, 50 mM KCl, 2 mM DTT) and the pooled fractions were flash frozen in liquid nitrogen in small aliquots for storage at -80°C.

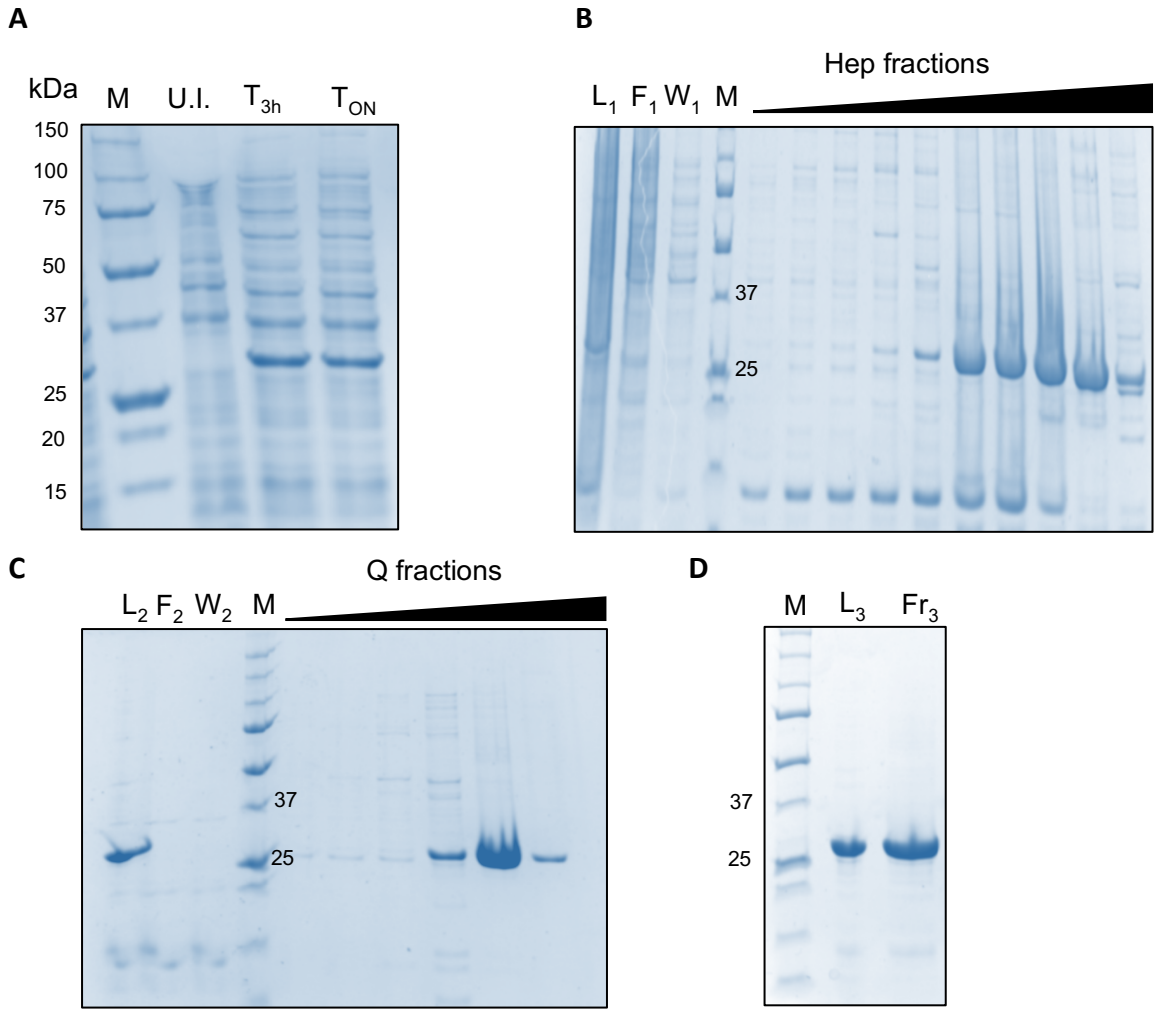


Figure 5-19: T5FEN Asp155Lys overexpression and purification

Analysis of T5FEN Asp155Lys overexpression and purification. (A) Overexpression before (lane U.I.), after 2 hours heat shock induction at 42°C (lane T_{3h}) and after overnight incubation at 20°C (lane T_{ON}). (B) Heparin HP (10 mL) column purification in KP8 buffer. Lane L_1, total protein loaded on to the column; lane F_1, flow-through; lane W_1, wash with buffer containing no NaCl; Lanes Hep fractions, 10 mL peak fractions collected over a 120 mL gradient ranging from 200 mM- 800 mM NaCl. The purest fractions were pooled for further purification. (C) Q HP (10 mL) column in Tris pH 8 buffer. Lane L_2, total protein loaded on to the column; lane F_2, flow through; lane W_2, wash with buffer containing 100 mM NaCl; lanes Q fractions, 10 mL peak fractions collected over a 120 mL gradient ranging from 200 mM – 800 mM NaCl. The purest fraction was taken forward for further purification. (D) Size exclusion chromatography. Lane L_3, protein loaded onto the column; lane Fr_3, pooled peak fractions. Analysis was performed by SDS-PAGE using 10% acrylamide gels with PrecisionPlus marker (BioRad) (lane M).

5.2.5.2. Co-crystallization of T5FEN Asp155Lys with flap JT2+2 substrate

Protein was mixed with the annealed JT2+2 substrate at a 1:1 ratio in the presence of 50 mM $MgCl_2$ or 50 mM $CaCl_2$ and crystal trials were set at a final protein concentration of 19 $mg.mL^{-1}$. Crystal trials were set using the sitting drop technique with three commercial screens (JCSG+, Proplex and Natrix). Crystal growth was observed within 4 days. Several promising conditions were selected for optimization which was carried out as previously described (section 2.11.4.).

5.2.6. T5FEN Asp155Lys structure

5.2.6.1. Data collection and processing

The best crystals were looped and data collected at the Diamond Light Source with beamline I03. In total, five sub 2 Å data sets were collected which were all processed into the same space group, P1, and had similar unit cell dimensions. These crystals were all obtained in very similar crystallization condition, with only slight variations in pH or precipitant concentration.

The best dataset was obtained from a crystal with the protein buffer 25 mM tris pH 8, 50 mM KCl, 50 mM $MgCl_2$ and 2 mM DTT. The crystallization condition was 0.1 M sodium citrate pH 6.5, 100 mM $MgCl_2$ and 16% (w/v) PEG 4000. The cryoprotectant was 20% glycerol and 80% mother liquor. A full data set of 3600 images was collected for this crystal (Resolution = 1.30 Å, Exposure = 0.015 s, Beamsize = 80x20 μm, Ω oscillation = 0.10 ^{o}, wavelength = 0.9789 Å, Transmission = 70.02%, Type = SAD) (Figure 5-20). The diffraction data was processed using the Xia2 dials automated pipeline. The crystal parameters and data processing statistics are shown in Table 5-5.

5.2.6.2. Molecular replacement

The data was imported into CCP4i2 and Matthews coefficient analysis was performed to predict the number of molecules per AU (Table 5-6). The highest probability was 2 molecules (99%) and this was used for MR.

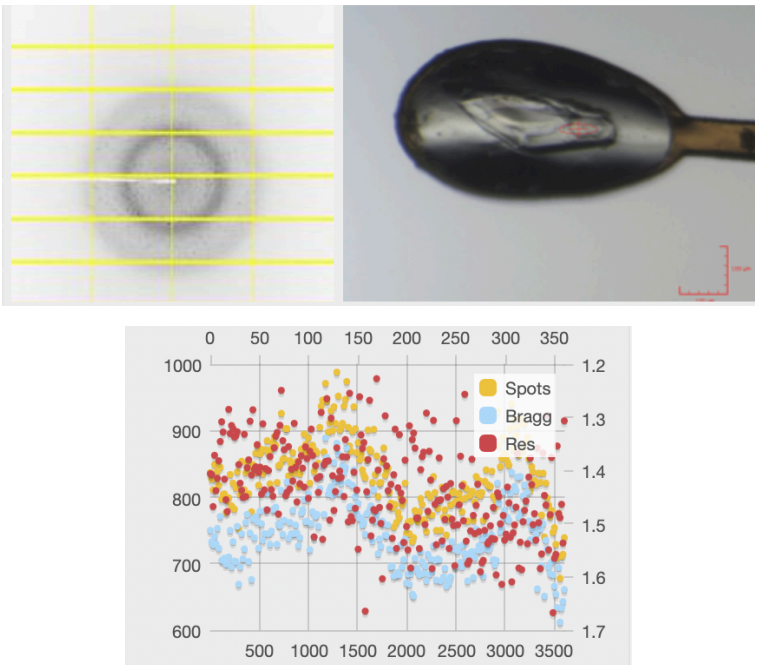


Figure 5-20:Data collection summary of T5FEN Asp155Lys

Data collection summary showing an example diffraction image, looped crystal snapshot and a *DISTL* plot.

Table 5-5: Data collection statistics for T5FEN Asp155Lys dataset

Data collection	**T5FEN Asp155Lys**
Wavelength (Å)	0.9789
BeamLine	I03
Resolution range (Å)	44.83 – 1.44 (1.49 – 1.44)
Space group	P 1
Unit cell (a, b, c, α, β, γ)	48.0, 60.3, 62.6, 64.1, 69.4, 78.0
Total reflections	364839 (33674)
Unique reflections	102638 (10056)
Multiplicity	3.6 (3.3)
Completeness (%)	95.6 (89.2)
Mean I/sigma(I)	8.82 (0.57)
Wilson B-factor	24.36
R-merge	0.05854 (1.147)
R-meas	0.06933 (1.372)
R-pim	0.03667 (0.743)
$CC_{1/2}$	0.995 (0.519)

Diffraction data was automatically processed by the Xia2 dials pipeline. Values in brackets indicate the data in the high resolution (outer) shell.

The model from MR was prepared from one molecule of a previous structure of WT T5FEN (PDB: 5HMM; AlMalki *et al.*, 2016). MR was performed using PHASER and two molecules were searched for (McCoy *et al.*, 2007). The top result was refined in Refmac5 and had a R_{work} and R_{free} of 0.251 and 0.277, respectively.

Table 5-6: Matthews probability of T5FEN Asp155Lys

Number of copies	Solvent %	Matthews coefficient	Matthews probability
1	74.77	4.87	0.01
2	49.53	2.44	0.99
3	24.30	1.62	0.00

5.2.6.3. Model building and refinement

The model was viewed in COOT to see if there was any un-modelled electron density which could indicate that DNA was present. However, unfortunately it did not appear that DNA was present for this dataset and closer inspection of other datasets collected also indicate that no DNA is present in the structures. Despite this, the model was still processed further as there is no published structure of T5FEN Asp155Lys, and this represents a high resolution, novel structure. After MR, the model appeared to fit well to the density. Any problem areas were deleted and rebuilt and several rounds of model building and refinement were performed. Water molecules, solvent and ions were placed. The final refined model had a R_{work} and R_{free} of 0.182 and 0.212, respectively.

5.2.6.4. Structure validation and final model of T5FEN Asp155Lys

The final model was uploaded to the MolProbity server for validation (Williams *et al.*, 2018). The report is shown in Figure 5-21. This is a high-resolution structure so density can clearly be seen for many water molecules and a Mg^{2+} ion is bound to the active site in each molecule in the AU. The Mg^{2+} ion is co-ordinated octahedrally with the oxygen of Asp130 and 5 water molecules. The active site of both molecules in the AU varies slightly and Lys155 is seen in different conformations and in chain B, two conformations are modelled, each with 50% occupancy.

All-Atom Contacts	Clashscore, all atoms:	4.62		92nd percentile* (N=479, 1.44Å ± 0.25Å)
	Clashscore is the number of serious steric overlaps (> 0.4 Å) per 1000 atoms.			
Protein Geometry	Poor rotamers	3	0.60%	Goal: <0.3%
	Favored rotamers	470	94.57%	Goal: >98%
	Ramachandran outliers	1	0.19%	Goal: <0.05%
	Ramachandran favored	532	98.52%	Goal: >98%
	Rama distribution Z-score	-0.55 ± 0.33		Goal: abs(Z score) < 2
	MolProbity score^	1.23		96th percentile* (N=3449, 1.44Å ± 0.25Å)
	Cβ deviations >0.25Å	0	0.00%	Goal: 0
	Bad bonds:	0 / 4650	0.00%	Goal: 0%
	Bad angles:	9 / 6302	0.14%	Goal: <0.1%
Peptide Omegas	Cis Prolines:	2 / 12	16.67%	Expected: ≤1 per chain, or ≤5%

Figure 5-21: MolProbity analysis of T5FEN Asp155Lys

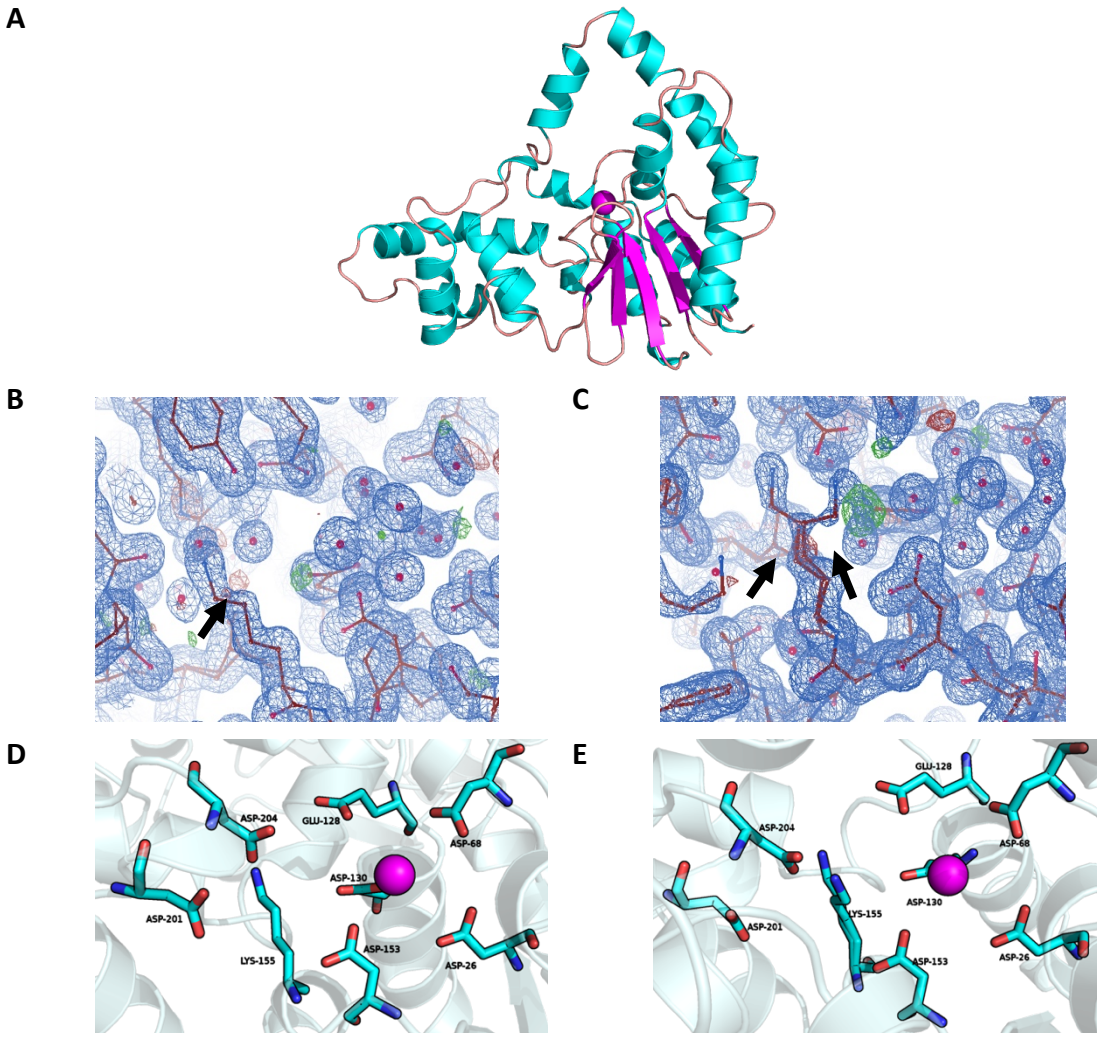


Figure 5-22: Final model of T5FEN Asp155Lys

(A) Final model shown as a cartoon. The protein is coloured in cyan (helices), magenta (beta-sheets) and light pink (loop regions) with Mg^{2+} ion bound to the active site (magenta sphere). (B-C) Composite (2Fo-Fc, blue, contoured at 1.0 RMSD) and difference (Fo–Fc, green/red, contoured at 3.0 RMSD) electron density maps of the active site of chain A (B) and chain B (C), showing the alterative conformations of Lys155 (indicated by a black arrow). In chain B model each conformation of Lys155 is modelled at 50% occupancy. Water molecules are shown as pink spheres and Mg^{2+} ion is the green sphere. (D-E) Active site residues of chain A (D) and chain B (E) with Mg^{2+} ion (magenta sphere).

5.3. Discussion

5.3.1. Crystallization of T7 gp6

As there is no published structure of T7 gp6, an initial aim of this project was to crystallize the protein to yield more information on the structural interactions of the protein. For this, extensive screening which varied commercial screens, additives, protein concentration, temperature and DNA substrates was performed. However, despite crystalline material in various morphologies being sent for data collection, a dataset of a reasonable resolution to be solved was never obtained. The vast screening of 1000s of conditions covering large chemical space, indicates that the issue of difficulty crystallizing and poor-quality crystals probably does not lie with the elusive condition not being identified. It has been estimated that around 80% of crystallizable proteins can be crystallized with only 50 different conditions (Jancarik and Kim, 1991).

Intrinsic characteristics of a protein can also affect its ability to crystallize. For example, flexible terminal regions may be disordered and could introduce heterogeneity into a protein structure and interfere with crystal lattice formation. There have been multiple documented examples of where truncation of N- or C-terminus regions have been utilized to engineer proteins which have a higher propensity to crystallize in addition to improved resolution (for a review see: Dale and Oefner, 2003). A truncated T5FEN mutant called T5FENΔ19, which lacks the first 19 amino acids has no effect on the structure, binding or cleavage capabilities of the enzyme but is more stable than the WT and has been used to generate several high resolution structures (Ceska *et al.*, 1996; Garforth and Sayers, 1997; AlMalki *et al.*, 2016). Truncation mutants could be used in this project and generated either by altered construct design during recombinant expression or by limited proteolysis (Dong *et al.*, 2007).

Specific surface residues can also impact on the crystallization properties of a protein. For example, studies have shown that mutagenesis of residues with large flexible side chains to smaller amino acids has generated quality diffracting crystals from proteins otherwise resistant to crystallization, due to a reduced conformational entropy (Derewenda, 2004). Cysteine residues can form disulphide bonds in the absence of a reducing agent and if solvent

exposed, can result in oligomerisation, which can inhibit crystallization. Sequence analysis of T7 gp6 reveals that there are 8 cysteines present, whilst analysis of the location of these on the predicted structure of the protein suggests that 4 of these are solvent exposed (Figure 5-23). Generation of the predicted structure will be discussed in more detail below (section 6.3.2.). Purification of T7 gp6 was carried out in the presence of the thiol-reducing agent DTT, and SEC gave one elution peak, indicating one oligomeric state. However, DTT has a short half-life in aqueous solutions containing oxygen which ranges from 40 hours at pH 6.5 to 1.4 hours at pH 8.5 at 20 °C and may result in oligomers forming after the crystal drops have been set up. An alternative reducing agent, Tris (2-carboxyethyl)phosphine hydrochloride (TCEP-HCl), with a greater half-life and improved stability was also used during later crystal trials but did not appear to improve the crystallization properties. Substitution mutations of cysteine residues have been successfully used in the literature, for example, mutation of the solvent exposed cysteine to alanine in a histidine ammonia-lyase protein abolished aggregation and generated crystals which diffracted to 1.8 Å (Schwede *et al.*, 1999). In the case of T7 gp6, a cysteine-free mutant (T7 gp6 8CA) was generated in which all 8 cysteines were substituted for alanine. However, the protein was insoluble and exhibited greatly reduced nuclease activity, indicating that these mutations affect both the structure and function of the protein (data not shown). Alternatively, mutation of only surface-exposed cysteines may provide a less severe phenotype and may warrant investigation.

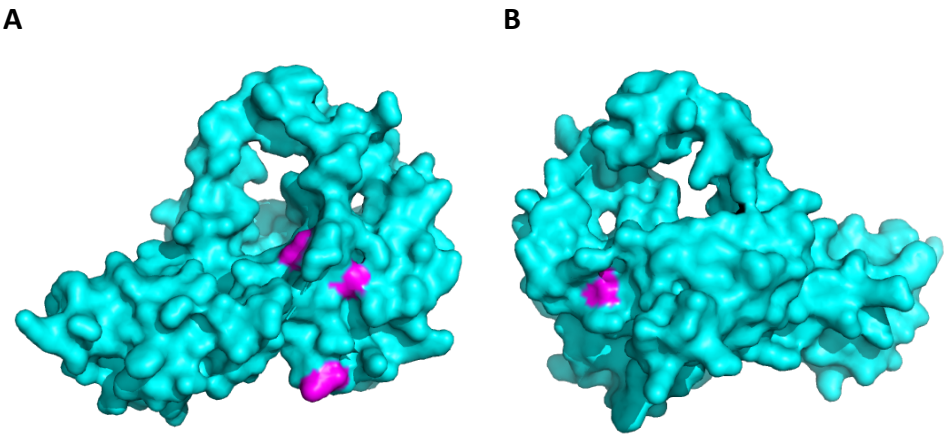


Figure 5-23: Predicted solvent exposed cysteine residues on T7 gp6

The surface exposed cysteine residues on the predicted structure of T7 gp6 have been shaded in magenta. (A) front view of protein. (B) Rotated 180° along Y-axis.

5.3.2. The predicted structure of T7 gp6

The structure of T7 gp6 was predicted from the amino acid sequence using the Phyre2 web server (Kelley *et al.*, 2015). This program detects homology between the submitted sequence and the sequence of a known structure to select a template for modelling. The top 15 results from this are all members of the FEN family with a 100% confidence. A top result uses T5FEN as the template and has an alignment coverage of 81% with residues 19-262 of T7 gp6 included in the model. The modelled structure was aligned with T5FEN (5HMM; Almalki *et al.*, 2016) using PyMOL and had a RMSD of 2.12 Å (Figure 5-24).

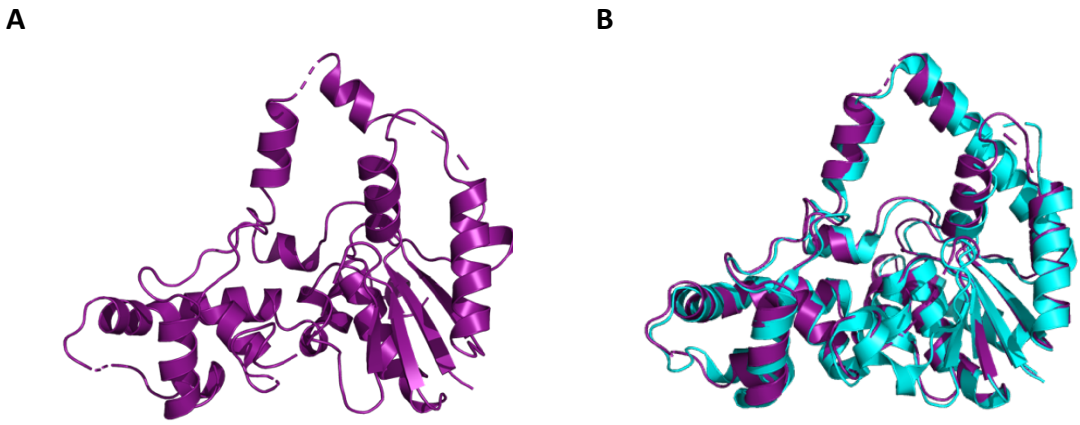


Figure 5-24: Phyre2 predicted structure of T7 gp6

The predicted structure of T7 gp6 was generated using the Phyre2 web server. (A) T7 gp6 predicted structure generated by the software which chose T5FEN (PDB: 1UT8) as the template. (B) Structural alignment with T5FEN (PDB: 5HMM) was carried out using PyMOL (RMSD: 2.12 Å)

More recently, machine learning has been used to develop a programme called AlphaFold2, which is widely regarded as the 'gold-standard' in protein structure prediction (Jumper *et al.*, 2021). The developers have released the full programme code along with a simplified version online as AlphaFold Colab (available at: AlphaFold.ipynb). This "collaborative note-book" version of AlphaFold2 was also used to generate a predicted structure of T7 gp6 (Figure 5-25). The overall protein architecture remains similar between both models, however, two regions of the alphafold2 model differ from the Phyre2 model and indeed the overall conserved structure of other FENs (see Figure 1-14 for structures of homologous FENs). These regions are residues 38 -50 and 220 – 250, which are structured into beta-sheets in the AlphaFold Colab model and correspond to regions on the predicted aligned error plot which have the

highest uncertainty of modelled position (Figure 5-25B). Furthermore, in the T5FEN crystal structure residues 36-40 could also not be modelled, indicating that conformational flexibility may exist in this region of FEN proteins (AlMalki *et al.*, 2016). Differences observed in regions of the AlphaFold Colab model between FENs may be due to the version of the programme used for this prediction which whilst easy to access and use, utilizes a smaller version of the protein database and does not use homologous structures which could result in a drop in the accuracy of the prediction for a small fraction of proteins compared to the full version of AlphaFold2. However, it is important to note that either structure could represent a more accurate model of T7 gp6 and complete validation of a theoretical protein structure requires experimental methods.

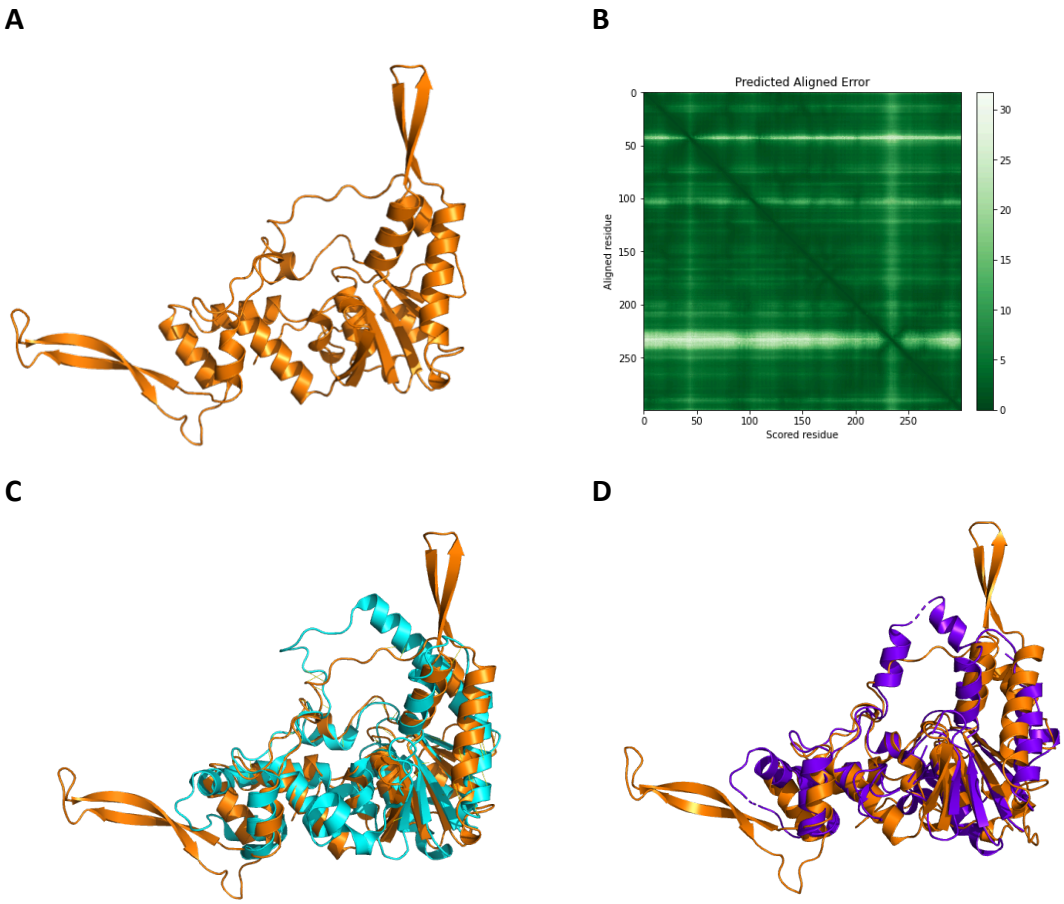


Figure 5-25: AlphaFold Colab prediction of T7 gp6 structure

(A) The collaborative notebook version of AlphaFold2 was used to generate a predicted structure of T7 gp6. (B) Predicted regions of error in the AlphaFold Colab model. (C) Structural alignment of the AlphaFold Colab model (orange) with T5FEN (PDB: 5HMM) using PyMOL (RMSD: 2.89 Å). (D) Structural alignment of the AlphaFold colab model with the Phyre2 model (purple) (RMSD: 4.79 Å)

5.3.3. Crystallization of TaqFEN Asp142Lys

Two co-crystal structures of TaqFEN with a double-flap DNA substrate are reported in this chapter. However, despite setting crystal trials of TaqFEN alone, no structure of the *apo* protein was obtained. A general observation was that TaqFEN: DNA complexes crystallized more readily than protein alone. This was also seen with the crystallization of *Streptococcus pneumoniae* polymerase I and FEN domain truncation mutant (Lau, 2017). Proteins are often stabilized when they are complexed with a substrate, nucleic acid, cofactor or small molecule and it has been shown that this enhanced conformational stability can increase the chances of successful crystallization (Hassell *et al.*, 2007). Disordered regions of a protein also affect conformational stability and proteins with disordered regions will be less likely to assemble into a crystal (Deller *et al.*, 2016). It is therefore likely that the interaction of TaqFEN with DNA stabilized the structure to help facilitate assembly of the crystal lattice.

5.3.4. The structures of TaqFEN Asp142Lys with double-flap DNA

TaqFEN was crystallized with double-flap DNA substrates. The aim of this was to capture the structural interactions between protein and DNA, including residues which correspond to the conserved archway of other FENs, of which no crystallographic model of these residues from Taq have been obtained to date. In this study, two structures of TaqFEN with DNA were refined and presented here. The structure of TaqFEN Asp142Lys:JT2+2 with Mg^{2+} was determined to 1.82 Å and TaqFEN Asp142Lys:JT2+5 with Ca^{2+} determined to 2.18 Å.

Structural alignment of each of the molecules from both structures, highlights the high similarity between these, with the slight differences in the downstream DNA position and the part of the archway which could be modelled, with an RMSD of ~0.3 Å (Figure 5-18). This is not surprising as both complexes were crystallized in similar conditions and in the same space group. Based on this high structural similarity, the higher resolution structure, TaqFEN Asp142Lys:JT2+2, was chosen for a more detailed structural analysis.

5.3.4.1. Structural comparison with Taq polymerase I

The structure of full-length Taq DNA polymerase I has been solved to a resolution of 2.4 Å (Kim *et al.*, 1995). This represents the only previously published structure which includes the

FEN domain and therefore, the models presented in this thesis provide the first crystallographic structural characterization of DNA interactions with the FEN domain of Taq DNA polymerase I. The structure of TaqFEN Asp142Lys aligns well with Taq polymerase I, with an RMSD of 1.06 Å (Figure 5-18). This superposition shows the best alignment is between the helices and beta-sheets that make up the core of the protein, whilst the loop regions have more structural variation. This may be because these are flexible regions or conformational changes that are induced due to DNA and ion interactions. Comparison of specific conserved motifs with published structures will be included in the discussion of the corresponding sections below.

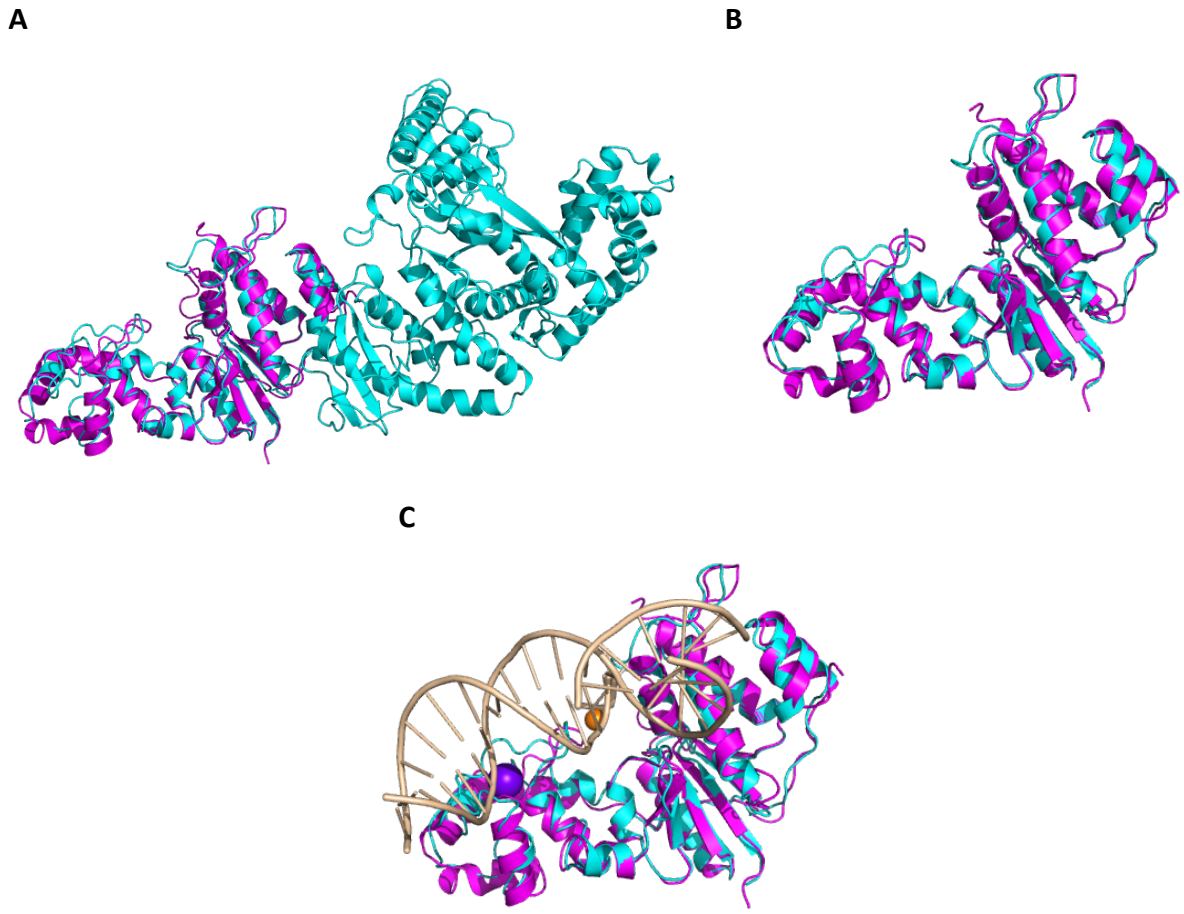


Figure 5-26: Comparison of structure to published Taq DNA polymerase I structure

TaqFEN Asp142Lys:JT2+2 (Chain B, magenta) was aligned to the structure of Taq DNA polymerase I (cyan; PDB: 1TAQ), with an RMSD of 1.06 Å (A) Alignment to entire Taq DNA polymerase I structure. (B) Alignment to the FEN domain of Taq DNA polymerase I (residues 10 - 285). (C) Alignment including DNA and metal ions from the TaqFEN structure (Mg^{2+}, orange sphere; K^{+}, purple sphere).

5.3.4.2. The conserved archway motif

The structure with JT2+2 DNA was the first obtained in this project and whilst most of the model could be built into the density, the residues which corresponded to a conserved archway motif observed in the crystal structure of other FENs could not be modelled, indicating that this is a flexible region of the protein. These are residues 70 - 88 on protein chain A and 77 - 87 on protein chain B (TaqFENAsp142Lys:JT2+2). In the published Taq DNA polymerase I structure there is a large overlap in the missing residues, with residues 69- 84 also missing here (Kim *et al.*, 1995), showing that this region of the protein has proved difficult to crystallize. This may indicate that it is hard to capture this region in a single, stable conformation.

With the aim of being able to model the entire protein molecule, crystal trials were set with other JT2 double-flap substrates (JT2+3 to JT2+5), which differed only by the 5' flap length. It was reasoned that a longer flap may facilitate different interactions between the protein and DNA which may stabilize these unmodeled residues. From these trials, the second structure, TaqFEN:JT2+5, was obtained. Unfortunately, these residues in this structure also could not be modelled and the density for the additional 5' nucleotides was also missing. It is unlikely that these missing nucleotides have been cleaved by the protein as TaqFEN Asp142Lys is an active site mutation and predicted to be catalytically inert, whilst the co-crystal structure was also obtained in the presence of Ca^{2+}, which cannot support the nuclease activity of FENs (Garforth *et al.*, 2001). An alternative explanation is that the 5' flap is flexible and the un-modelled nucleotides are in different positions/ conformations throughout the crystal lattice, so no clear density for these can be seen in the difference maps.

Crystallization captures a static snapshot of a molecule and the conditions may not provide the correct environment for the stabilization of flexible regions of the protein and DNA into one conformation. Re-screening crystallization conditions and alternative substrates with varying nucleotide sequences, dsDNA regions, overhang or nicks may capture the complex in an alternative state in which the entire protein molecule can be modelled. Assays such as the EMSA or thermal shift could be used to screen substrates and conditions which provide the tightest binding or most stable complexes.

5.3.4.3. The active site

Structural comparison with the FEN active site of Taq polymerase I is shown in Figure 5-27. In this structure a Zn^{2+} ion is bound to a WT active site in site I and is co-ordinated by the active site residues Asp18, Asp119 and Asp142. In the structure obtained in this project, Asp142 is replaced by Lys142 and Asp119 points away from the active site.

Typically, in the FEN active site, 2 or 3 divalent metal ions bind to facilitate nuclease activity (Zheng *et al.*, 2002; Syson *et al.*, 2008). The active site of the TaqFEN Asp142Lys structures obtained here have no divalent ions bound to active site residues, despite a $MgCl_2$ or $CaCl_2$ concentration of 50 mM in the crystallization buffer, to give a final drop concentration of 25 mM. The mutation aspartic acid to lysine introduces a positive charge to the active site and removes a negative charge, resulting in a charge change of +2 compared to the WT protein. The aim of this mutation was to repel the binding of positively charged catalytic ions whilst providing a positive charge that can interact with the negatively charged DNA backbone, to generate a mutant that is catalytically inert but retains DNA-binding capacity. Therefore, it is not surprising that there are fewer ions than observed in the WT.

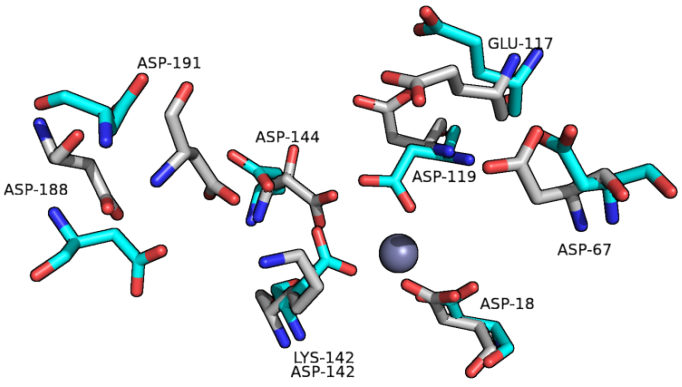


Figure 5-27: Comparison of TaqFEN Asp142Lys active site with Taq polymerase I

The structure of TaqFEN Asp142Lys:JT2+2 (chain A) was aligned with Taq polymerase I (PDB: 1TAQ). Active site residues are shown as sticks (grey, TaqFEN Asp142Lys; cyan, Taq pol). A Zn^{2+} ion (grey sphere) is bound to the active site of Taq pol. The Asp142Lys mutation is shown.

Three divalent metal ions bind to the T5FEN active site. As shown in the crystal structure of WT T5FEN, site I accommodates metal ion 1 (M1) and metal ion 2 (M2), whilst site II accommodates metal ion 3 (M3) (PDB: 5HMM). The active-site mutation used here, Asp142Lys, is equivalent to the T5FEN mutant Asp153Lys, which has also been successfully

crystallized alone (PDB:5HML) and bound to an overhang DNA substrate (PDB: 5HNK) (AlMalki *et al.*, 2016). Comparison of the active site residues are shown in Figure 5-28, highlighting the conserved structural similarity of the FEN active site. In these T5FEN structures, Mg^{2+} ions are bound to active site residues. In the active site of the *apo* structure of T5FEN Asp153Lys, Lys153 occupies the M1 binding site, and M2 is occupied by a water molecule, whilst a single Mg^{2+} is bound, at position M3. In the co-crystal structure, Lys153 occupies position M2 whilst M1 and M3 are each bound to Mg^{2+} ions (AlMalki *et al.*, 2016), indicating this lysine is structurally flexible within the active site. In the structure obtained here, Lys142 occupies a similar position as Lys153 in the DNA-bound T5 structure, although it is not certain if three metal-ion binding sites are present in the TaqFEN active site.

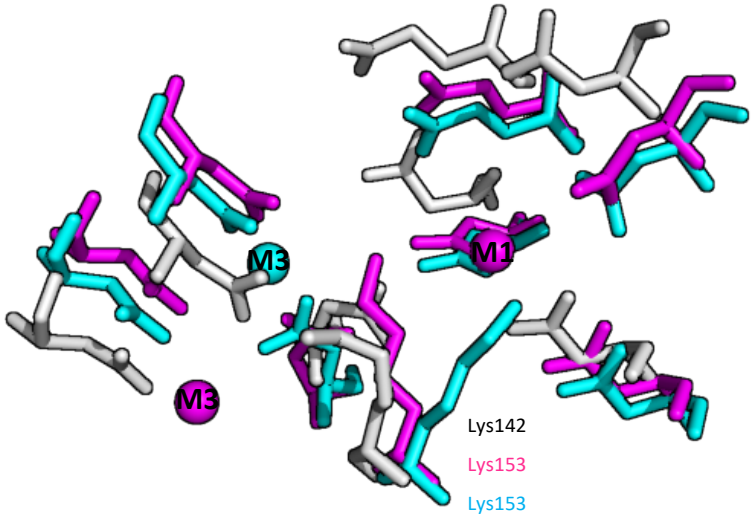


Figure 5-28: Active site alignment of with T5FEN

Active site residues from TaqFEN Asp142Lys:JT2+2 (chain A, shown in white) were aligned with the active site residues of the structures of T5FEN Asp153Lys (PDB: 5HML, shown in cyan) and T5FEN Asp153Lys:5OV4 (PDB: 5HNK, shown in magenta). Mg^{2+} ions are bound to the T5FEN active site on different positions in each structure and are shown as sphere of the corresponding colour.

This indicates that under the right conditions, divalent metal ions can still bind to the active site of an aspartic acid to lysine active-site mutant. Future work to capture the active site with bound metals may include studies with WT TaqFEN, as the active site should permit the binding of catalytic ions. Crystallization of the WT protein with DNA should be carried out in the presence of Ca^{2+}, so as to prevent the cleavage of DNA. Alternatively, higher concentrations of divalent salts could be included in setting up the crystal trials. Crystal

structures of the WT and active site mutant T5FEN which had metal ions bound to the active site were obtained with 100 mM $MgCl_2$ final concentration in the crystallization drop (AlMalki *et al.*, 2016).

5.3.4.4. The helix-3-turn helix motif

In the structures obtained here, a K^+ ion is bound to each molecule via the structurally-conserved H3TH motif and also interacts with the DNA phosphate backbone (Figure 5-12). Protein residues which co-ordinate the K^+ ion are Thr186, Leu185, Val196, Ile199. No K^+ ions are bound in the structure of Taq polymerase and structural comparisons with the H3TH motif show that two residues that interact with K+ move position by as much as ~8 Å to accommodate this interaction (Figure 5-29).

Potassium chloride has been shown to stabilize the binding of FENs to DNA including T5FEN and FEN-family member *E. coli* ExoIX (Garforth *et al.*, 2001; Anstey-Gilbert *et al.*, 2013), this is probably mediated, at least in part by binding to the H3TH to bridge the interaction between protein and DNA. A K^+ ion is observed bound to this motif in the crystal structures of multiple other FENs co-crystallized with DNA (Anstey-Gilbert *et al.*, 2013; AlMalki *et al.*, 2016; Tsutakawa *et al.*, 2017).

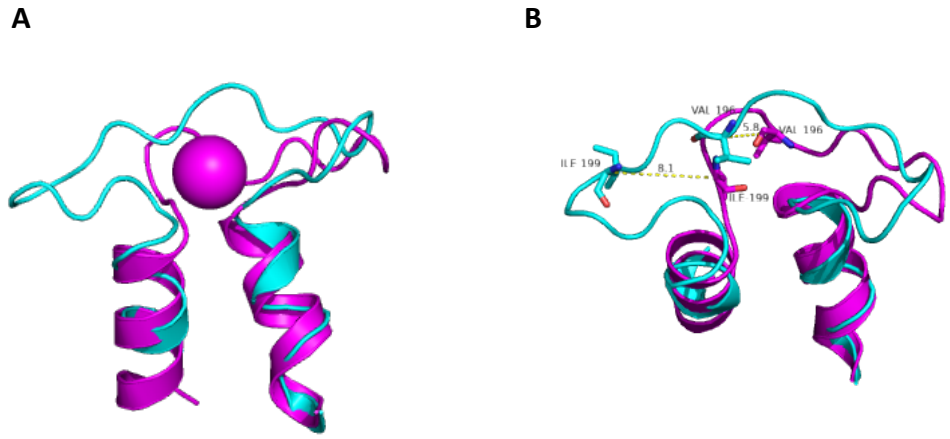


Figure 5-29: Comparison of the helix-3-turn-helix motif with Taq polymerase

The conformational differences between the helix-3-turn-helix (H3TH) motif of TaqFEN Asp142Lys:JT2+2 with Taq polymerase (PDB: 1TAQ). (A) Structural alignment of H3TH motifs (Magenta, TaqFEN Asp142Lys:JT2+2; cyan, Taq polymerase I). (B) Movement of VAL196 and ILE199 to co-ordinate the bound K^+ ion. Distances are shown in Å.

6.3.4.5. DNA interactions and proposed mechanism

In the models presented here, the DNA sits in close proximity to positively charged surface residues on TaqFEN (Figure 5-9). Several FEN:DNA co-crystal structures have been published and the most recent structural data suggests that FENs interact with 5' ssDNA by threading through a flexible archway motif (AlMalki *et al.*, 2016; Shi *et al.*, 2017; Tsutakawa *et al.*, 2017). To date, two 'fully-threaded' crystal structures have been published, threading is highly suggested in a third and in other structures DNA appears poised to enter through an opening in this region (Figure 5-31). The structure of T4 RNase H has been solved with a pseudo-Y DNA substrate. In this structure, extensive contacts are made between the 5' ssDNA and archway residues, strongly suggesting that there is threading, but not all of the residues corresponding to this motif were modelled (Devos *et al.*, 2007). Definitive crystallographic evidence of threading has been provided for T5FEN with an overhang substrate (PDB: 5HNK) and human FEN-1 with a double-flap substrate (PDB: 5KSE) (AlMalki *et al.*, 2016; Tsutakawa *et al.*, 2017). In these structures, the 5' ssDNA passes through an ordered helical archway. It has been proposed that FENs interact with 5' flaps through a 'disorder-thread-order' mechanism, where the archway adopts a disordered conformation to allow threading to occur and is then ordered for catalysis (Patel *et al.*, 2012). Whilst the structural evidence presented here cannot completely confirm this mechanism for TaqFEN, the position of DNA and interaction with the region above the active site do imply this.

In these structures, the 5' flap interacts with an Mg^{2+} / Ca^{2+} ion which is co-ordinated by Ser190 and water molecules. Furthermore, there are hydrogen bond interactions between the DNA backbone with Asn192 and Ser190 whilst His28 stacks with the base of nucleotide 3 (Figure 5-30). These interactions appear to enforce a looped conformation to the ssDNA. The ssDNA in other co-crystal structures also appears to adopt a looped conformation, and the DNA bases are pointed up away from the active site (see Figure 5-31 and Figure 5-32), although in these structures, metal ions are not involved in these interactions. In the threaded T5FEN (PDB: 5HNK) and human FEN-1 (PDB: 5KSE) structures metal ions are instead bound to active site residues. It may be that the TaqFEN:DNA structures presented in this chapter represent a 'pre-threaded' complex and the divalent ion may be displaced to the active site as the ssDNA is threaded. However, this binding of divalent ions could also represent a secondary or non-

specific binding site that has arisen due to the high concentrations used in crystallization (50 mM divalent salt was included in the protein crystallization buffer).

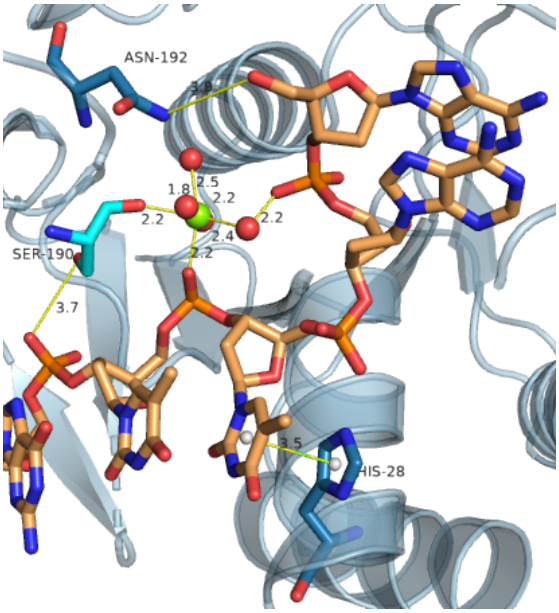


Figure 5-30: Examples of interactions involving the 5' flap

TaqFEN Asp142Lys: JT2+2 (chain B). Interactions between the protein chain (blue), oligonucleotide JT2+2 of the flap substrate (gold), Mg^{2+} ion (green sphere) and water molecules (red sphere) are shown. The Mg^{2+} ion is co-ordinated octahedrally and the interactions occur directly between protein or DNA and indirectly via water molecules. Asn192 and Ser190 form hydrogen bonds with DNA phosphate. His28 forms a stacking interaction with the DNA base. Interactions detected and figure prepared using the PLIP server (Adasme *et al.*, 2021).

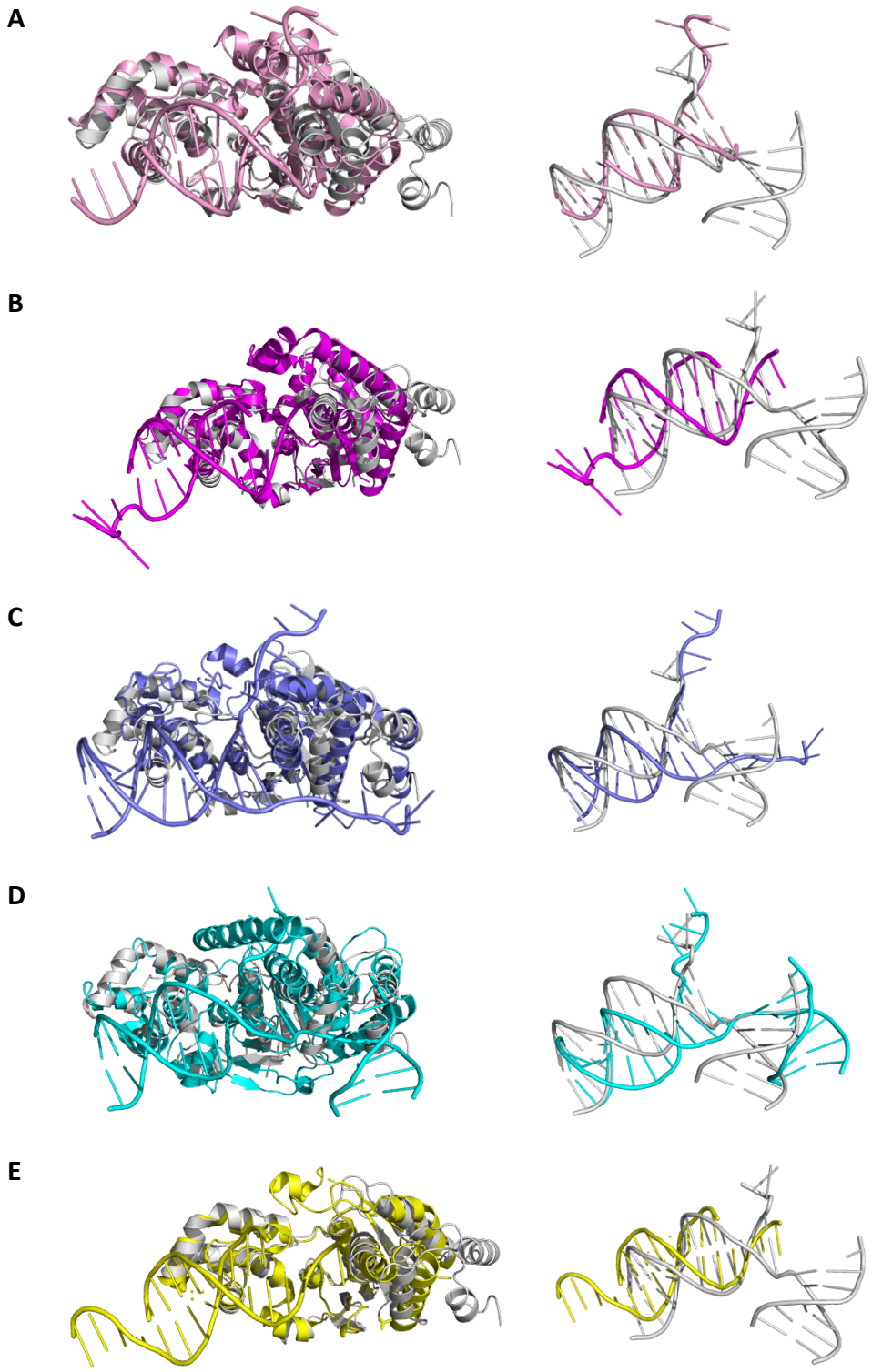


Figure 5-31: Structural comparison of protein: DNA interactions with other FEN structures

Structural alignment of TaqFEN Asp142Lys:JT2+2 (chain B, grey) with the structure of FENs bound to DNA substrates deposited in the PDB. In the left hand side images, the DNA from the aligned FEN is included. In the corresponding right hand side image, the aligned DNA from both molecules is shown. (A) T5FEN fully threaded complex (PDB: 5HNK, light pink). (B) T5FEN 'pre-threaded' complex (PDB: 5HNK, magenta). (C) T4 RNase H (PDB: 2IHN, purple). (D) Human FEN-1 (PDB: 5KSE, cyan). (E) *E.coli* ExoIX (PDB: 3ZDB).

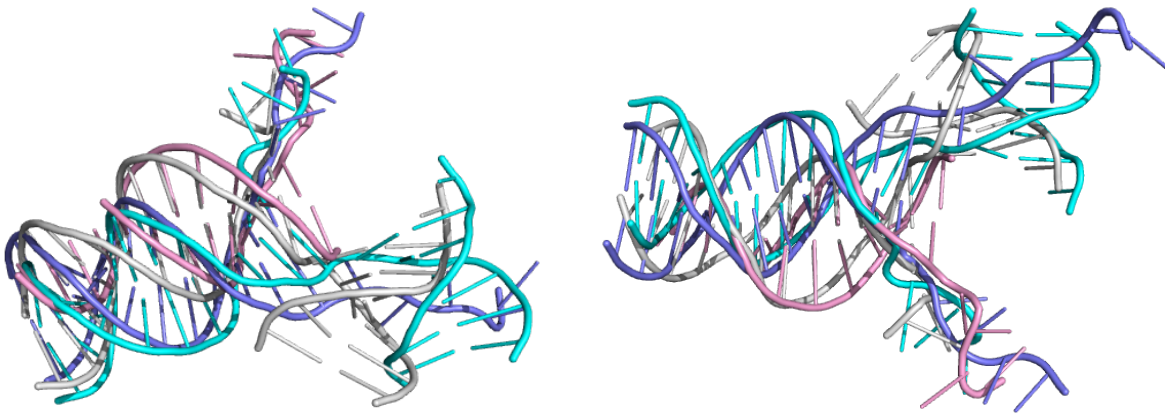


Figure 5-32: Structural compariosns of bound DNA

The structure of TaqFEN bound to JT2+2 was aligned to the co-crystal structures of human FEN-1 (PDB: 5KSE, cyan), T5FEN (PDB: 5HNK, pink), T4 RNaseH (PDB: 2IHN, purple), in which threading is implicated. The alignment of the DNA from these structures is shown.

5.3.5. Prediction of archway residue positions in TaqFEN

Dr Zied Hosni, a collaborator at the University of Sheffield used molecular dynamics (MD) simulations to predict possible conformations of the missing residues. The collaborator was provided with the final refined model of TaqFEN Asp142Lys: JT2+2 which had residues 70 - 88 removed on both chains. The structure of 10 results were returned for further analysis. These were structurally aligned with the original crystal structure provided in PyMOL and had an RMSD of 0.6 – 1.3 Å (Figure 5-33). As can be seen in the structural alignment, these missing residues appear to adopt a variety of conformations, indicating that this is a highly mobile region of the protein and this could explain why the corresponding electron density was not observed in the crystal structure as the electron density from any individual conformation would be represented at a fraction of that of the well-defined nearby backbone carbons.

The missing residues corresponding to the archway motif in other FENs and in the crystal structures of T5FEN and human FEN-1, DNA is observed passing through this. In the MD simulations presented here, the DNA here does not appear to traverse the equivalent residues. Instead, DNA is looped up and appears primed and positioned to enter, so this structure possible represents a ‘pre-threaded’ complex.

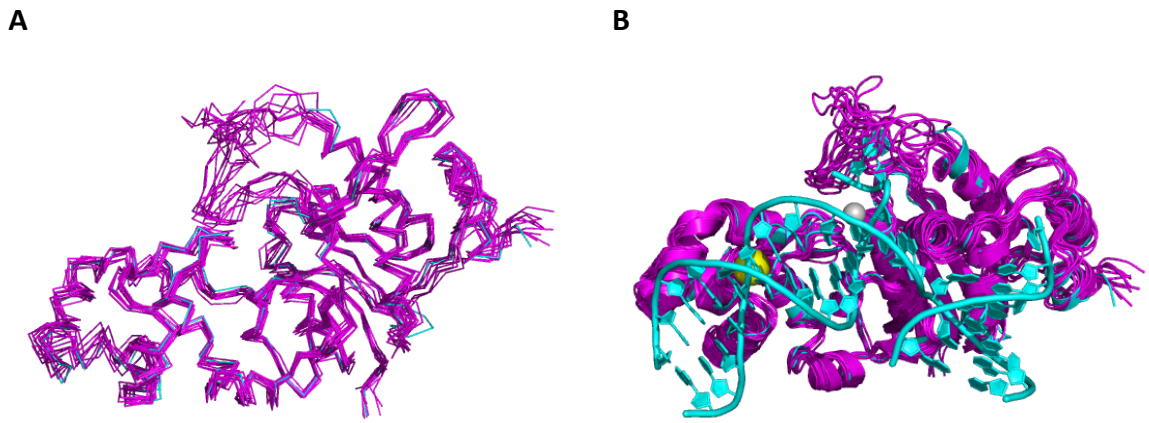


Figure 5-33: Molecular dynamic simulation of TaqFEN missing residues

The position of residues 70 - 88 of the TaqFEN Asp142Lys: JT2+2 structure was modelled by molecular dynamic (MD) simulations (Dr Z Hosni). (A) Structural alignment of protein molecules from the top 10 MD results (magenta ribbon) with the crystal structure with residues 70 – 88 missing (cyan ribbon). (B) Structural alignment of protein molecules of MD simulations (magenta cartoon) with protein and DNA of the crystal structure (cyan cartoon). Ions K^+ (yellow sphere) and Mg^{2+} (grey sphere) from the crystal structure are included in the alignment. All structures align with an RMSD between 0.6 – 1.3 Å.

5.3.6. Co-crystallization of T5FEN Asp155Lys with DNA

There is currently no published structure of T5FEN with full flap DNA substrate, rather the published structures are with overhang substrates. In order to capture upstream and downstream dsDNA interactions, crystal trials of T5FEN Asp155Lys with the JT2+2 flapped substrate were set. Whilst a total of five sub 2 Å datasets were obtained, no DNA was present in any of the structures. All of these datasets had very similar unit cell dimensions, were processed into the same space group and came from crystals which were obtained in very similar conditions, so it is not surprising that DNA is absent from all the structures. Regardless, the best dataset was obtained at 1.44 Å and is a novel structure of *apo* T5FEN Asp155Lys whilst also representing the highest resolution T5FEN data set compared to the published results and was therefore refined and presented in this thesis.

The substrate used for co-crystallization was the JT2+2 double-flap substrate consisting of a 4 nt 5’ flap and a 1 nt 3’ flap. Double-flap substrates have been successfully co-crystallized with TaqFEN (presented above) and human FEN-1 (Tsutakawa *et al.*, 2017). Structural studies of FENs have demonstrated that a 3’ binding pocket is conserved in FENs from higher organisms

but not in bacteriophages (Chapados *et al.*, 2004; Friedrich-Heineken and Hübscher, 2004; Finger *et al.*, 2009) and this may explain the difficulty in co-crystallizing this complex. However, studies have shown that the presence of a 3' flap appears to have little effect on the kinetics (K_M and k_{cat}) of a T5FEN reaction (Kaiser *et al.*, 1999). Furthermore, binding studies in T7FEN demonstrate that a single-flap and double-flap substrate are bound by approximately the same binding affinity (section 3.2.4.). However, these experiments were carried out in the presence of Ca^{2+}, which has been shown to enhance FEN: DNA interactions (Feng *et al.*, 2004).

This protein has previously been crystallized with a 3' overhang substrate in the presence of Ca^{2+} to generate an enzyme: pseudo product complex (AlMalki *et al.*, 2016). The use of Ca^{2+} may also help in this case. Adding DNA at a molar excess over protein may also help favour complex formation. Co-crystallization of the restriction endonuclease *Eco*RI with DNA was successful where DNA was at a 7-fold excess, although general recommendations suggest that DNA should be included at a 1.2- to 1.5-fold excess (Krauss *et al.*, 2013). Alternatively, other substrates may prove more successful, such as single-flap or nicked constructions. Techniques such as the electrophoretic mobility shift assay (EMSA) or biolayer interferometry, both of which allow determination of dissociation constants, could be useful to screen conditions and substrates which form the tightest complexes.

During the final stages of this project, preliminary investigation of co-crystallization with a single-flap substrate which contains a 1 nt nick in 3' upstream DNA was undertaken. A crystal from the screening plates diffracted to 3.2 Å. Molecular replacement with the T5FEN:5OV4 as the search model (PDB:5HNK) appeared to successfully place both protein and DNA within the density. Optimization was attempted to try and improve the resolution but has so far proved unsuccessful. However, this was probably due to the optimizations using a stock of protein that had been in -80°C storage for over 6 months prior to setting these plates, whilst initial screening was performed with freshly purified protein. Unfortunately, due to time constrains, this optimization was not re-attempted but this substrate could be promising for future work.

5.3.7. The structure of T5FEN Asp155Lys

The T5FEN Asp155Lys structure determined here aligns well to both WT T5FEN (PDB: 5HMM: RMSD = 0.236Å) and T5FEN Asp155Lys:3OV6:Ca^{2+} (PDB: 5HP4 RMSD = 0.442Å), with slight

differences observed in the flexible loop regions. In this structure, one Mg^{2+} ion is co-ordinated at M1 in binding site 1 (Figure 5-34), whilst Lys155 occupies the M2 binding site. The active site architecture observed here is very similar to that of Asp155Lys:3OV6:Ca^{2+}, except that Ca^{2+} ion occupies the M1 site.

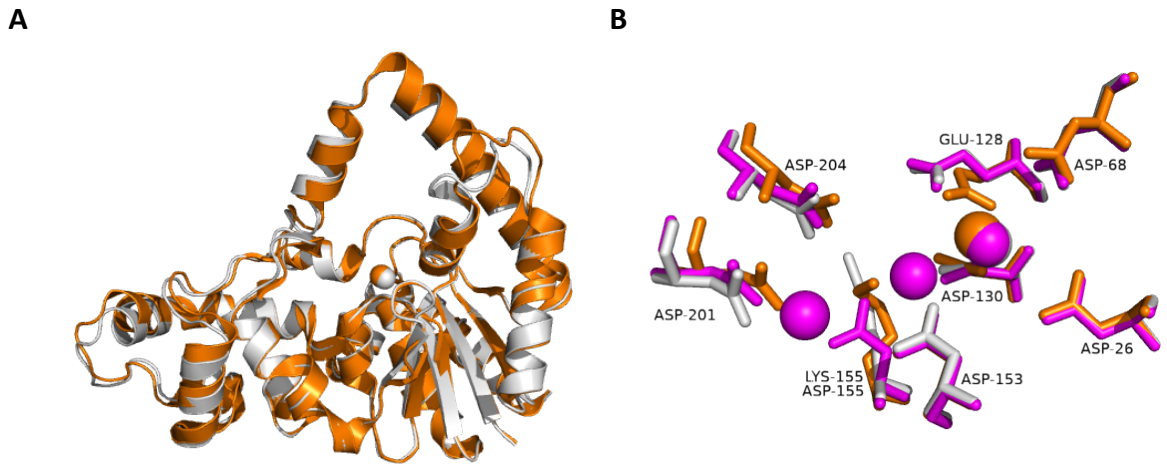


Figure 5-34: Comparison of T5FEN Asp155lys crystal structure

(A) Structural alignment of T5FEN Asp155Lys determined in this project (grey) with the published structure of T5FEN Asp155Lys determined in complex with DNA (PDB: 5HP4). The M1 site is occupied by an Mg^{2+} ion (grey sphere) or a Ca^{2+} ion (orange sphere) in the structures. (B) Active site residues of these are shown as well as WT T5FEN (PDB: 5HMM, magenta), where three Mg^{2+} ions are bound at sites M1 – M3.

The structures presented here show that many aspects of FEN-DNA interactions appear to be conserved between homologues. However, it is clear that there are differences between these structures which may be exploitable in engineering molecules for future optimizations of various molecular biological applications.

Chapter 6 – Discussion

6.1. Summary

In general, flap endonucleases are a well characterized family of proteins. The ability of these enzymes to cleave nucleic acids in a structure-specific manner, independent of sequence, has led to their exploitation by molecular biologists for over 30 years (Sayers *et al.*, 1988; Sayers *et al.*, 1996; Lyamichev *et al.*, 1999; Kiss-Toth *et al.*, 2001). Even today, use of these enzymes remains common-place in a molecular biology laboratory whilst more sophisticated applications have also been developed. Bacterial and bacteriophage FENs are the most popularly used within this field, likely due to higher levels of exonuclease activity compared to eukaryotic and archaeal counterparts (Lee and Wilson, 1999; Kaiser *et al.*, 1999; Williams *et al.*, 2007). Despite all living organisms and even some viruses possessing this protein, a few FEN homologues dominate the molecular biology and biotechnology landscape. For example, bacteriophage T5FEN has become an archetypal phage FEN since it was first purified from phage-infected cells in 1966 (Paul and Lehman, 1966). This protein has been extensively characterized both *in vivo* and *in vitro* and a variety of uses of this enzyme have been described including clean-up of DNA plasmid preparations, molecular cloning, aptamer selection, topoisomerase inhibitor screening and CRISPR-Cas fusion proteins (Sayers *et al.*, 1996; Alkhamis *et al.*, 2020; Wu *et al.*, 2020; Deng and Leng, 2021; Qu *et al.*, 2018). However, the knowledge surrounding other FEN homologues is more limited, and whilst this family share broadly comparable activities, there are some notable differences between individual proteins, as discussed in Chapter 1, including in their structures, such as the active site and archway architecture and in enzymatic activities such as divalent-metal-ion preferences and *in vitro* substrate specificities. Hence the choice of FEN is dependent on the individual application.

Biochemical and structural studies of T7 gp6 has been neglected relative to the bacteriophage T5 orthologue. Whilst early literature characterized the function of the enzyme *in vivo* and the exonuclease activity *in vitro*, only one primary research article has been published in the last 20 years which looks at both exonuclease and endonuclease activity of this protein (Mitsunobu *et al.*, 2014), which means more modern biochemical analyses have been omitted.

Despite this, T7 gp6 has several characterised uses in electrochemical detection methods (Xia *et al.*, 2020; Li *et al.*, 2021; Zhou *et al.*, 2021). The Binx Health IO point-of-care testing device utilizes T7 gp6 and Taq polymerase in a highly sensitive and specific nucleic-acid diagnostic test (Van Der Pol *et al.*, 2020). However, there is a large gap in application-specific knowledge regarding the use of this protein, both in the biochemistry of the protein and optimization of the use of it within the device. This led to the development of this PhD project in which the overarching aim was to improve the characterization of flap endonucleases in order to enhance molecular biology techniques as well as the IO device. This project looked mainly at T7 gp6 but also studied other relevant FENs to enable progression of certain areas.

Initially, wild-type T7 gp6 was produced and biochemically characterized using the UV, FRET and electrophoretic mobility shift assays (EMSA) (Chapter 3). The catalytic parameters were determined with the FRET assay using endonuclease and exonuclease substrates. This method was chosen as it gives real-time results, has improvements in safety over traditional ^{32}P methods and facilitates stable labelling to allow relatively long-term storage and standardisation of DNA substrates. The parameters determined *via* this method are comparable to published data with ^{32}P considering the impact of reaction conditions such as pH and cations on T7 gp6 reactions (Chapter 3) (Table 6-1). However, the obvious draw-back of this method is that only one cleavage event is measured and cleavage products are not visualised. Therefore, the actual turnover rate is likely to be higher than that reported here, especially for exonuclease activity. The UV and FRET assays were used to characterize optimum reaction conditions (pH, NaCl, KCl, divalent salts) and shows that T7 gp6 exhibits a marked sensitivity to these. Whilst the relevance of these finding to enhance the IO device and molecular biology techniques remains to be seen, this work serves as a good starting point for future studies in terms of enhancing speed or reducing the cost of goods. The binding affinity for the substrates was also determined. Whilst originally attempted using fluorescently-labelled DNA with an anisotropy assay, the lack of sensitivity and mechanical issues with the equipment excluded this method. Therefore, the EMSA with ^{32}P DNA was used to provide the sensitivity required. The low concentration of DNA (K_D >> [DNA]) in this technique is important to avoid the "titration regime" which can distort the calculated K_D by over a factor of 10 compared to the true value (Jarmoskaite *et al.*, 2020). This also means that to accurately calculate K_D, prior knowledge of K_D is required. Based on previous studies

involving T5FEN, the K_D was assumed to be in the nanomolar range (Dervan *et al.*, 2002), meaning that sub-nanomolar concentration of DNA was required. Whilst the calculated binding affinity determined in this thesis was up to 10-fold higher for T7 than that of T5FEN via fluorescent anisotropy and ^{32}P methods (Dervan *et al.*, 2002; Zhang, 2012), it is still comparable to the K_D of other FEN homologues which have been shown to range from nanomolar to micromolar affinity (Nolan *et al.*, 1996; Dervan *et al.*, 2002; Finger *et al.*, 2009; Zhang, 2012; Mitsunobu *et al.*, 2014; Oates, 2016).

Table 6-1: Binding and catalytic activity of T7 gp6 compared to previously published data

			Catalytic parameters			
	Method	**K_D (nM)**	**K_M (nM)**	**k_{cat} (min^{-1})**	**k_{cat}/K_M (min^{-1}.nM^{-1})**	**Substrate**
T7 gp6	FRET	ND	104.2 ± 5.7	6.89 ± 0.10	0.0662 ± 0.00374	Single-flap
	^{32}P[a]	ND	26.2[a]	0.57[a]	0.0217[a]	Overhang
	^{32}P	194 ± 11.1	ND	ND	ND	Single-flap
T5FEN	FRET	13.4[b]	32[b]	185[b]	6[b]	Overhang
	^{32}P	85 ± 6[c]	70 ± 10[c]	101 ± 10[c]	1.4[c]	Overhang
*Sp*FEN	FRET	ND	1100[d]	490[d]	0.449[d]	Single-flap
	BLI	1400[d]	ND	ND	ND	Overhang
hFEN-1	FC	7.5[e]	ND	6	ND	Single-flap
	dHPLC	ND	44[f]	164[f]	3.7[f]	Double-flap

a, (Mitsunobu *et al.*, 2014); b, (Zhang, 2012); c, (Dervan *et al.*, 2002); d, (Oates, 2016); e, (Nolan *et al.*, 1996); f, (Finger *et al.*, 2009). ND, data not reported; BLI, biolayer interferometry; FC, flow cytometry; dHPLC, denaturing high-performance liquid chromatography.

Mutations involving residues at and around the active site were investigated as a method to engineer T7 gp6 proteins with altered nuclease activity. The mutations studied in this thesis are shown in Figure 6-1. Catalytic activity was assessed via the UV and FRET assays and binding affinity via the EMSA (Table 6-2). Production of active-site mutants was performed in order to obtain an enzyme that is catalytically inert but able to bind DNA and therefore would be suitable to set crystal trials of the protein: substrate complex in the presence of catalytically relevant metal ions. Of the three active-site mutants generated, two were shown to be catalytically inert (Asp160Lys and Asp162Lys) whilst the other (Asp202Lys) retained low levels of nuclease activity (Chapter 4). This implies that the metal-ion binding sites composed of

active-site residues contribute differently to nuclease activity. This is largely in agreement with previous studies on T5FEN regarding the nuclease-dependence of active site residues, which indicated that only site I is required for endonuclease activity but both for exonuclease activity (Feng *et al.*, 2004). The T7 gp6 active site mutants bound DNA with a similar affinity compared to the WT protein, as determined by EMSA (Chapter 4). Previous work on T5FEN, showed that the equivalent mutants had an increased binding affinity (AlMalki *et al.*, 2016). More work should be done to confirm if this effect is due to the reaction conditions used, assay format or if it is enzyme specific. For example, analogous mutations in T7 gp6 may introduce a clash into the active site which could affect DNA binding. Alternatively, the reaction conditions used here could be close to the optimum for the enzyme so perhaps the mutations could offer little improvement.

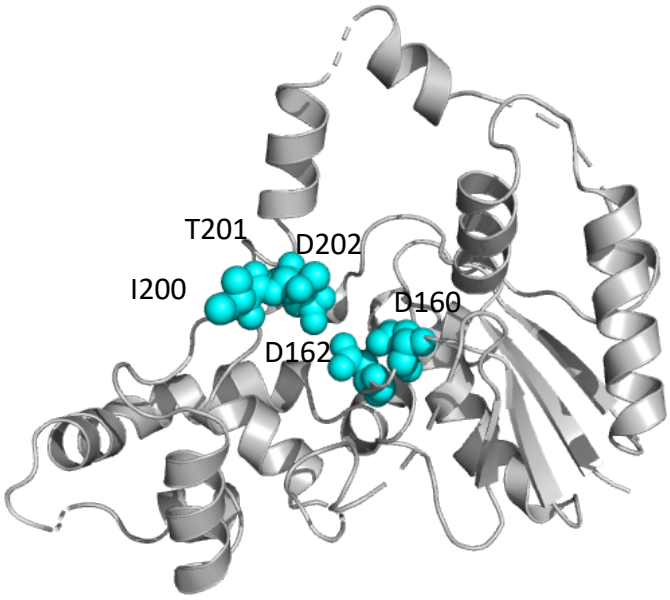


Figure 6-1: T7 gp6 mutations studied in this thesis

Residues that are mutated in this thesis are shown as cyan spheres on the predicted structure of T7 gp6. These residues are at and around the active site. The predicted structure of T7 gp6 was determined using the Phyre2 server (Kelley *et al.*, 2015).

Mutations within the helix-3-turn helix (H3TH) motif were investigated as way of altering the ratio of exonuclease and endonuclease activity. A mutation in this region was initially investigated in a previous study involving a chemical thiol modification of a cysteine residue in T5FEN at position 202 (residue 200 in T7 gp6), which resulted in a selective loss of endonuclease activity (Zhang, 2012). We set out to mimic this modification using amino acid substitutions, with amino acids with bulky side-chains, as chemical modification was not feasible due to T7 gp6 containing 8 cysteine residues. For this, two mutants were generated: Ile200Arg and Thr201Trp. For comparison, two new mutations corresponding to their T5FEN counterparts were also generated: Leu202Arg and Gly203Trp. Biochemical analysis of these

mutants indicated that Thr201Trp and Gly203Trp both exhibited significantly compromised exonuclease and endonuclease activity as determined by the UV and FRET assay. The activity of T7 gp6 Ile200Arg and Leu202Arg were more comparable to their WT counterpart and therefore kinetic parameters were determined (K_M, k_{cat}; Table 4-2). The specificity constant (k_{cat}/K_M) of Ile200Arg with all substrates is improved compared to the WT but this is not observed for T5 Leu202Arg. Binding studies in T7 gp6 indicated that Ile200Arg binds a single-flap substrate but not a nicked substrate with an increased affinity compared to the WT, although the binding affinities could not be determined for the T5FEN proteins. The tighter binding induced by the arginine substitution could be due to the introduction of a positive charge which interacts with the negatively-charged DNA phosphate backbone. The reasons for the differential effects in T7 gp6 and T5FEN warrant further investigation. Enzymes with altered selectivity may prove useful in the IO device and other molecular biology techniques due to a more specific activity and the Ile200Arg mutant appears to be an improved enzyme. However, whether this will translate through to improved techniques remains to be seen as planned studies at the Binx research facilities were cancelled due to the COVID-19 pandemic and it was not possible to reschedule the work within the time available.

Table 6-2: Binding and catalytic activity of T7 gp6 wild type and mutant proteins

A **Single-flap**

	K_D (nM)	Catalytic parameters		
		K_M (nM)	k_{cat} (min^{-1})	k_{cat}/K_M ($min^{-1}.nM^{-1}$)
Wild type	243 ± 53.0	104.2 ± 5.7	6.89 ± 0.10	0.0662 ± 0.00374
Asp160Lys	170 ± 77.7	ND	ND	ND
Asp162Lys	177 ± 17.2	ND	ND	ND
Asp202Lys	328 ± 121	ND	ND	ND
Ile200Arg	60.9 ± 16.9	22.1 ± 2.6	3.60 ± 0.183	0.162 ± 0.0208
Thr201Trp	>2000	ND	ND	ND

B **Nicked**

Protein	K_D (nM)	Catalytic parameters		
		K_M (nM)	k_{cat} (min^{-1})	k_{cat}/K_M ($min^{-1}.nM^{-1}$)
Wild type	194 ± 11.1	83.4 ± 3.2	14.7 ± 0.34	0.176 ± 0.00783
Asp160Lys	75.1 ± 34.4	ND	ND	ND
Asp162Lys	167 ± 52.6	ND	ND	ND
Asp202Lys	99.5 ± 29.4	ND	ND	ND
Ile200Arg	141 ± 29.1	324. 2 ± 54.0	110.4 ± 13.7	0.340 ± 0.0707
Thr201Trp	87.4 ± 11.9	ND	ND	ND

Errors indicate SEM (n=3); ND, result not determined.

Two inactive mutants (Asp160Lys, Asp162Lys) were taken forward with crystal trials in an attempt to solve the structure of T7 gp6. Having a protein structure could help identify residues which are important to the function and structure of the enzyme. By comparing structures of the protein alone and when bound to DNA, residues which interact with DNA or flexible regions could be identified which could aid structure-guided engineering of improved proteins. However, despite extensive screening >1000 conditions, no structure of T7 gp6 was obtained. To progress with the project, two homologous FENs were studied. These are the N-terminal FEN domain of *Thermus aquaticus* DNA polymerase I (TaqFEN) and T5FEN. Two novel TaqFEN:DNA structures are presented in this thesis (Chapter 5) and provide the first structural insights of the interactions of this homologue with DNA, which we hypothesize represents a 'pre-threaded' complex. Additionally, a novel high-resolution structure of the T5FEN active-site mutant (Asp155Lys) was obtained.

6.2. Future work

The main area of future work in relation to this project lies with the translation of the results into biotechnology applications. This project was originally developed to include a three-month placement within the laboratory of the industrial partner. The aim of this placement was to examine the impact of the optimum reaction conditions and generated mutants on the device as all the results presented here were obtained with laboratory substrates which possess a defined structure. Within the research laboratory at Binx Health, the individual stages of the IO device can be separated and the detection stage, which utilizes T7 gp6 would be examined. The reaction conditions and enzyme would be altered within this stage and parameters such as time till result and specificity would be measured.

To obtain the structure of T7 gp6 various approaches could be attempted. Mutations could be engineered, for example, N- or C-terminal truncations or replacing amino acids with large flexible side chains with smaller amino acids. Alternatively, various substrates could be investigated which include different lengths of dsDNA and single-stranded regions. Cysteine residues can cause oxidation sensitivity of proteins which results in a requirement for a reducing agent such as DTT throughout all stages of protein handling. A modified enzyme with no cysteines may prove more resilient, for both structural and biochemical studies. Whilst the T7 gp6 8CA cysteine-free mutant exhibits a severe defect in catalytic activity, mutation to alternative residues such as serine or threonine could be investigated.

Other future work could look at repurposing current molecular biology techniques with other understudied and underused homologues which may offer improvements in efficiency. Preliminary studies conducted during this project explored the use of T7 gp6 within a modified Gibson-style cloning technique. This method utilizes a low concentration of T5FEN to join overlapping DNA fragments without the requirement of DNA polymerase and DNA ligase, whilst also improving the efficiency of the original method (Xia *et al.*, 2019). In these preliminary studies, the *mCHERRY* gene was cloned into the pUC18 plasmid, so expression was under the control of the lactose-inducible bacteriophage T7 promoter. The gene was designed to have 20 nt overlapping ends with the multiple cloning sequence of the linearized plasmid.

These studies indicated that T7 gp6 was able to facilitate the cloning of this gene, as visualised by pink colonies and confirmed by colony PCR of the *mCHERRY* gene. The number of positive colonies obtained was comparable to that of T5FEN. More work should be carried out to calculate the efficiency of both these enzymes within the technique. Using different homologs within popular laboratory techniques may be of benefit as FENs with improved thermostability or altered pH-dependence could be more compatible with specific requirements of a technique. For example, T7 gp6 lacks GEN activity (Sayers *et al.*, 1988) and so may be suitable for Gibson cloning approaches. Alternatively, T5FEN proteins such as the Leu202Cys mutant chemically modified with thiol-reactive compounds (DTNB), which lack endonuclease activity may also prove more efficient.

To date, no co-crystal structure of a FEN with an exonuclease substrate, such as a nicked or blunt end dsDNA, has been obtained. Therefore, it would be interesting if future structural studies are directed towards this. Especially given that most molecular biological uses of FENs involve the exonuclease activity and these studies may identify important residues for this activity which could be altered to improve the enzyme.

6.3. Concluding remarks

Rapid diagnostics have already been shown to have a real impact. In response to the COVID-19 pandemic, more than 350 PCR testing kits are commercially available, so there is a huge societal need for optimized diagnostics (Garg *et al.*, 2020). Furthermore, a recent systematic review and meta-analysis including 5,920 patients found that when combined with antimicrobial stewardship, molecular rapid diagnostic tests decreased mortality by 36% in patients with blood stream infections, reduced the time to therapy by 5.03 hours and length of hospital stay by 2.48 days (Timbrook *et al.*, 2017). Rapid point-of-care (PoC) detection of bacterial infection is important as it provides information regarding antibiotic resistance and treatment options, enabling the correct treatment to be administered correctly as early as possible. This is important as studies have indicated that up to >80% of antibiotics that are prescribed are unnecessary or inappropriate (Luyt *et al.*, 2014; Afari-Asiedu *et al.*, 2020). There is a correlation between antibiotic use and drug resistance (Goossens *et al.*, 2005; Tan *et al.*, 2015; Yang *et al.*, 2018), so there is an ever-growing need for improved antibiotic stewardship

since the threat of antibiotic resistance is becoming an ever more pressing issue. This is especially important as The World Health Organization (2019) warns that the current antibiotics in clinical development are not sufficient to combat the rise in antimicrobial resistance.

This work contributes to the body of knowledge surrounding FENs and provides a starting point for optimization reaction conditions for use in the IO device and other molecular biology techniques. Optimization of techniques such as diagnostics is important for multiple reasons. The first is that the cost of recombinant proteins is high, so if reaction conditions can be altered which result in less enzyme required, then manufacturing costs will be reduced, whilst time to result could also be decreased. The global polymerase chain reaction global market is predicted to be worth USD 25.3 billion by 2028 (Grand View Research, 2021), therefore the use of optimized conditions or engineered enzymes could represent a large saving and or a commercial advantage.

References

Adasme, M.F., Linnemann, K.L., Bolz, S.N., Kaiser, F., Salentin, S., Haupt, V.J., Schroeder, M. (2021) PLIP 2021: expanding the scope of the protein–ligand interaction profiler to DNA and RNA. *Nucleic Acids Research*. 49(W1), W530–W534.

Afari-Asiedu, S., Oppong, F.B., Tostmann, A., Ali Abdulai, M., Boamah-Kaali, E., Gyaase, S., Agyei, O., Kinsman, J., Hulscher, M., Wertheim, H.F.L., Asante, K.P. (2020) Determinants of Inappropriate Antibiotics Use in Rural Central Ghana Using a Mixed Methods Approach. *Frontiers in Public Health*. 8.

Alkhamis, O., Yang, W., Farhana, R., Yu, H., Xiao, Y. (2020) Label-free profiling of DNA aptamer-small molecule binding using T5 exonuclease. *Nucleic Acids Research*. 48(20), e120–e120.

Allawi, H.T., Dahlberg, J.E., Olson, S., Lund, E., Olson, M., Ma, W.-P., Takova, T., Neri, B.P., Lyamichev, V.I. (2004) Quantitation of microRNAs using a modified Invader assay. *RNA*. 10(7), 1153–1161.

Allen, L.M., Hodskinson, M.R.G., Sayers, J.R. (2009) Active site substitutions delineate distinct classes of eubacterial flap endonuclease. *Biochemical Journal*. 418(Pt 2), 285–292.

AlMalki, F.A., Flemming, C.S., Zhang, J., Feng, M., Sedelnikova, S.E., Ceska, T., Rafferty, J.B., Sayers, J.R., Artymiuk, P.J. (2016) Direct observation of DNA threading in flap endonuclease complexes. *Nature Structural & Molecular Biology*. 23(7), 640–646.

Anstey-Gilbert, C.S., Hemsworth, G.R., Flemming, C.S., Hodskinson, M.R.G., Zhang, J., Sedelnikova, S.E., Stillman, T.J., Sayers, J.R., Artymiuk, P.J. (2013) The structure of Escherichia coli ExoIX—implications for DNA binding and catalysis in flap endonucleases. *Nucleic Acids Research*. 41(17), 8357–8367.

Balakrishnan, L., Bambara, R.A. (2013) Okazaki Fragment Metabolism. *Cold Spring Harbor Perspectives in Biology*. 5(2), a010173–a010173.

Barnes, C.J., Wahl, A.F., Shen, B., Park, M.S., Bambara, R.A. (1996) Mechanism of Tracking and Cleavage of Adduct-damaged DNA Substrates by the Mammalian 5′- to 3′-Exonuclease/Endonuclease RAD2 Homologue 1 or Flap Endonuclease 1*. *Journal of Biological Chemistry*. 271(47), 29624–29631.

Bayliss, C.D., Sweetman, W.A., Moxon, E.R. (2005) Destabilization of tetranucleotide repeats in Haemophilus influenzae mutants lacking RnaseHI or the Klenow domain of PolI. *Nucleic Acids Research*. 33(1), 400–408.

Benchling (2021) *Benchling* [Online] Available from: https://benhling.com [Accessed December 22, 2021

Bhagwat, Medha, Hobbs, L.J., Nossal, N.G. (1997) The 5′-Exonuclease Activity of Bacteriophage T4 RNase H Is Stimulated by the T4 Gene 32 Single-stranded DNA-binding Protein, but Its Flap Endonuclease Is Inhibited*. *Journal of Biological Chemistry*. 272(45), 28523–28530.

Bhagwat, M., Meara, D., Nossal, N.G. (1997) Identification of residues of T4 RNase H required for catalysis and DNA binding. *The Journal of Biological Chemistry*. 272(45), 28531–28538.

Bishop, J.S., Guy-Caffey, J.K., Ojwang, J.O., Smith, S.R., Hogan, M.E., Cossum, P.A., Rando, R.F., Chaudhary, N. (1996) Intramolecular G-quartet Motifs Confer Nuclease Resistance to a Potent Anti-HIV Oligonucleotide (∗). *Journal of Biological Chemistry*. 271(10), 5698–5703.

Bock, L.C., Griffin, L.C., Latham, J.A., Vermaas, E.H., Toole, J.J. (1992) Selection of single-stranded DNA molecules that bind and inhibit human thrombin. *Nature*. 355(6360), 564–566.

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

Castellazzi, M., Brachet, P., Eisen, H. (1972) Isolation and characterization of deletions in bacteriophage λ residing as prophage in E. coli K12. *Molecular and General Genetics MGG*. 117(3), 211–218.

Center, M.S., Studier, F.W., Richardson, C.C. (1970) The Structural Gene for a T7 Endonuclease Essential for Phage DNA Synthesis*. *Proceedings of the National Academy of Sciences of the United States of America*. 65(1), 242–248.

Ceska, T.A., Sayers, J.R., Stier, G., Suck, D. (1996) A helical arch allowing single-stranded DNA to thread through T5 5’-exonuclease. *Nature*. 382(6586), 90–93.

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

Choi, B., Rempala, G.A., Kim, J.K. (2017) Beyond the Michaelis-Menten equation: Accurate and efficient estimation of enzyme kinetic parameters. *Scientific Reports*. 7(1), 17018.

Cowtan, K. (2006) The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallographica Section D: Biological Crystallography*. 62(9), 1002–1011.

Dale, G.E., Oefner, C. (2003) The protein as a variable in protein crystallization. *J. Struct. Biol*, 88–97.

DeJesus, M.A., Gerrick, E.R., Xu, W., Park, S.W., Long, J.E., Boutte, C.C., Rubin, E.J., Schnappinger, D., Ehrt, S., Fortune, S.M., Sassetti, C.M., Ioerger, T.R. (2017) Comprehensive Essentiality Analysis of the Mycobacterium tuberculosis Genome via Saturating Transposon Mutagenesis. *mBio*. 8(1), e02133-16.

Deller, M.C., Kong, L., Rupp, B. (2016) Protein stability: a crystallographer’s perspective. *Acta Crystallographica. Section F, Structural Biology Communications*. 72(Pt 2), 72–95.

Deng, Z., Leng, F. (2021) A T5 Exonuclease-Based Assay for DNA Topoisomerases and DNA Intercalators. *ACS Omega*. 6(18), 12205–12212.

Derewenda, Z.S. (2004) Rational Protein Crystallization by Mutational Surface Engineering. *Structure*. 12(4), 529–535.

Dervan, J.J., Feng, M., Patel, D., Grasby, J.A., Artymiuk, P.J., Ceska, T.A., Sayers, J.R. (2002) Interactions of mutant and wild-type flap endonucleases with oligonucleotide substrates suggest an alternative model of DNA binding. *Proc Natl Acad Sci U S A*. 99(13), 8542–7.

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

Dianov, G., Lindahl, T. (1994) Reconstitution of the DNA base excision-repair pathway. *Current biology: CB*. 4(12), 1069–1076.

Díaz, A., Lacks, S.A., López, P. (1992) The 5′ to 3′ exonuclease activity of DNA polymerase I is essential for Streptococcus pneumoniae. *Molecular Microbiology*. 6(20), 3009–3019.

Dong, A., Xu, X., Edwards, A.M. (2007) In situ proteolysis for protein crystallization and structure determination. *Nature Methods*. 4(12), 1019–1021.

Emsley, P., Lohkamp, B., Scott, W.G., Cowtan, K. (2010) Features and development of Coot. *Acta Crystallographica Section D: Biological Crystallography*. 66(Pt 4), 486–501.

Feng, M., Patel, D., Dervan, J.J., Ceska, T., Suck, D., Haq, I., Sayers, J.R. (2004) Roles of divalent metal ions in flap endonuclease-substrate interactions. *Nat Struct Mol Biol*. 11(5), 450–6.

Finger, L.D., Blanchard, M.S., Theimer, C.A., Sengerová, B., Singh, P., Chavez, V., Liu, F., Grasby, J.A., Shen, B. (2009) The 3′-Flap Pocket of Human Flap Endonuclease 1 Is Critical for Substrate Binding and Catalysis*. *Journal of Biological Chemistry*. 284(33), 22184–22194.

Fraser, M.J. (1980) [33] Purification and properties of Neurospora crassa Endo-exonuclease, an Enzyme which can be converted to a single-strand specific endonuclease. In *Methods in Enzymology*. Nucleic Acids Part I. Academic Press, pp. 255–263.

Frenkel, G.D., Richardson, C.C. (1971) The Deoxyribonuclease Induced after Infection of Escherichia coli by Bacteriophage T5: I. CHARACTERIZATION OF THE ENZYME AS A 5′-EXONUCLEASE. *Journal of Biological Chemistry*. 246(15), 4839–4847.

Friedrich-Heineken, E., Hübscher, U. (2004) The Fen1 extrahelical 3′-flap pocket is conserved from archaea to human and regulates DNA substrate specificity. *Nucleic Acids Res*. 32(8), 2520–8.

Fukushima, S., Itaya, M., Kato, H., Ogasawara, N., Yoshikawa, H. (2007) Reassessment of the In Vivo Functions of DNA Polymerase I and RNase H in Bacterial Cell Growth. *Journal of Bacteriology*. 189(23), 8575–8583.

Garforth, S.J., Ceska, T.A., Suck, D., Sayers, J.R. (1999) Mutagenesis of conserved lysine residues in bacteriophage T5 5′-3′ exonuclease suggests separate mechanisms of endo-and exonucleolytic cleavage. *Proc Natl Acad Sci U S A*. 96(1), 38–43.

Garforth, S.J., Patel, D., Feng, M., Sayers, J.R. (2001) Unusually wide co-factor tolerance in a metalloenzyme; divalent metal ions modulate endo–exonuclease activity in T5 exonuclease. *Nucleic Acids Research*. 29(13), 2772–2779.

Garforth, S.J., Sayers, J.R. (1997) Structure-specific DNA binding by bacteriophage T5 5'-->3' exonuclease. *Nucleic Acids Research*. 25(19), 3801–3807.

Garg, A., Ghoshal, U., Patel, S.S., Singh, D.V., Arya, A.K., Vasanth, S., Pandey, A., Srivastava, N. (2020) Evaluation of seven commercial RT-PCR kits for COVID-19 testing in pooled clinical specimens. *Journal of Medical Virology*, 10.1002/jmv.26691.

Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M.R., Appel, R.D., Bairoch, A. (2005) Protein Identification and Analysis Tools on the ExPASy Server. In J. M. Walker, ed. *The Proteomics Protocols Handbook*. Totowa, NJ: Humana Press, pp. 571–607.

Giacalone, M.J., Gentile, A.M., Lovitt, B.T., Berkley, N.L., Gunderson, C.W., Surber, M.W. (2006) Toxic protein expression in Escherichia coli using a rhamnose-based tightly regulated and tunable promoter system. *BioTechniques*. 40(3), 355–364.

Gibson, D.G., Young, L., Chuang, R.-Y., Venter, J.C., Hutchison, C.A., Smith, H.O. (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature Methods*. 6(5), 343–345.

Goossens, H., Ferech, M., Vander Stichele, R., Elseviers, M., ESAC Project Group (2005) Outpatient antibiotic use in Europe and association with resistance: a cross-national database study. *Lancet*. 365(9459), 579–87.

Grand View Research (2021) Polymerase Chain Reaction Market Size Report, 2028. [online]. Available from: https://www.grandviewresearch.com/industry-analysis/polymerase-chain-reaction-market?utm_source=prnewswire&utm_medium=referral&utm_campaign=hc_17-nov-20&utm_term=polymerase-chain-reaction-market&utm_content=rd [Accessed January 7, 2022].

Green, L.S., Jellinek, D., Bell, C., Beebe, L.A., Feistner, B.D., Gill, S.C., Jucker, F.M., Janjić, N. (1995) Nuclease-resistant nucleic acid ligands to vascular permeability factor/vascular endothelial growth factor. *Chemistry & Biology*. 2(10), 683–695.

Green, M.R., Sambrook, J. (2019) Isolation of DNA Fragments from Polyacrylamide Gels by the Crush and Soak Method. *Cold Spring Harbor Protocols*. 2019(2), pdb.prot100479.

Groll, D.H., Jeltsch, A., Selent, U., Pingoud, A. (1997) Does the Restriction Endonuclease EcoRV Employ a Two-Metal-Ion Mechanism for DNA Cleavage? *Biochemistry*. 36(38), 11389–11401.

Grossman, T.H., Kawasaki, E.S., Punreddy, S.R., Osburne, M.S. (1998) Spontaneous cAMP-dependent derepression of gene expression in stationary phase plays a role in recombinant expression instability. *Gene*. 209(1–2), 95–103.

Hakem, R. (2008) DNA-damage repair; the good, the bad, and the ugly. *The EMBO Journal*. 27(4), 589–605.

Hall, J.G., Eis, P.S., Law, S.M., Reynaldo, L.P., Prudent, J.R., Marshall, D.J., Allawi, H.T., Mast, A.L., Dahlberg, J.E., Kwiatkowski, R.W., de Arruda, M., Neri, B.P., Lyamichev, V.I. (2000) Sensitive detection of DNA polymorphisms by the serial invasive signal amplification reaction. *Proceedings of the National Academy of Sciences of the United States of America*. 97(15), 8272–8277.

Harrington, J.J., Lieber, M.R. (1994) The characterization of a mammalian DNA structure-specific endonuclease. *The EMBO Journal*. 13(5), 1235–1246.

Hassell, A.M., An, G., Bledsoe, R.K., Bynum, J.M., Carter, H.L., Deng, S.-J.J., Gampe, R.T., Grisard, T.E., Madauss, K.P., Nolte, R.T., Rocque, W.J., Wang, L., Weaver, K.L., Williams, S.P., Wisely, G.B., Xu, R., Shewchuk, L.M. (2007) Crystallization of protein–ligand complexes. *Acta Crystallographica Section D: Biological Crystallography*. 63(Pt 1), 72–79.

Hengen, P.N. (1996) Methods and reagents: Eliminating ghost bands from plasmid preps. *Trends in Biochemical Sciences*. 21(11), 441–442.

Hobbs, L.J., Nossal, N.G. (1996) Either bacteriophage T4 RNase H or Escherichia coli DNA polymerase I is essential for phage replication. *Journal of Bacteriology*. 178(23), 6772–6777.

Hosfield, D.J., Mol, C.D., Shen, B., Tainer, J.A. (1998) Structure of the DNA repair and replication endonuclease and exonuclease FEN-1: coupling DNA and PCNA binding to FEN-1 activity. *Cell*. 95(1), 135–46.

Hwang, K.Y., Baek, K., Kim, H.Y., Cho, Y. (1998) The crystal structure of flap endonuclease-1 from Methanococcus jannaschii. *Nat Struct Biol*. 5(8), 707–13.

Jancarik, J., Kim, S.-H. (1991) Sparse matrix sampling: a screening method for crystallization of proteins. *Journal of Applied Crystallography*. 24(4), 409–411.

Jarmoskaite, I., AlSadhan, I., Vaidyanathan, P.P., Herschlag, D. (2020) How to measure and evaluate binding affinities S. Deindl & J. Kuriyan, eds. *eLife*. 9, e57264.

Joyce, C.M., Grindley, N.D. (1984) Method for determining whether a gene of Escherichia coli is essential: application to the polA gene. *Journal of Bacteriology*. 158(2), 636–643.

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S.A.A., Ballard, A.J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., Back, T., Petersen, S., Reiman, D., Clancy, E., Zielinski, M., Steinegger, M., Pacholska, M., Berghammer, T., Bodenstein, S., Silver, D., Vinyals, O., Senior, A.W., Kavukcuoglu, K., Kohli, P., Hassabis, D. (2021) Highly accurate protein structure prediction with AlphaFold. *Nature*. 596(7873), 583–589.

Kaiser, M.W., Lyamicheva, N., Ma, W., Miller, C., Neri, B., Fors, L., Lyamichev, V.I. (1999) A comparison of eubacterial and archaeal structure-specific 5’-exonucleases. *J Biol Chem*. 274(30), 21387–94.

Kao, H.I., Henricksen, L.A., Liu, Y., Bambara, R.A. (2002) Cleavage specificity of Saccharomyces cerevisiae flap endonuclease 1 suggests a double-flap structure as the cellular substrate. *J Biol Chem*. 277(17), 14379–89.

Keefe, A.D., Pai, S., Ellington, A. (2010) Aptamers as therapeutics. *Nature Reviews Drug Discovery*. 9(7), 537–550.

Kelley, L.A., Mezulis, S., Yates, C.M., Wass, M.N., Sternberg, M.J. (2015) The Phyre2 web portal for protein modeling, prediction and analysis. *Nat Protoc*. 10(6), 845–58.

Kerr, C., Sadowski, P.D. (1972) Gene 6 Exonuclease of Bacteriophage T7: I. PURIFICATION AND PROPERTIES OF THE ENZYME. *Journal of Biological Chemistry*. 247(1), 305–310.

Kim, Y., Eom, S.H., Wang, J., Lee, D.S., Suh, S.W., Steitz, T.A. (1995) Crystal structure of Thermus aquaticus DNA polymerase. *Nature*. 376(6541), 612–6.

Kisker, C., Kuper, J., Van Houten, B. (2013) Prokaryotic Nucleotide Excision Repair. *Cold Spring Harbor Perspectives in Biology*. 5(3), a012591.

Kiss-Toth, E., Dower, S.K., Sayers, J.R. (2001) A Method for Enhancing the Transfection Efficiency of Minipreps Obtained from Plasmid cDNA Libraries. *Analytical Biochemistry*. 288(2), 230–232.

Klungland, A., Lindahl, T. (1997) Second pathway for completion of human DNA base excision-repair: reconstitution with purified proteins and requirement for DNase IV (FEN1). *EMBO J*. 16(11), 3341–8.

Konrad, E.B., Lehman, I.R. (1974) A conditional lethal mutant of Escherichia coli K12 defective in the 5’ leads to 3’ exonuclease associated with DNA polymerase I. *Proceedings of the National Academy of Sciences of the United States of America*. 71(5), 2048–2051.

Krauss, I.R., Merlino, A., Vergara, A., Sica, F. (2013) An Overview of Biological Macromolecule Crystallization. *International Journal of Molecular Sciences*. 14(6), 11643–11691.

Kucherlapati, M., Yang, K., Kuraguchi, M., Zhao, J., Lia, M., Heyer, J., Kane, M.F., Fan, K., Russell, R., Brown, A.M.C., Kneitz, B., Edelmann, W., Kolodner, R.D., Lipkin, M., Kucherlapati, R. (2002) Haploinsufficiency of Flap endonuclease (Fen1) leads to rapid tumor progression. *Proceedings of the National Academy of Sciences*. 99(15), 9924–9929.

de Laat, W.L., Jaspers, N.G., Hoeijmakers, J.H. (1999) Molecular mechanism of nucleotide excision repair. *Genes & Development*. 13(7), 768–785.

Lau, R. (2017) Characterisation of the streptococcal DNA polymerase I flap endonuclease domain. *PhD thesis, The University of Sheffield*.

Lechner, R.L., Engler, M.J., Richardson, C.C. (1983) Characterization of strand displacement synthesis catalyzed by bacteriophage T7 DNA polymerase. *The Journal of Biological Chemistry*. 258(18), 11174–11184.

Ledford, M., Friedman, K.D., Hessner, M.J., Moehlenkamp, C., Williams, T.M., Larson, R.S. (2000) A Multi-Site Study for Detection of the Factor V (Leiden) Mutation from Genomic DNA Using a Homogeneous Invader Microtiter Plate Fluorescence Resonance Energy Transfer (FRET) Assay. *The Journal of Molecular Diagnostics*. 2(2), 97–104.

Lee, B.-I., Wilson, D.M. (1999) The RAD2 Domain of Human Exonuclease 1 Exhibits 5′ to 3′ Exonuclease and Flap Structure-specific Endonuclease Activities*. *Journal of Biological Chemistry*. 274(53), 37763–37769.

Lee, M., Miller, R.C. (1974) T7 exonuclease (gene 6) is necessary for molecular recombination of bacteriophage T7. *J Virol*. 14(5), 1040–8.

Li, S., Fu, Z., Wang, C., Shang, X., Zhao, Y., Liu, C., Pei, M. (2021) An ultrasensitive and specific electrochemical biosensor for DNA detection based on T7 exonuclease-assisted regulatory strand displacement amplification. *Analytica Chimica Acta*. 1183, 338988.

Lin, T., Zhang, L., Jiang, D., Wu, L., Chen, K., Li, L., Jin, C., Li, Z., Oger, P. (2022) Biochemical characterization and mutational analysis of a novel flap endonuclease 1 from Thermococcus barophilus Ch5. *The International Journal of Biochemistry & Cell Biology*, 106154.

Liu, R., Qiu, J., Finger, L.D., Zheng, L., Shen, B. (2006) The DNA–protein interaction modes of FEN-1 with gap substrates and their implication in preventing duplication mutations. *Nucleic Acids Research*. 34(6), 1772–1784.

Liu, Y., Kao, H.I., Bambara, R.A. (2004) Flap endonuclease 1: a central component of DNA metabolism. *Annu Rev Biochem*. 73, 589–615.

Livak, K.J., Flood, S.J., Marmaro, J., Giusti, W., Deetz, K. (1995) Oligonucleotides with fluorescent dyes at opposite ends provide a quenched probe system useful for detecting PCR product and nucleic acid hybridization. *Genome Research*. 4(6), 357–362.

Luyt, C.E., Bréchot, N., Trouillet, J.L., Chastre, J. (2014) Antibiotic stewardship in the intensive care unit. *Crit Care*. 18(5), 480.

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

Lyamichev, V., Brow, M.A.D., Varvel, V.E., Dahlberg, J.E. (1999) Comparison of the 5′ nuclease activities of Taq DNA polymerase and its isolated nuclease domain. *Proceedings of the National Academy of Sciences*. 96(11), 6143–6148.

Lyamichev, Victor, Mast, A.L., Hall, J.G., Prudent, J.R., Kaiser, M.W., Takova, T., Kwiatkowski, R.W., Sander, T.J., de Arruda, M., Arco, D.A., Neri, B.P., Brow, M.A.D. (1999) Polymorphism identification and quantitative detection of genomic DNA by invasive cleavage of oligonucleotide probes. *Nature Biotechnology*. 17(3), 292–296.

Marras, S.A., Kramer, F.R., Tyagi, S. (2002) Efficiencies of fluorescence resonance energy transfer and contact-mediated quenching in oligonucleotide probes. *Nucleic Acids Res*. 30(21), e122.

Marshall, C.J., Santangelo, T.J. (2020) Archaeal DNA Repair Mechanisms. *Biomolecules*. 10(11), 1472.

McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., Read, R.J. (2007) Phaser crystallographic software. *Journal of Applied Crystallography*. 40(4), 658–674.

Mitsunobu, H., Zhu, B., Lee, S.J., Tabor, S., Richardson, C.C. (2014) Flap endonuclease activity of gene 6 exonuclease of bacteriophage T7. *J Biol Chem*. 289(9), 5860–75.

Mitsunobu, Hitoshi, Zhu, B., Lee, S.-J., Tabor, S., Richardson, C.C. (2014b) Flap endonuclease of bacteriophage T7: Possible roles in RNA primer removal, recombination and host DNA breakdown. *Bacteriophage*. 4, e28507.

Moyer, R.W., Rothe, C.T. (1977) Role of the T5 gene D15 nuclease in the generation of nicked bacteriophage T5 DNA. *Journal of Virology*. 24(1), 177–193.

Mueser, T.C., Nossal, N.G., Hyde, C.C. (1996) Structure of bacteriophage T4 RNase H, a 5' to 3' RNA-DNA and DNA-DNA exonuclease with sequence similarity to the RAD2 family of eukaryotic proteins. *Cell*. 85(7), 1101–12.

Murante, R.S., Rust, L., Bambara, R.A. (1995) Calf 5' to 3' Exo/Endonuclease Must Slide from a 5' End of the Substrate to Perform Structure-specific Cleavage (∗). *Journal of Biological Chemistry*. 270(51), 30377–30383.

Murshudov, G.N., Skubák, P., Lebedev, A.A., Pannu, N.S., Steiner, R.A., Nicholls, R.A., Winn, M.D., Long, F., Vagin, A.A. (2011) REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallographica Section D: Biological Crystallography*. 67(Pt 4), 355–367.

Nakai, H., Richardson, C.C. (1988) The effect of the T7 and Escherichia coli DNA-binding proteins at the replication fork of bacteriophage T7. *Journal of Biological Chemistry*. 263(20), 9831–9839.

Nakamaye, K.L., Eckstein, F. (1986) Inhibition of restriction endonuclease Nci I cleavage by phosphorothioate groups and its application to oligonucleotide-directed mutagenesis. *Nucleic Acids Research*. 14(24), 9679–9698.

Nolan, J.P., Shen, B., Park, M.S., Sklar, L.A. (1996) Kinetic Analysis of Human Flap Endonuclease-1 by Flow Cytometry. *Biochemistry*. 35(36), 11668–11676.

Oates, S. (2016) Molecular Characterisation of the Trypanosoma brucei Flap Endonuclease. *PhD thesis, The University of Sheffield*.

O'Connor, C.D., Timmis, K.N. (1987) Highly repressible expression system for cloning genes that specify potentially toxic proteins. *Journal of Bacteriology*. 169(10), 4457–4462.

Ogawa, T., Okazaki, T. (1980) Discontinuous DNA Replication. *Annual Review of Biochemistry*. 49(1), 421–457.

Ogawa, T., Okazaki, T. (1984) Function of RNase H in DNA replication revealed by RNase H defective mutants of Escherichia coli. *Molecular and General Genetics MGG*. 193(2), 231–237.

Ozaki, K., Ohnishi, Y., Iida, A., Sekine, A., Yamada, R., Tsunoda, T., Sato, Hiroshi, Sato, Hideyuki, Hori, M., Nakamura, Y., Tanaka, T. (2002) Functional SNPs in the lymphotoxin-α gene that are associated with susceptibility to myocardial infarction. *Nature Genetics*. 32(4), 650–654.

Pan, W., Craven, R.C., Qiu, Q., Wilson, C.B., Wills, J.W., Golovine, S., Wang, J.F. (1995) Isolation of virus-neutralizing RNAs from a large pool of random sequences. *Proceedings of the National Academy of Sciences of the United States of America*. 92(25), 11509–11513.

Parrish, J.Z., Yang, C., Shen, B., Xue, D. (2003) CRN-1, a Caenorhabditis elegans FEN-1 homologue, cooperates with CPS-6/EndoG to promote apoptotic DNA degradation. *EMBO J*. 22(13), 3451–60.

Patel, D., Tock, M.R., Frary, E., Feng, M., Pickering, T.J., Grasby, J.A., Sayers, J.R. (2002) A Conserved Tyrosine Residue Aids Ternary Complex Formation, but not Catalysis, in Phage T5 Flap Endonuclease. *Journal of Molecular Biology*. 320(5), 1025–1035.

Patel, N., Atack, J.M., Finger, L.D., Exell, J.C., Thompson, P., Tsutakawa, S., Tainer, J.A., Williams, D.M., 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(10), 4507–19.

Paul, A.V., Lehman, I.R. (1966) The deoxyribonucleases of Escherichia coli. VII. A deoxyribonuclease induced by infection with phage T-5. *J Biol Chem*. 241(14), 3441–51.

Pickering, T.J., Garforth, S, Sayers, J.R., Grasby, J.A. (1999) Variation in the steady state kinetic parameters of wild type and mutant T5 5'-3'-exonuclease with pH. Protonation of Lys-83 is critical for DNA binding. *J Biol Chem*. 274(25), 17711–7.

Pickering, T.J., Garforth, S J, Thorpe, S.J., Sayers, J.R., Grasby, J.A. (1999) A single cleavage assay for T5 5'-->3' exonuclease: determination of the catalytic parameters forwild-type and mutant proteins. *Nucleic Acids Research*. 27(3), 730–735.

Qu, B., Ni, Y., Lempp, F.A., Vondran, F.W.R., Urban, S. (2018) T5 Exonuclease Hydrolysis of Hepatitis B Virus Replicative Intermediates Allows Reliable Quantification and Fast Drug Efficacy Testing of Covalently Closed Circular DNA by PCR. *Journal of Virology*. 92(23), e01117-18.

Remaut, E., Stanssens, P., Fiers, W. (1981) Plasmid vectors for high-efficiency expression controlled by the PL promoter of coliphage lambda. *Gene*. 15(1), 81–93.

Rosano, G.L., Ceccarelli, E.A. (2014) Recombinant protein expression in Escherichia coli: advances and challenges. *Frontiers in Microbiology*. 5, 172.

Rosenthal, A.L., Lacks, S.A. (1977) Nuclease detection in SDS-polyacrylamide gel electrophoresis. *Analytical Biochemistry*. 80(1), 76–90.

Ruscito, A., DeRosa, M.C. (2016) Small-Molecule Binding Aptamers: Selection Strategies, Characterization, and Applications. *Frontiers in Chemistry*. 4.

Rye, C., Wise, R., Jurukovski, V., DeSaix, J., Choi, J., Avissar, Y. (2016) *Biology*. Houston, Texas: OpenStax.

Sadowski, P.D., Kerr, C. (1970) Degradation of Escherichia coli B deoxyribonucleic acid after infection with deoxyribonucleic acid-defective amber mutants of bacteriophage T7. *J Virol*. 6(2), 149–55.

Sancar, A. (1996) DNA excision repair. *Annual Review of Biochemistry*. 65, 43–81.

Saro, F.J.L. de, O'Donnell, M. (2001) Interaction of the β sliding clamp with MutS, ligase, and DNA polymerase I. *Proceedings of the National Academy of Sciences*. 98(15), 8376–8380.

Sayers, J.R. (1994) Computer Aided Identification of a Potential 5'-3' Exonuclease Gene Encoded by Escherichia coli. *Journal of Theoretical Biology*. 170(4), 415–421.

Sayers, J.R., Eckstein, F. (1991) A single-strand specific endonuclease activity copurifies with overexpressed T5 D15 exonuclease. *Nucleic Acids Research*. 19(15), 4127–4132.

Sayers, J.R., Eckstein, F. (1990) Properties of overexpressed phage T5 D15 exonuclease. Similarities with Escherichia coli DNA polymerase I 5'-3' exonuclease. *Journal of Biological Chemistry*. 265(30), 18311–18317.

Sayers, J.R., Evans, D., Thomson, J.B. (1996) Identification and Eradication of a Denatured DNA Isolated during Alkaline Lysis-Based Plasmid Purification Procedures. *Analytical Biochemistry*. 241(2), 186–189.

Sayers, J.R., Krekel, C., Eckstein, F. (1992) Rapid high-efficiency site-directed mutagenesis by the phosphorothioate approach. *Biotechniques*. 13(4), 592–6.

Sayers, J.R., Schmidt, W., Eckstein, F. (1988) 5'-3' exonucleases in phosphorothioate-based oligonucleotide-directed mutagenesis. *Nucleic Acids Res*. 16(3), 791–802.

Schomacher, L., Chong, J.P.J., McDermott, P., Kramer, W., Fritz, H.-J. (2009) DNA uracil repair initiated by the archaeal ExoIII homologue Mth212 via direct strand incision. *Nucleic Acids Research*. 37(7), 2283–2293.

Schutzbank, T.E., Jarvis, C., Kahmann, N., Lopez, K., Weimer, M., Yount, A. (2007) Detection of High-Risk Papillomavirus DNA with Commercial Invader-Technology-Based Analyte-Specific Reagents following Automated Extraction of DNA from Cervical Brushings in ThinPrep Media. *Journal of Clinical Microbiology*. 45(12), 4067–4069.

Schwede, T.F., Bädeker, M., Langer, M., Rétey, J., Schulz, G.E. (1999) Homogenization and crystallization of histidine ammonia-lyase by exchange of a surface cysteine residue. *Protein Engineering*. 12(2), 151–153.

Sengerová, B., Tomlinson, C., Atack, J.M., Williams, R., Sayers, J.R., Williams, N.H., Grasby, J.A. (2010) Brønsted Analysis and Rate-Limiting Steps for the T5 Flap Endonuclease Catalyzed Hydrolysis of Exonucleolytic Substrates. *Biochemistry*. 49(37), 8085–8093.

Serwer, P., Watson, R.H., Son, M. (1990) Role of gene 6 exonuclease in the replication and packaging of bacteriophage T7 DNA. *J Mol Biol*. 215(2), 287–99.

Shen, B., Nolan, J.P., Sklar, L.A., Park, M.S. (1997) Functional analysis of point mutations in human flap endonuclease-1 active site. *Nucleic Acids Research*. 25(16), 3332–3338.

Shi, Y., Hellinga, H.W., Beese, L.S. (2017) Interplay of catalysis, fidelity, threading, and processivity in the exo- and endonucleolytic reactions of human exonuclease I. *Proc Natl Acad Sci U S A*. 114(23), 6010–6015.

Shin, Y.-K., Amangyeld, T., Nguyen, T.A., Munashingha, P.R., Seo, Y.-S. (2012) Human MUS81 complexes stimulate flap endonuclease 1. *The FEBS Journal*. 279(13), 2412–2430.

Shivji, M.K., Podust, V.N., Hübscher, U., Wood, R.D. (1995) Nucleotide excision repair DNA synthesis by DNA polymerase epsilon in the presence of PCNA, RFC, and RPA. *Biochemistry*. 34(15), 5011–5017.

Singh, A., Upadhyay, V., Upadhyay, A.K., Singh, S.M., Panda, A.K. (2015) Protein recovery from inclusion bodies of Escherichia coli using mild solubilization process. *Microbial Cell Factories*. 14(1), 41.

Singh, P., Zheng, L., Chavez, V., Qiu, J., Shen, B. (2007) Concerted Action of Exonuclease and Gap-dependent Endonuclease Activities of FEN-1 Contributes to the Resolution of Triplet Repeat Sequences (CTG) - and (GAA) -derived Secondary Structures Formed during Maturation of Okazaki Fragments. *Journal of Biological Chemistry*. 282(6), 3465–3477.

Speers, D.J. (2006) Clinical applications of molecular biology for infectious diseases. *Clin Biochem Rev*. 27(1), 39–51.

Studier, F.W. (1969) The genetics and physiology of bacteriophage T7. *Virology*. 39(3), 562–574.

Syson, K., Tomlinson, C., Chapados, B.R., Sayers, J.R., Tainer, J.A., Williams, N.H., Grasby, J.A. (2008) Three metal ions participate in the reaction catalyzed by T5 flap endonuclease. *J Biol Chem*. 283(42), 28741–6.

Tan, C.-K., Tang, H.-J., Lai, C.-C., Chen, Y.-Y., Chang, P.-C., Liu, W.-L. (2015) Correlation between antibiotic consumption and carbapenem-resistant Acinetobacter baumannii causing health care-associated infections at a hospital from 2005 to 2010. *Journal of Microbiology, Immunology and Infection*. 48(5), 540–544.

Till, M., Robson, A., Byrne, M.J., Nair, A.V., Kolek, S.A., Shaw Stewart, P.D., Race, P.R. (2013) Improving the Success Rate of Protein Crystallization by Random Microseed Matrix Screening. *Journal of Visualized Experiments : JoVE*. (78), 50548.

Timbrook, T.T., Morton, J.B., McConeghy, K.W., Caffrey, A.R., Mylonakis, E., LaPlante, K.L. (2017) The Effect of Molecular Rapid Diagnostic Testing on Clinical Outcomes in Bloodstream Infections: A Systematic Review and Meta-analysis. *Clin Infect Dis*. 64(1), 15–23.

Tishkoff, D.X., Boerger, A.L., Bertrand, P., Filosi, N., Gaida, G.M., Kane, M.F., Kolodner, R.D. (1997) Identification and characterization of Saccharomyces cerevisiae EXO1, a gene encoding an exonuclease that interacts with MSH2. *Proc Natl Acad Sci U S A*. 94(14), 7487–92.

Tock, M.R., Frary, E., Sayers, J.R., Grasby, J.A. (2003) Dynamic evidence for metal ion catalysis in the reaction mediated by a flap endonuclease. *The EMBO Journal*. 22(5), 995–1004.

Tsutakawa, S.E., Classen, S., Chapados, B.R., Arvai, A.S., Finger, L.D., Guenther, G., Tomlinson, C.G., Thompson, P., Sarker, A.H., Shen, B., Cooper, P.K., Grasby, J.A., Tainer, J.A. (2011) Human flap endonuclease structures, DNA double-base flipping, and a unified understanding of the FEN1 superfamily. *Cell*. 145(2), 198–211.

Tsutakawa, S.E., Thompson, M.J., Arvai, A.S., Neil, A.J., Shaw, S.J., Algasaier, S.I., Kim, J.C., Finger, L.D., Jardine, E., Gotham, V.J.B., Sarker, A.H., Her, M.Z., Rashid, F., Hamdan, S.M., Mirkin, S.M., Grasby, J.A., Tainer, J.A. (2017) Phosphate steering by Flap Endonuclease 1 promotes 5'-flap specificity and incision to prevent genome instability. *Nat Commun*. 8, 15855.

Uson, M.L., Ghosh, S., Shuman, S. (2017) The DNA Repair Repertoire of Mycobacterium smegmatis FenA Includes the Incision of DNA 5' Flaps and the Removal of 5' Adenylylated Products of Aborted Nick Ligation. *Journal of Bacteriology*. 199(17).

Vagin, A., Teplyakov, A. (2010) Molecular replacement with MOLREP. *Acta Crystallographica Section D: Biological Crystallography*. 66(1), 22–25.

Vaisman, A., McDonald, J.P., Noll, S., Huston, D., Loeb, G., Goodman, M.F., Woodgate, R. (2014) Investigating the mechanisms of ribonucleotide excision repair in Escherichia coli. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*. 761, 21–33.

Valdez-Cruz, N.A., Caspeta, L., Pérez, N.O., Ramírez, O.T., Trujillo-Roldán, M.A. (2010) Production of recombinant proteins in E. coli by the heat inducible expression system based on the phage lambda pL and/or pR promoters. *Microbial Cell Factories*. 9(1), 18.

Van Der Pol, B., Taylor, S.N., Mena, L., Lebed, J., McNeil, C.J., Crane, L., Ermel, A., Sukhija-Cohen, A., Gaydos, C.A. (2020) Evaluation of the Performance of a Point-of-Care Test for Chlamydia and Gonorrhea. *JAMA Netw Open*. 3(5), e204819.

Vinograd, J., Lebowitz, J. (1966) Physical and Topological Properties of Circular DNA. *The Journal of General Physiology*. 49(6), 103–125.

Wanner, B.L., Kodaira, R., Neidhardt, F.C. (1978) Regulation of lac operon expression: reappraisal of the theory of catabolite repression. *Journal of Bacteriology*. 136(3), 947–954.

WHO (2019) 2019 antibacterial agents in clinical development: an analysis of the antibacterial clinical development pipeline. [online]. Available from: https://www.who.int/publications-detail-redirect/9789240000193 [Accessed January 15, 2022].

Williams, C.J., Headd, J.J., Moriarty, N.W., Prisant, M.G., Videau, L.L., Deis, L.N., Verma, V., Keedy, D.A., Hintze, B.J., Chen, V.B., Jain, S., Lewis, S.M., Arendall III, W.B., Snoeyink, J., Adams,

P.D., Lovell, S.C., Richardson, J.S., Richardson, D.C. (2018) MolProbity: More and better reference data for improved all-atom structure validation. *Protein Science*. 27(1), 293–315.

Williams, R., Sengerová, B., Osborne, S., Syson, K., Ault, S., Kilgour, A., Chapados, B.R., Tainer, J.A., Sayers, J.R., Grasby, J.A. (2007) Comparison of the catalytic parameters and reaction specificities of a phage and an archaeal flap endonuclease. *Journal of molecular biology*. 371(1), 34–48.

Wood, R.D. (1997) Nucleotide Excision Repair in Mammalian Cells *. *Journal of Biological Chemistry*. 272(38), 23465–23468.

Wu, Y., Yuan, Q., Zhu, Y., Gao, X., Song, J., Yin, Z. (2020) Improving FnCas12a Genome Editing by Exonuclease Fusion. *The CRISPR Journal*. 3(6), 503–511.

Xia, L.-Y., Li, M.-J., Wang, H.-J., Yuan, R., Chai, Y.-Q. (2020) Novel Single-Enzyme-Assisted Dual Recycle Amplification Strategy for Sensitive Photoelectrochemical MicroRNA Assay. *Analytical Chemistry*. 92(21), 14550–14557.

Xia, Y., Li, K., Li, J., Wang, T., Gu, L., Xun, L. (2019) T5 exonuclease-dependent assembly offers a low-cost method for efficient cloning and site-directed mutagenesis. *Nucleic Acids Research*. 47(3), e15–e15.

Xu, Y., Potapova, O., Leschziner, A.E., Grindley, N.D., Joyce, C.M. (2001) Contacts between the 5’ nuclease of DNA polymerase I and its DNA substrate. *J Biol Chem*. 276(32), 30167–77.

Yang, P., Chen, Y., Jiang, S., Shen, P., Lu, X., Xiao, Y. (2018) Association between antibiotic consumption and the rate of carbapenem-resistant Gram-negative bacteria from China based on 153 tertiary hospitals data in 2014. *Antimicrobial Resistance & Infection Control*. 7(1), 137.

Zauli, D.A.G. (2019) *PCR and Infectious Diseases*. IntechOpen.

Zhang, J. (2012) Biochemical Studies on T5 Exonuclease. *PhD thesis, The University of Sheffield*.

Zhang, Q., Yin, K., Liu, G., Li, S., Li, M., Qiu, J.-L. (2020) Fusing T5 exonuclease with Cas9 and Cas12a increases the frequency and size of deletion at target sites. *Science China Life Sciences*. 63(12), 1918–1927.

Zhao, G., Tang, S., Li, J., Hu, T., Guan, Y. (2014) Effects of cations on small fragment of DNA polymerase I using a novel FRET assay. *Acta Biochim Biophys Sin (Shanghai)*. 46(8), 659–67.

Zheng, H., Cooper, D.R., Porebski, P.J., Shabalin, I.G., Handing, K.B., Minor, W. (2017) CheckMyMetal: a macromolecular metal-binding validation tool. *Acta Crystallographica. Section D, Structural Biology*. 73(Pt 3), 223–233.

Zheng, L., Jia, J., Finger, L.D., Guo, Z., Zer, C., Shen, B. (2011) Functional regulation of FEN1 nuclease and its link to cancer. *Nucleic Acids Research*. 39(3), 781–794.

Zheng, L., Li, M., Shan, J., Krishnamoorthi, R., Shen, B. (2002) Distinct Roles of Two Mg + Binding Sites in Regulation of Murine Flap Endonuclease-1 Activities. *Biochemistry*. 41(32), 10323–10331.

Zheng, L., Zhou, M., Chai, Q., Parrish, J., Xue, D., Patrick, S.M., Turchi, J.J., Yannone, S.M., Chen, D., Shen, B. (2005) Novel function of the flap endonuclease 1 complex in processing stalled DNA replication forks. *EMBO reports*. 6(1), 83–89.

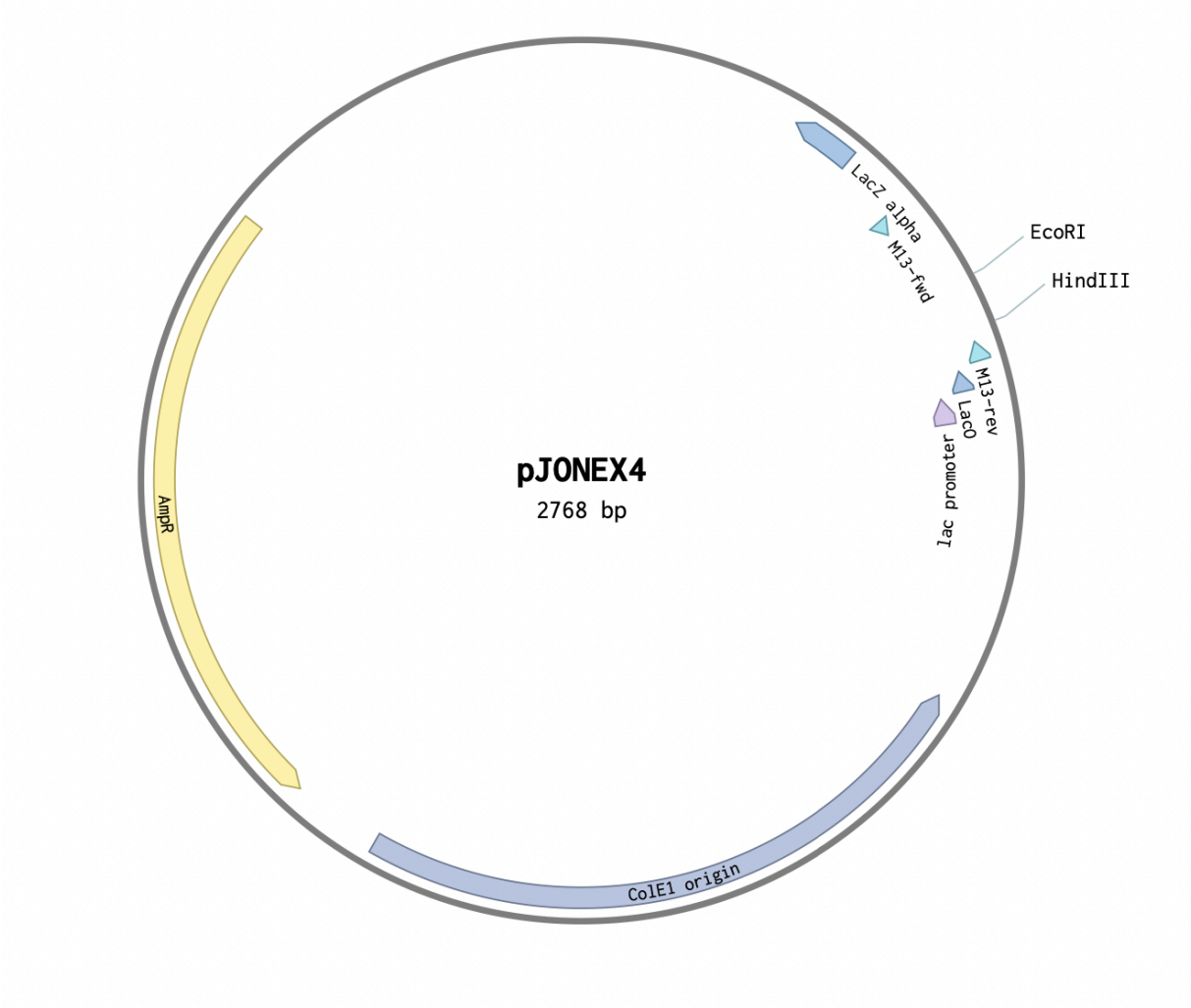
Zhou, Q.-Y., Ma, R.-N., Hu, C.-L., Sun, F., Jia, L.-P., Zhang, W., Shang, L., Xue, Q.-W., Jia, W.-L., Wang, H.-S. (2021) A novel ratiometric electrochemical biosensing strategy based on T7 exonuclease-assisted homogenous target recycling coupling hairpin assembly-triggered double-signal output for the multiple amplified detection of miRNA. *The Analyst*. 146(8), 2705–2711.

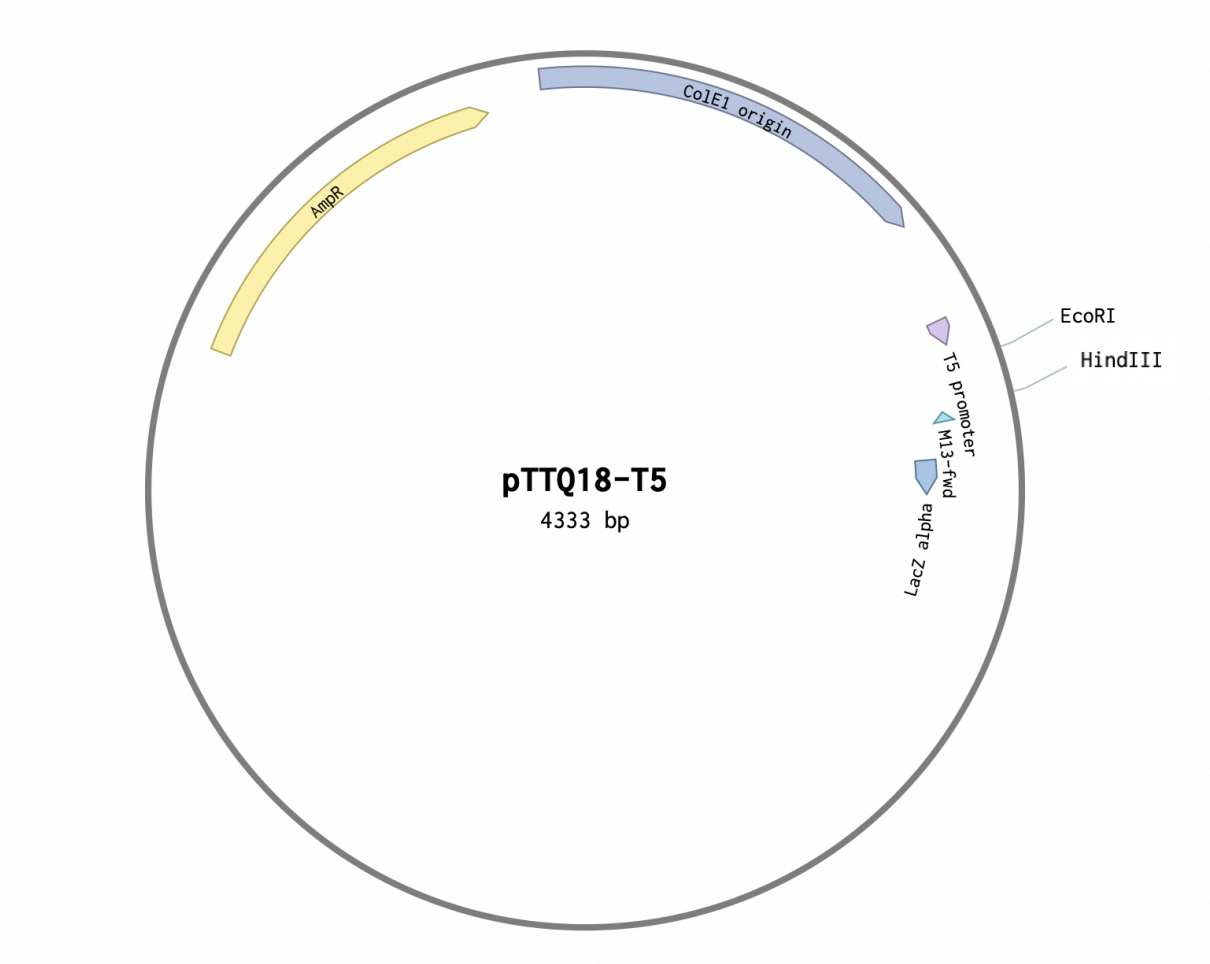
Zoller, M.J., Smith, M. (1982) Oligonucleotide-directed mutagenesis using M13-derived vectors: an efficient and general procedure for the production of point mutations in any fragment of DNA. *Nucleic Acids Research*. 10(20), 6487–6500.

Appendix A: Vector maps

Prepared using Benchling (Benchling, 2021).

Restriction sites used for cloning are shown on the plasmid maps.




ColE1 origin
AmpR
EcoRI
HindIII
T5 promoter
M13-fwd
lacZ alpha
pTTQ18-T5
4333 bp