The University Of Sheffield.

Structural and inhibition studies on Human Flap Endonuclease

BY:
Stavroula Louka

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

August, 2021

Abstract

Flap endonuclease 1 (FEN1) is a paradigm of structure-specific metallonuclease family that selectively catalyzes the removal of single-stranded 5′-flaps, occurring *in vivo* primarily in Okazaki fragment maturation and DNA repair. Enzymatic malfunction of FEN1 has dramatic consequences for eukaryotic genome stability. Since overexpression of FEN1 has been reported in multiple cancers, FEN1 inhibition is proposed to have a therapeutic potential. Without any clinical trials related to FEN1 inhibition registered to date, small molecule inhibitors are now recognized as possible leads for the development of novel anticancer drugs. Aiming at the identification of small molecule FEN1 inhibitors at an *in silico* stage, a dual approach of virtual and physical screening was followed. Manipulation of active site metal charges led to a well-performing docking strategy under +2 active site metal charges that can effectively predict the great majority of *in vitro* confirmed hits and used for cost-effective FEN1 inhibitor identification. In addition, following a collaborative hit expansion study an inhibitor of HsFEN1 with IC_{50} of ~1.7 μM was identified for future inhibitor development studies, expanding the reported to date range of FEN1 inhibitors. Crystallographic analyses also led to two HsFEN1:product DNA complex structures supporting the dsDNA substrate's ability to bend 100°, "trapping" of the 1-nt long 3′-flap and positioning of 5′-flap towards the active site because of HsFEN1 binding at the DNA bend. The presence of $CaCl_2$ allowed some enzymatic activity and facilitated crystallization of a product complex whose structure was determined at 2.3 Å. Whilst this project focuses on human FEN1, microbial FEN inhibitors identified by the University of Sheffield (UoS) spin-out DeFENition Ltd primarily for antibiotics development were examined within the context of the current thesis for their induced cytotoxicity. Although structure-activity relationships fell outside the scope of the current thesis, this first ever collected cytotoxicity dataset provided key information for future toxicity studies and downstream compound development.

Acknowledgements

It only takes one over numerous experiential stories, published in Nature's "The PhD" collection, to realize that each PhD comes with its own challenges and its own highs and lows. Each PhD reflects a unique drama and/or success story, carrying its own backdrop of stress, struggles and resilience. Who thought, 4-years earlier, that this PhD, along with many others of course, would have been submitted in the middle of a global pandemic crisis? Who thought that access to the laboratory facilities would have been highly limited in the final year of this PhD project, where most students typically speed up and intensify to plug empty holes of their research? Who can blame a PhD student for worrying about thesis submission, novelty of data, writing quality, etc., while we are witnessing history being made?

COVID-19 has definitely overturned the world. While we are all being adapted to the new era of "social distancing", I have been lucky enough to be supported by people, whose help has been critical for the completion of this PhD. People, who I have been pleased to meet and collaborate with, and who I would like to particularly thank.

A big thank you goes to my supervisors, Dr Helen E. Bryant and Prof Jon R. Sayers, who supported me ever since the very beginning, and the University of Sheffield for funding this PhD project through one of its internal scholarships. Special thanks to you Jon, for all the support, guidance and skills developed throughout my PhD studies. I am thankful for every single opportunity you offered me, which, despite the final year's struggles, kept me motivated, stimulated my curiosity and kept reminding me why I choose to undertake a PhD at first place. Thanks for your patience and understanding in numerous occasions.

Also, thanks to everyone in Sayers' group. Each one of you has taught me something different. Of particular mention, I would like to thank Dr Sarbendra L. Pradhananga, who has been a great support inside and outside lab (especially thanks for the data analysis, interpretation and GraphPad Prism tips!). Also, thanks to Dr Domen Zafred, whose help has been critical for crystallizing and solving the structures of HsFEN1. Thank you to our research technician, Mrs Janine Phipps, who in addition to her help, has always been an incredible and cheerful human being. Thank you to the PhD students, Emma Brudenell, Fiona Macleod and Arturo Landeros. It has been a pleasure working in the same lab you.

Thank you to DeFENition Ltd, and its founding director Prof Jon R. Sayers, for the high-throughput screening of small-molecule compounds and for funding the final part of my PhD project on the toxicity studies of FEN1 inhibitors. Special thanks to Dr Yannick Lacroix and Dr Jacqueline Fok, with whom I collaborated for the data analysis and interpretation.

Thank you to everyone from the X-ray crystallography lab in the Department of Molecular Biology and Biotechnology, University of Sheffield, for sharing their facilities and providing technical support. Special thanks to Dr Sam R. Dix and Prof Patrick J. Baker for structure refinement support.

Thank you to Dr Gian Marco Ghiandoni and his supervisor Prof Valerie J. Gillet (Information School, University of Sheffield), for the work on the identification of FEN1 inhibitors.

Thank you to Dr Spencer Collis, Dr Ruth H Thompson and the wider Collis, Thompson and Bryant groups for their valuable feedback during our lab meeting presentations.

My greatest thanks go to my family, my true driving force, who supported me throughout my studies, and laid the foundations for an unstoppable, lifelong and curiosity-driven learning journey. Thank you to mum and dad, Stella Stylianou and Andreas Fotiou†, to my grandparents, Stavroula and Angelis Stylianou and my sister, Loukia Louka.

Lastly, thanks to my 12 years on boyfriend, Valentinos Andreou, who ever since we graduated high school has been relocating to support me with every single step I take. Thank you for so patiently tolerating watching me puzzling over protein crystallization, wrangling over my non-diffracting crystals and suffering my PhD's drama! "It's a yes, finally".

Dedicated to my father, Andreas Fotiou, who passed away on the 12th of June 2020 due to end-stage kidney failure, while the present PhD thesis was still in progress. Submission and, eventually, completion of this PhD is first of all your and mum's achievement. Thank you for everything you taught me, the opportunities you have given me and the dreams, worth fighting for.

Abbreviations

HsFEN1	(Human) flap endonuclease-1
Å	angstrom
AP	apurinic/apyrimidinic site
Apc	adenomatous polyposis coli gene
APE1	AP endonuclease-1
BER	base excision repair
BIR	break-induced replication
BLAST	Basic local alignment search tool
Cdc4	cell division control 4 gene
CpG	cytosine-phosphate-guanidine
D/Asp	aspartic acid
DNA	Deoxyribonucleic acid
DSB	double-strand DNA breaks
E/Glu	glutamic acid
EXO1	exonuclease-1
Fen1	Flap endonuclease-1 gene
FRET	fluorescence resonance energy transfer
GEN	gap endonuclease activity
GEN-1	gap endonuclease-1
H/His	histidine
H2TH	helix-two turn-helix
HR	homologous recombination
HTC-116	human colorectal carcinoma cell line
I/Ile	isoleucine
IR/UV	ionizing/ultraviolet radiation
K/Lys	lysine
k_{cat}	turnover number
L/Leu	leucine
Lig I	DNA ligase I
Me_A^{2+}/ Me_B^{2+}/ Me_C^{2+}	first/ second/ third divalent metal ion
MMS	methyl mathanesulfonate
Mre11a	meiotic recombination 11-like protein A gene

mRNA	messenger RNA
NCBI	National Center for Biotechnology Information
NT	nucleotide
ORC	origin replication complex
PCNA	proliferating cell nuclear antigen
PCR	polymerase chain reaction
PDB	Protein Data Bank
PilT	protein incorporated later into tight junctions
PIN	PilT N-terminal domain
Pol α, δ, ε	DNA polymerase α, δ, ε
R/Arg	arginine
Rad27	radiation 27 gene
RCF	replication factor C
RCSB	Research Collaboratory of Structural Bioinformatics
RNA	ribonucleic acid
RPA	replication protein A
S/Ser	serine
SAM-	sterile alpha motif-
siRNA	small inhibitory RNA
SNPs	single nucleotide polymorphisms
SP/LP-BER	short-patch/long-patch-BER
ss/ds	single-stranded/double-stranded
T/Thr	threonine
Taq	*Thermus aquaticus*
TMZ	temozolomide
WT	wild-type
XPG	xeroderma pigmentosum complementation group G
Y/Tyr	tyrosine
MSI	microsatellite instability
DTT	dithiothreitol
PEI	Poly(ethyleneimine)
SP	sulphopropyl
Q	quaternary ammonium

CV	column volumes
PES	polyethersulfone
TCGA	The Cancer Genome Atlas
FRET	Förster Resonance Energy Transfer
DIALS	Diffraction Integration for Advanced Light Resources
TEMED	N,N,N',N'-tetramethylethylenediamine
FPLC	Fast Protein Liquid Chromatography
EDTA	ethylenediaminetetraacetic acid
W/Trp	tryptophan
SDS	sodium deoxycholate
PAGE	polyacrylamide gel electrophoresis
E. coli	*Escherichia coli*
Cy3	cyanine 3
FAM	carboxyfluorescein
WHO	World Health Organisation
IACG	UN Interagency Coordination Group

List of Figures

List of Tables

Table of Contents

Chapter 1 – Introduction

1.1 Flap endonuclease (FEN) I mediates genome integrity: Roles in DNA replication, repair and telomere stability

1.1.1 Role of FEN in DNA replication

In eukaryotic cells, DNA synthesis occurs at the replication fork in a semi-discontinuous manner. Initially, at the replication origin, the DNA duplex is opened up by the action of six helicase subunits bound together, which form the origin replication complex (ORC) (Figure 1.1). Once the replication fork is formed, replication protein A (RPA) binds to single-stranded (ss) DNA stabilizing the template strands, which are replicated simultaneously by discrete initiation events. These include addition of a single RNA primer on the leading strand and multiple ones on the lagging strand, synthesised by RNA polymerase in complex with DNA polymerase α (Pol α), known as primase. Pol α is then displaced by Pol ε on the leading strand and Pol δ on the lagging strand under the action of replication factor C (RCF)/proliferating cell nuclear antigen (PCNA) (Figure 1.1). Pol α and δ then catalyse DNA elongation. Because of the antiparallel nature of DNA, the leading strand is replicated continuously in 5′ to 3′ direction, whereas lagging strand replication occurs in segments, in a process known as Okazaki fragment synthesis (Burgers, 2009; Nazarkina *et al*., 2008) (Figure 1.1). While Pol δ approaches and displaces the RNA primer at the 5′ end of the previous Okazaki segment, 5′ branches referred as flaps are formed (Burgers, 2009; Maga *et al*., 2001) (Figure 1.1). Precise removal of these 5′-flaps is vital to ensure efficient ligation of Okazaki fragments by DNA ligase I (Lig I) and thus, to maintain genome fidelity. Approximately 50 million primers on the lagging strand are estimated to require removal per human cell cycle (Burgers, 2009).

Two distinct mechanisms have been proposed to be involved in this process, also referred as Okazaki fragment maturation, dependent on the action of flap structure-specific endonuclease 1, FEN, alone or in synergy with Dna2 (Burgers, 2009; Rossi and Bambara, 2006). As the 5′-end of the elongated lagging strand reaches the downstream RNA primer, ≤2 ribonucleotides are displaced by Pol δ, resulting in short 5′-flaps (Figure 1). These, are removed by the sole action of FEN. For longer RNA primers, the remaining ribonucleotides are removed by consecutive Pol δ/FEN cycles, until the primer is fully degraded (Garg *et al*., 2004). On the other hand, extensive nucleotide displacement by Pol δ results in longer 5′-flaps, which cannot be cleaved by FEN alone. Biochemical analyses on ss flaps, ≥30-nucleotides long, revealed an alternative mechanism that takes advantage of the complementary nuclease activities of FEN/Dna2 (Kao *et al*., 2004). Longer 5′-flaps bound by RPA seem to promote endonucleolytic cleavage by Dna2, which results to shorter flaps capable of being cleaved by FEN (Kao *et al*., 2004) (Figure 1.1).

Removal of RNA primers, solely by FEN, is likely the predominant pathway for higher eukaryotic organisms. The most important evidence for this, is firstly that *Fen1* deficient murine embryos die *in utero* (Kucherlapati *et al*., 2002), unlike the homologous *Rad27*$^{-/-}$

Saccharomyces cerevisiae mutants, and secondly, that human Dna2 is localised to mitochondria and does not participate in mammalian DNA replication (Zheng *et al*., 2008). In contrast in lower organisms, although cell growth rates are decelerated and sensitivity to duplication mutations is increased in *Rad27*$^{-/-}$ cells (Greene et al., 1999), synthetic lethality has been shown to be caused in *Rad27*$^{-/-}$ yeast cells mated with *Pol δ* mutations attenuating its exonuclease activity (Jin *et al*., 2001). Subsequent studies, showed that these cells could be rescued when *Dna2* is overexpressed, suggesting a critical role of FEN/Dna2 pathway in lower organisms (Jin *et al*., 2003).

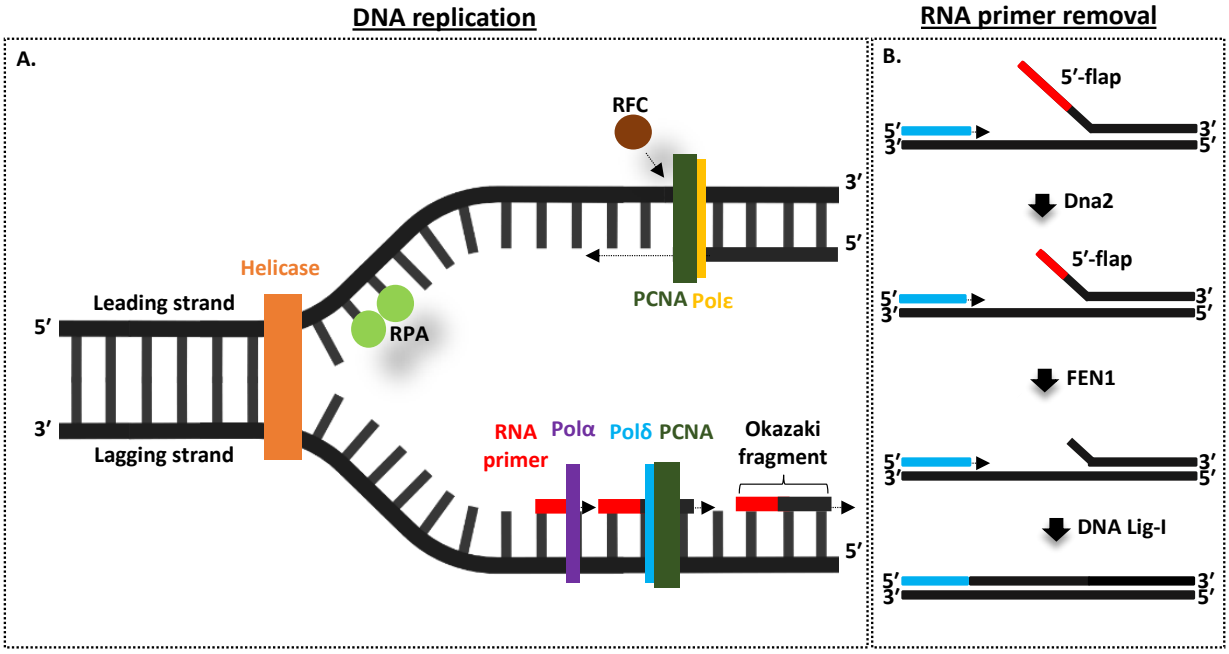


Figure 1.1 Role of FEN1 in DNA replication. (A) Schematic representation of DNA replication in eukaryotes. **(B)** Removal of RNA primers by FEN/Dna2 pathway. While Pol δ (light blue) catalyses strand displacement, 5′-flaps are generated (red). Long 5′-flaps (≥30-nt) are processed by FEN/Dna2. Initially, Dna2 catalyses the endonucleolytic cleavage of these flaps, resulting in shorter 5′-flaps which can then be cleaved by FEN.

1.1.2 Role of FEN in DNA repair

Several DNA repair mechanisms have evolved in different organisms to deal with DNA lesions, caused by the cumulative effect of endogenous and/or exogenous factors. By definition, endogenous factors refer to those generated within the cells such as products of metabolism, whereas exogenous refers to environmental risk factors including exposure to ionizing/ultraviolet radiation (IR/UV) and toxic chemicals (Nazarkina *et al*., 2008). Defective repair of these lesions often results in functional consequences, which threaten genomic integrity and lead to diverse complex diseases, including cancer. FEN is a crosstalk protein involved in a number of DNA repair pathways, which interacts with other crucial proteins of the repair machinery to ensure sufficient DNA repair (Nazarkina *et al*., 2008).

Single base damage could be induced by either endogenous or exogenous factors and represent the commonest type of DNA lesions. These are mainly repaired through base excision repair

(BER), which in turn is subdivided in short-patch (SP) and long-patch (LP) (Nazarkina *et al*., 2008). In both pathways, DNA glycosylase recognises the damaged site and cleaves the relevant N-glycosidic bond to remove the damaged base. This cleavage results in a gap on the DNA backbone, known as an apurinic/apyrimidinic (AP) site depending on the missing base. APE1, which is an AP endonuclease, cleaves the 5′ phosphodiester bond of the AP site to create 3′-hydroxyl and 5′-phosphate ends (Whitaker and Freudenthal, 2018). At this point the two pathways differentiate. In SP-BER, the resynthesized nucleotide containing the correct base, is added by Pol β and DNA ends are covalently joined by DNA ligase III. In LP-BER. Pol δ, along with RCF/PCNA, or Pol β catalyse strand displacement 5′ to the AP site and the formed 5′-flap is cleaved by FEN (Kim *et al*., 1998). The resulting ends are finally joined together by DNA ligase III.

FEN, is also required to unblock stalled replication forks. Normally during DNA replication, six helicase subunits, known as the replicative maschinery, act in synergy to untwist the DNA double helix. This process leads to an opened, Y-shaped DNA structure, called a replication fork, that allows DNA replication. DNA lesions caused by impaired genomic replication/repair mechanisms, can alter the progressive movement of stalled replication forks by setting up barriers that challenge the DNA replication machinery (Nazarkina *et al*., 2008). Inability of the replication fork to overcome these barriers can lead to fork stalling and even collapse of the fork resulting in one ended double-strand DNA breaks (DSB), deletions or expansions that threaten fork and telomere stability and cell viability.

Through its gap endonuclease activity (GEN), FEN1 catalyzes the cleavage of the template strand of a gapped replication fork. These are mimicking stalled replication forks and therefore FEN1 seems to play a critical role in DSB repair and unblocking stalled replication forks. Using DNA bubble substrates representing stalled replication fork structures, Zheng *et al*. (2005) identified that FEN1 is capable of processing DNA bubble structures at ss-ds junction. Subsequent co-immunoprecipitation assays in HeLa cells, showed that this cleavage, is initiated by FEN1 interactions with Werner syndrome helicase (WRN) (Zheng *et al*., 2005). In response to chemotherapeutic drugs, and particularly to camptothecin, FEN1/WRN co-localise at stalled replication forks producing 5′-ssDNA substrates, which are processed by FEN1 once RPA associates (Zheng *et al*., 2005). Efficient cleavage of ssDNA regions, formed in stalled replication forks, is critical to create DSB/recombination structures, which could be repaired through break-induced replication (BIR) pathway of recombinational repair.

1.1.3 Role of FEN in telomere stability

In addition to the roles of FEN in DNA replication and repair, both lagging and leading strand telomere stability were also found to be highly dependent on the canonical function of FEN. Parenteau and Wellinger provided the first direct evidence of FEN-mediated telomere stability in *Saccharomyces cerevisiae* through its involvement in telomeric lagging strand replication (Parenteau and Wellinger, 1999; 2002). Deletion of *RAD27* gene, encoding the yeast FEN homologue, resulted in high telomere heterogeneity and accumulation of single-stranded G-

rich overhang. These, in addition to the fact that no accumulation of single-stranded C-rich overhangs was detected in leading strand suggested initially a unique role of FEN1 in telomeric lagging strand replication.

Comparable studies on higher eukaryotes have also confirmed an indirect role of FEN1 in telomere maintenance (Sampathi *et al*., 2009). Using co-immunoprecipitation and a non-radioactive Telomere Repeat Amplification Protocol (TRAP) analysis, Sampathi *et al*., (2009) provided the first evidence for the existence of an *in vivo* FEN1-telomerase complex in telomerase-positive cancer cells and normal human somatic cells, suggesting a role in regulation of mammalian telomerase activity. Interactions between FEN1 and telomerase were found to be mediated by telomere DNA and the human Telomerase Reverse Transcriptase (TERT) subunit. In an effort to functionally characterize the importance of that indirect interaction between human FEN1 and TERT, FEN1 deficiency in mouse embryonic fibroblasts was shown to retard telomerase elongation and cell growth, without however altering the telomeric G-overhangs. FEN1 specifically targets single-stranded G4 DNA 5′-flaps, independently of whether these are telomeric or not, consistent to its role in lagging strand replication (Vallur and Maizels, 2010). In addition, telomeric G-rich 5′-flaps were also found to be excised by both FEN1 and its 5′-nuclease superfamily homologue, exonuclease I (EXO1).

Complete replication of telomeric leading strand also requires the synergistic function of FEN1. On the most recent study investigating the functional role of FEN1 in telomere maintenance and fidelity, the canonical function of FEN1 was found to extend from telomeric lagging to leading strand replication (Teasley *et al*., 2015). Replication stress induced by co-directional collisions likely to occur between RNA polymerase and replisome results in intermediate RNA:DNA flap structures, resembling those observed in Okazaki fragment synthesis. Accumulation of such structures, and therefore double strand breaks, promotes telomere fragility. FEN1 was shown to act on the leading strand solving these double strand breaks, re-initiating replication fork and preventing telomere fragility, specifically derived from leading strand replication flaws.

1.1.4 The fundamental nature of FEN to life

Since its initial discovery as a 5′ to 3′ exonuclease at the N-terminal domain of *Escherichia coli* DNA polymerase I (Pol I) and its critical roles in DNA replication and repair, 5′ to 3′ exonucleases were thereafter identified as members of a wider family of evolutionary conserved metallonucleases, named flap endonuclease (FEN) superfamily (Harrington and Lieber, 1994). Multiple cross-species studies, spanning in all three kingdoms of life (Eukaryotes, Bacteria and Archaea), including some viruses, have demonstrated the presence of such a biologically important 5′ to 3′ exonuclease.

Bacterial DNA Pol I is typically composed from the C-terminal “Klenow” fragment, which exhibits the polymerase and 3′ to 5′ activity, and the N-terminal 5′ to 3′ exonuclease. Aiming to address the importance of each DNA Pol α domain, encoded by *polA* gene, in bacterial cell

viability, Diaz *et al*., (1992) generated three distinct *Streptococcus pneumoniae* mutants, targeting different sites of Pol I. Chromosomal transformation of these mutants showed that mutants encoding a truncated Pol I, defective in polymerase activity but with retained FEN1 activity, were transformed with efficiency analogous to that of WT *polA* gene. On the other hand, transformation of mutants, defective in FEN activity, was dramatically decreased (100,000-fold), independently of whether polymerase activity was retained or not. These data suggested that cell viability of *Streptococcus pneumoniae* depends on the 5′ to 3′ exonuclease activity of FEN and not the polymerase activity *per se* (Diaz *et al*., 1992)

Disruption of the 5′ to 3′ exonuclease activity of a T4 bacteriophage FEN homologue, historically known as T4 RNase H, also resulted in impaired DNA replication and decreased phage growth (Hobbs and Nossal, 1996). However, when combined with deletion of DNA Pol I N-terminus, phage progeny was found to be totally arrested supporting the indispensable requirement of a 5′ to 3′ exonuclease activity for phage DNA replication. Disruption of the 5′ to 3′ exonuclease activity, related to the N-terminus of DNA Pol α, typically results to lethal phenotypes similar to those seen in *rad27* null mutants, a gene also known to encode a FEN homologue (Reagan *et al*., 1995; Sommers *et al.*, 1995). Overexpression of the 5′ to 3′ exonuclease domain of *Escherichia coli* DNA Pol α in *Saccharomyces cerevisiae* was shown to suppress these phenotypes, demonstrating clearly that 5′-nucleases can compensate the 5′ to 3′ exonuclease activity of FEN1 (Sun *et al.*, 2002).

In contrast to *Streptococcus pneumoniae*, analogous studies in *Escherichia coli* and *Bacillus subtilis* indicated that Pol I is dispensable for bacterial cell growth, although it can lead to growth phenotypes of temperature sensitivity and high mutation rates (Fukushima *et al*., 2007; Nagata *et al*., 2002). Further exploration of these conflicting findings with regards to the essential nature of FEN orthologs led to the identification of a synthetic lethal partner of *polA*. *Ypc*P gene, encoding a second protein with a FEN-like domain in *Bacillus subtilis*, was shown to compensate the 5′ to 3′ nuclease activity of FEN (Fukushima *et al.*, 2007). Similarly *xni* and *exo1* genes were shown to compensate the 5′ to 3′ nuclease activity of FEN in *E. coli* and *Saccharomyces cerevisiae* (Fukushima *et al.*, 2007; Tishkoff *et al*., 1997). Typically, a 5′ to 3′ nuclease activity is required for cell viability and growth, independently of whether that is exhibited from the protein encoded by polA, *ypc*P or other genes.

Genetic knockout studies in mammalian organisms focusing on the essential nature of FEN and their likely *in vivo* roles in DNA replication and repair, reported that double-knockout mice, homozygous for deletion of both *Fen1* alleles ($Fen1^{-/-}$) could not be obtained (Kucherlapati *et al*., 2002). Consistently, intercrosses of $Fen1^{+/-}$ heterozygotes, conducted on a separate study also resulted in early *in utero* lethality of $Fen1^{-/-}$ mice embryos, due to complete inhibition of DNA replication (Larsen *et al*., 2003). $Fen1^{+/-}$ heterozygotes on the other hand, were found to be initially phenotypically normal, but with a mild predisposition to cancer (Kucherlapati *et al*., 2002; Larsen *et al*., 2003). No distinguishable differences with regards to fertility, morphology, histology and life span were identified compared to their WT littermates. When in conjunction though with mutations in adenomatous polyposis coli gene (Apc^{1638N}), a

tumour suppressor gene, double heterozygotes were shown to have reduced survival median and substantially increased rates of tumour development and progression (Kucherlapati *et al*., 2002). Although the indispensable nature of FEN for viability and survival of mammalian organisms suggested by these knockout experiments was not surprising at the time, haploinsufficiency and the link to cancer development and progression pointed out a new window for cancer therapeutics.

1.2 Structural specificity of FEN substrate recognition

Unlike most endonucleases, FEN1 homologs and the wider members of 5′-nuclease superfamily target and process an array of branched DNA structures in a sequence-independent manner (Garforth and Sayers, 1997; Harrington and Lieber, 1995). Such structures arise *in vivo* as intermediates at multiple stages of DNA metabolism, and processing of those with a 5′ to 3′ polarity can occur either endo- or exonucleolytically, depending on the 5′-nuclease catalysing the reaction and the type of the targeted DNA substrate. Despite the unified preference in ssDNA-dsDNA junctions and the conserved two-metal hydrolysis reaction mechanism, the individual substrate specificities and cleavage efficiencies are extraordinarily diverse (Table 1.1).

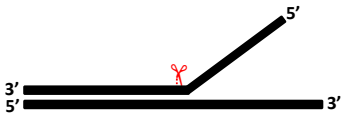
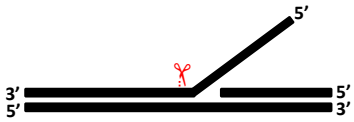
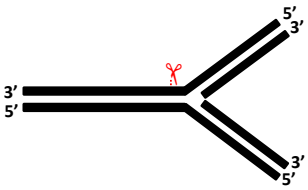
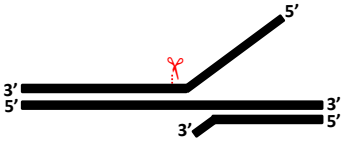
Double-stranded DNA substrates with a single-stranded 5′-flap were predominantly described as the natural substrates of FEN1, the prototypical member of 5′-nuclease superfamily. The binding efficiency of substrates having two dsDNA regions on either side of a single-stranded nick was shown to be significantly increased (100-fold), compared to pseudo-Y structures with single dsDNA regions and substrates with a 5′-overhang, lacking completely the 5′-flap strand (Harrington and Lieber, 1995). Binding and cleavage efficiency were also found to be affected by the respective gap length between the ssDNA-dsDNA junction of the 5′-flap strand and the 3′-end of the adjacent strand. Although hydrolysis of 5′-flap substrates with double duplex regions was shown to occur independently of the gap length at efficiencies higher than those observed for pseudo-Y substrates, gap length does affect the efficiency of processing. Gap lengths of only 1-nt are recognised as optimum, which when combined with a 1-nt long 3′-flap in addition to 5′-flap (double-flap substrate) maximise FEN1 efficiency.

The requirement of a 1-nt long 3′-flap for optimum efficiency has been confirmed by multiple studies on eukaryotic and archaeal FEN1, suggesting an important role in FEN1 binding and stabilisation of FEN1:DNA complexes. Double-flap substrates bearing a 1-nt long 3′-flap were cleaved by FEN1 with higher efficiency compared to longer 3′-flaps (Table 1.1; Friedrich-Heineken and Hubscher, 2004; Harrington and Lieber, 1995; Kao *et al*., 2002). This is attributed to the resulting nicked dsDNA product being immediately ligatable. Cleavage of the phosphodiester bond exclusively 1-nt into the ssDNA-dsDNA junction of the 5′-flap end leads to a 5′-monoester, which in the presence of a 1-nt long 3′-flap and its free hydroxyl group can be immediately ligated, without the interference of additional DNA repair pathways. Double-flap DNA substrates with a nick that bears a ssDNA/RNA 5′-flap and an unannealed 3′-terminal

nucleotide (3′-flap) are now recognised and the optimal substrates for FEN1 (Table 1.1; Harrington and Lieber, 1995; Kao *et al*., 2002).

Circular ssDNA, such as bubble and heterologous loop substrates, are also recognised by FEN1 enzymes for cleavage, through its most recently described GEN activity (Table 1.1; Zheng *et al*., 2005). Initially, endonucleolytic cleavage of nicked DNA was described as a feature of bacteriophage T5FEN, but later studies on *Homo sapiens* (Hs)FEN1 revealed an efficient endonucleolytic cleavage of DNA bubble structures, derived *in vivo* at stalled replication forks (Sayers and Eckstein, 1991; Zheng *et al*., 2005).

Table 1.1 Overview of recognized FEN1 DNA substrates

DNA/RNA structure	Assay[1]	Reference(s)
 Recessed 5′-end	EMSA	(Harrington and Lieber, 1994; Shen *et al*., 2005)
 Pseudo-Y	EMSA	(Harrington and Lieber, 1994; Shen *et al*., 2005)
 5′-Flap	EMSA & Native PAGE	(Bornarth *et al*., 1999; Harrington and Lieber, 1994; Zheng *et al*., 2005; Shen *et al*., 2005)
 Y-structure	Native PAGE	(Zheng *et al*., 2005; Shen *et al*., 2005)
 Bubble structure	Native PAGE	(Zheng *et al*., 2005; Shen *et al*., 2005)
 Double flap	EMSA	(Gloor *et al*., 2010; Shen *et al*., 2005)

[1]Experimental assay used to determine cleavage of the corresponding DNA/RNA substrates by HsFEN1. EMSA, refers to Electrophoretic Mobility Shift Assay. Scissors are indicative of FEN1 cleavage site.

1.3 Structural and biochemical conservation among FEN superfamily

1.3.1 FEN structural core and spatial conservation of common functional motifs

Aiming to resolve the paradoxes of substrate recognition, binding and catalysis within the members of 5′-nuclease superfamily, structural studies on FEN enzymes from diverse organisms have today elucidated a common architecture. Comparative analyses of these models have provided unprecedented insights into the level of structural conservation among FEN proteins, while key functional motifs, likely to underline the differences in substrate specificity, have been identified.

Thermus aquaticus FEN represents the first high-resolution structure of a FEN protein published, which was solved as part of the entire full-length PolI domain (Kim *et al*., 1995). This was then followed by a series of FEN structures spanning from eukaryotes to archaea, bacteria and viruses, suggesting a conserved metal ion-dependent reaction mechanism in addition to the remarkably similar domain organization (Table 1.2; Figure 1.2).

Despite their sequence diversity, the core domains of all structurally characterised FENfamily member enzymes fold into a SAM-/PIN-like arrangement, with α-helices surrounding a packed parallel and antiparallel β-sheet, composed particularly of seven β-strands in HsFEN1. These surrounding α-helices (α2, α4 and α5) span a flexible helical gateway, just above a carboxylate-rich active site. Asp34, Asp86, Glu158, Glu160, Glu179, Asp181 and Asp233 are conserved active site residues among eukaryotic and archaeal FEN1, critical for metal ion coordination and catalysis. These seven active site carboxylates are often classified into two categories, depending on the position of the coordinated metal: metal ion binding sites 1 (Asp86) and 2 (Asp179 and Asp181). Glu160 is the key residue bridging the two metals, whereas, unlike Asp86, Glu179 and Asp181, Glu158 and Asp233 do not interact directly with the active site metals, but only through water bridges (Sakurai *et al*., 2005).

Table 1.2 Published apo structures of FEN homologs and the wider human members of the 5′-nuclease superfamily to date

PDB ID	Protein	Organism		Active site metals	Metal ions coordination distance (Å)	Resolution (Å)	Reference
1TAQ	Taq polymerase	*Thermus aquaticus*	Bacteria	Zn^{2+}	-	-	(Kim *et al*., 1995)
1TFR	T4FEN/ T4 RNase H	*Escherichia phage T4*	Virus	Mg^{2+}, Mg^{2+}	6.3	2	(Mueser *et al*., 1996)
1UT5	T5FEN	*Escherichia phage* T5	Virus	Mn^{2+}, Mn^{2+}	8	2.75	(Feng et al., 2004a)
1B43	FEN	*Pyrococcus furiosus*	Archaea	-	-	-	(Hosfield *et al*., 1998)
1A77	FEN	*Methanococcus jannaschii*	Archaea	Mg^{2+}, Mg^{2+}	5	2	(Hwang *et al*., 1998)
1MC8	FEN	*Pyrococcus horikoshii*	Archaea	-	-	3.1	(Matsui *et al*., 2002)
1RXV	FEN	*Archaeglobus fulgidus*	Archaea	-	-	2.5	(Chapados *et al*., 2004)
2IZO	FEN	*Sulfolobus solfataricus*	Archaea	Zn^{2+}, Mg^{2+}, Mg^{2+}	9.8	2.9	(Dore *et al*., 2006)
1UL1	FEN1	*Homo sapiens*	Eukaryotes	Mg^{2+}, Mg^{2+}	3.6	2.9	(Sakurai *et al*., 2005)
3QEB	EXO1	*Homo sapiens*	Eukaryotes	Mn^{2+}, Mn^{2+}	4.1	3	*(Orans et al*., 2011)
3ORY	FEN	*Desulfurococcus amylolyticus*	Archaea	-	-	2	(Mase *et al*., 2011)
3ZD9	ExoIX	*Escherichia coli*	Bacteria	-	-	2	(Anstey-Gilbert *et al*., 2013)
4WA8	FEN	Methanopyrus kandleri	Archaea	-	-	2.2	(Shah *et al*., 2015)
5CNQ	GEN1	*Homo sapiens*	Eukaryotes	Mn^{2+}, Mn^{2+}	5.4	3	(Lee *et al*., 2015)
6VBH	XPG	*Homo sapiens*	Eukaryotes	-	-	2	(Tsutakawa *et al*., 2020)

All published structures shown here have been determined using X-ray crystallography, based on the available information deposited in PDB database. Structures are shown in a chronological order, starting with the oldest. Enzymes complexed with DNA for which apo structures are not available have also been included to the table (Date accessed: July, 2021).

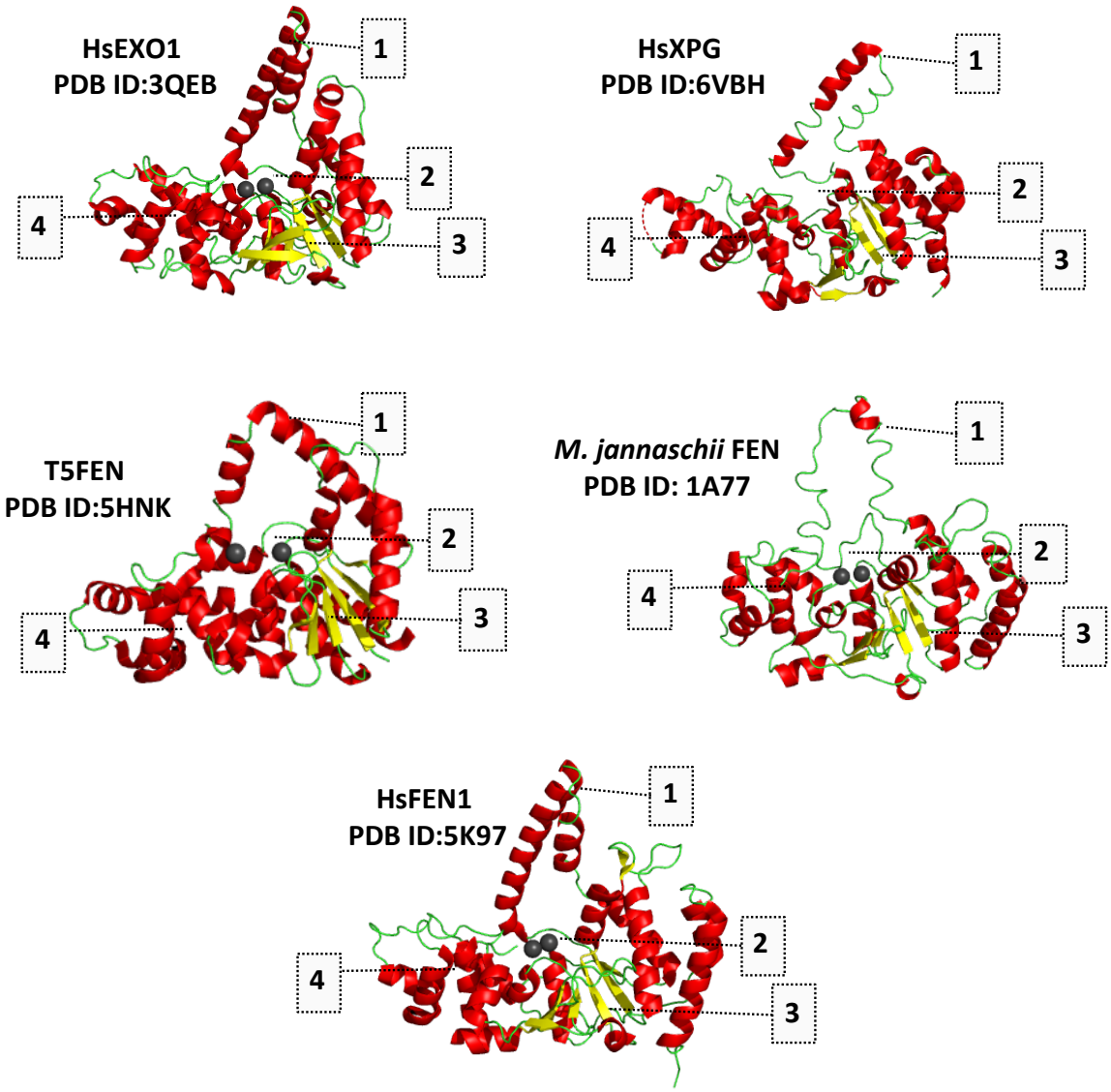


Figure 1.2 Structural similarities across HsFEN1 homologs. Cartoon representation of HsFEN1 and selected HsFEN1 homologue structures with α-helices, β-sheets and loops shown in red, green and yellow, respectively (orthologs: T5FEN and *Methanococcus jannaschii* FEN; paralogs: HsEXO1 and HsXPG). Numbers 1 to 4 correspond to the conserved among HsFEN1 homologs structural motifs, involving the important for catalysis helical gateway (1), the rich in carboxylates active site (2), the packed parallel and antiparallel β-sheets composing the central protein core (3) and the characteristic helix-turn-helix motif (4). Grey spheres are indicative of the active site metals, as these appear in the corresponding protein structures. HsFEN1 and homologue structures were downloaded from the online PDB depository and rendered using PyMOL™ version 2.3.5 Schrödinger, LLC (HsFEN1 – Tsutakawa *et al*., (2011); HsEXO1 – Orans *et al*., (2011); HsXPG – (Tsutakawa *et al*., (2020); T5FEN – AlMalki *et al*., (2016); *M. jannaschii* FEN – Hwang *et al*., (1998)).

HsFEN1 binding towards the 3′-end of the dsDNA template is mediated by a K^+ bridge, which is coordinated with the helix-two turn-helix (H2TH, Leu218-His253) motif, while substrate recognition occurs due to the ability of 5′-flaps to bend 100° at a phosphate diester linkage of dsDNA junction (Figure 1.3). This DNA bending is mediated by several interactions, including

mainly residues Ser197, Thr195 and Glu198 of β-pin (β6 and β7 loops), the hydrophobic wedge, consisting from α2 helix, β2 and β3 loops, and the 3′-binding pocket (Figure 1.3).

An unannealed 3′-flap is typically considered necessary to obtain sealable nicks, which will enable DNA ligation. The hydroxyl group of the 3′-flap end forms hydrogen bonds with Lys314 and Thr61 in HsFEN1, increasing not only the interaction surface between the two, but also the subsequent ligation efficiency. Thus, explaining the structural preference of HsFEN1, and the wider higher organism FEN1 enzymes, for double-flap substrates. Repulsion forces developed between a negatively charged block (Glu56-Glu59), referred as "acid block", and DNA phosphate backbone ensures that only the unannealed 3′-flap passes by the 3′-binding pocket. Therefore, allowing the so called helical gateway to act as an entry point for a contingent 5′-flap threading (Figure 1.3).

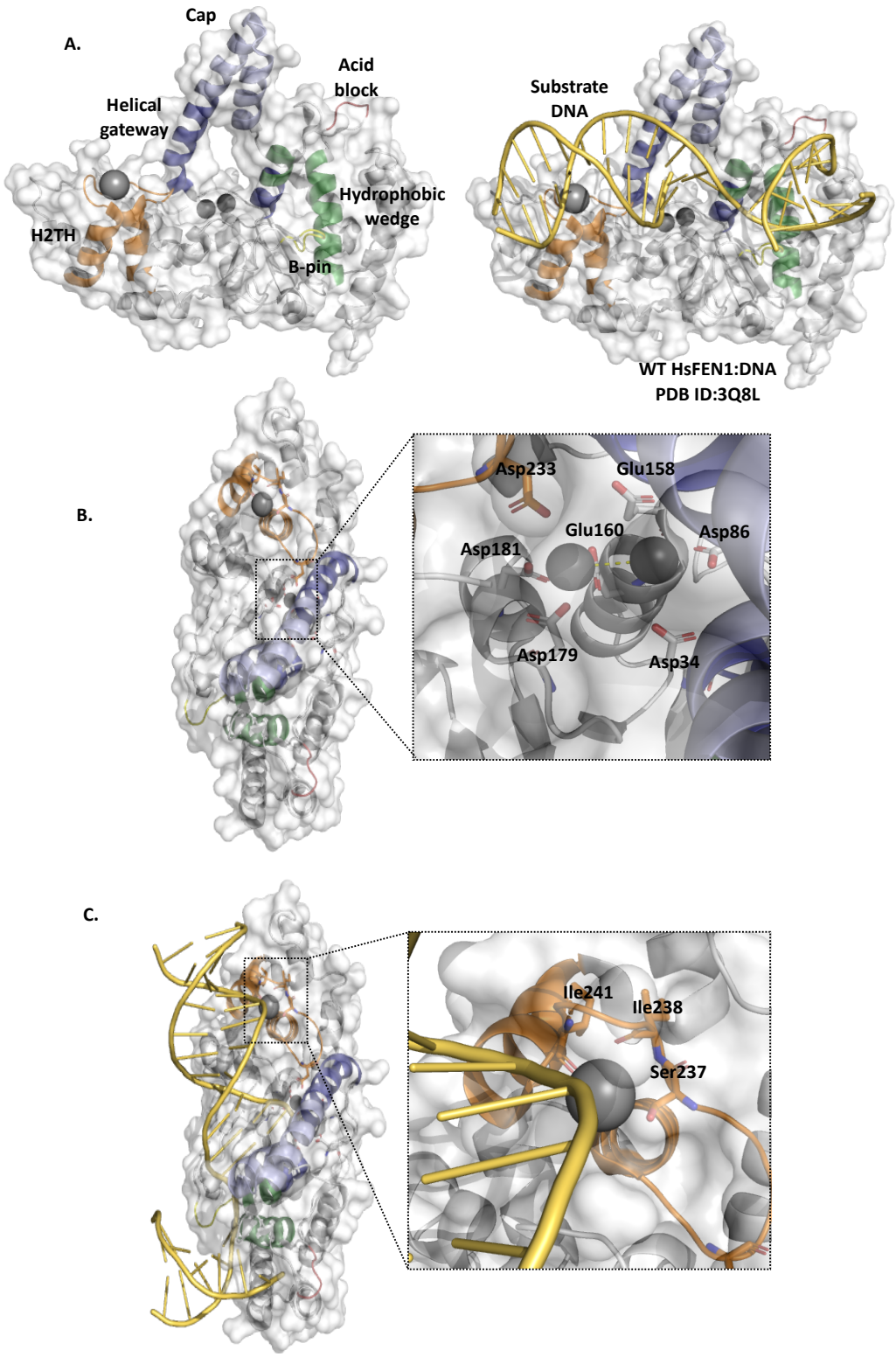


Figure 1.3 Structure of HsFEN1. (A) Cartoon and surface representation of HsFEN1 structure with/without DNA bound, with the key for DNA recognition and binding elements colored. Key structural elements involve the interacting with upstream dsDNA H2TH (orange), the only ssDNA permitting helical gateway (blue), the flexible cap (light blue), the hydrophobic wedge (green) and its forming acid block (red) inhibiting 1-nt long 3'-flap, and the interacting with downstream dsDNA β-pin (yellow). Potassium and the two active site metal ions mediating interactions between HsFEN1 and DNA are shown in grey spheres. (B)

Close-up view of the rich in carboxylates HsFEN1 active site with the key negatively charged residues shown in sticks. (C) Close-up view of H2TH motif interacting with dsDNA through K^+. Figures were generated in PyMOL™ (Schrödinger, LLC), using the deposited in PDB online database 3Q8L structure (HsFEN1 – Tsutakawa *et al*., (2011)).

1.3.2 FEN1 helical gateway

With the dsDNA of the 5′-flap substrates bound on either side of the dual metal active site, protruding helices generate a structurally conserved helical gateway, also referred to as helical clamp. Based on the available structures of FEN1 homologs and the members of 5′-nuclease superfamily, this typically consists of 3-4 helices. Helices extended by both sides of the dual metal active site form the base of the helical gateway, which ends in a structurally flexible cap.

Firstly identified by Ceska *et al*., (1996) in a 2.5 Å structure of a bacteriophage FEN1 homolog, T5 D15 exonucelase, (T5FEN), the helical gateway was proposed to have a critical role in ssDNA specificity and recognition of ssDNA-dsDNA junctions. With a diameter capable of accommodating only ssDNA and not dsDNA, catalysis by FEN1 homologs and members of 5′-nuclease superfamily was initially suggested to be mediated by ssDNA threading through the helical gateway. Indeed, co-crystal structures of FEN1 enzymes with DNA substrates showed helical gateways with diameters <20 Å, smaller than the required diameter for dsDNA passage. The mobile cap was also found to exist in two different structural confirmations: (a) a completely disordered arrangement in the DNA-free structures and (b) an ordered arrangement in DNA-bound structures with annealed 5′-flaps pointing towards the helical gateway (Figure 1.4; AlMalki *et al*., 2016; Tsutakawa *et al*., 2011). This indicated a potential entry point for the ssDNA, while the conserved transitioning from a “disordered-to-ordered” conformation seemed to highly explain the specificity of 5′-nuclease superfamily members to ssDNA-dsDNA junctions. Despite their common preference to ssDNA-dsDNA junctions, 5′-nuclease superfamily members were shown to target remarkably different DNA structures as their optimal substrates. This internal to the 5′-nuclease superfamily specificity is mostly attributed to the highly different amino acid sequence, composing the cap (helices α2, α4 and α5), and its controlled regional rearrangement determining the orientation of unique, but key, residues in DNA catalysis.

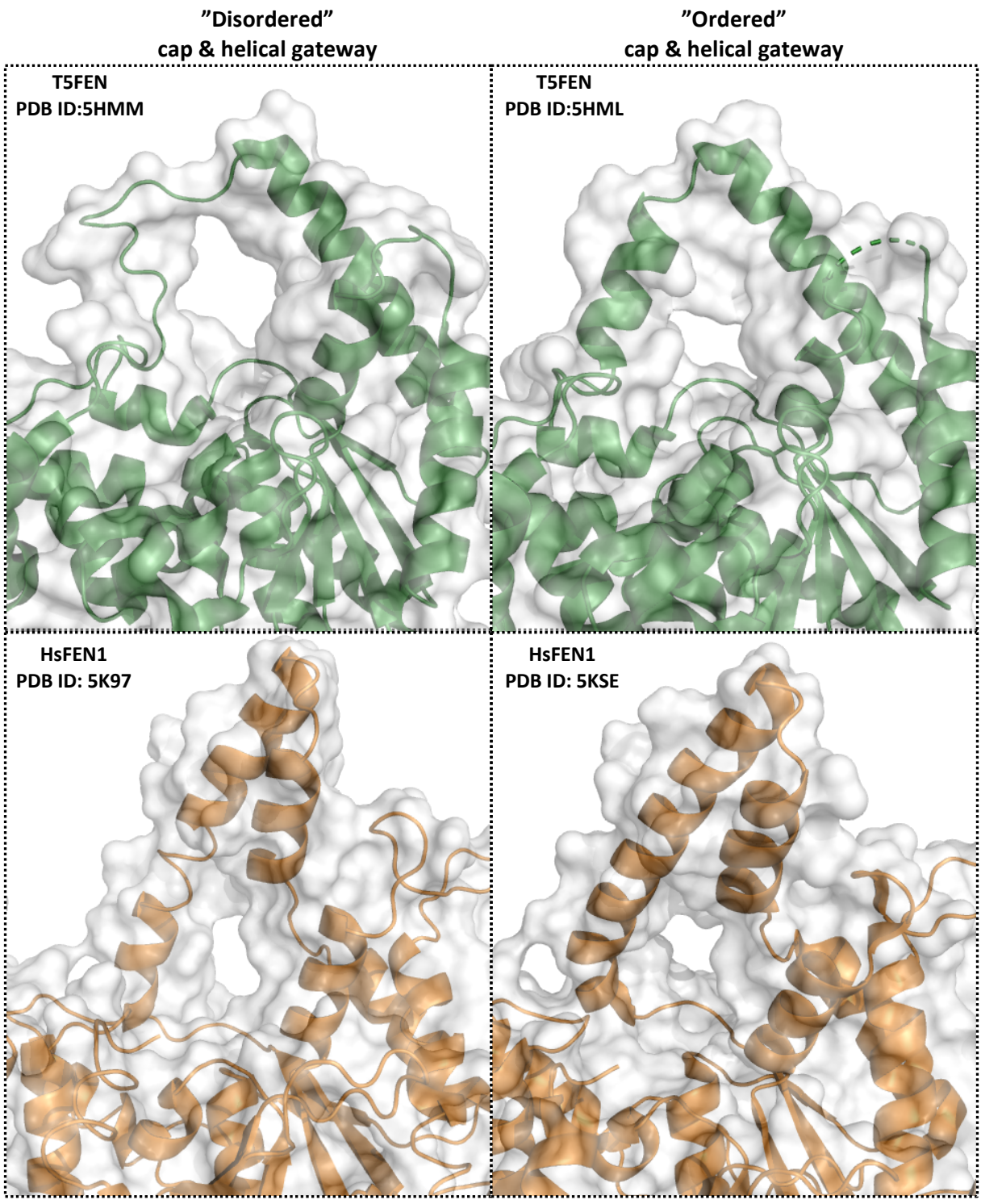


Figure 1.4 “Disordered-to-ordered” helical arch conformations in HsFEN1 and T5FEN enzymes. Close-up views of T5FEN (green cartoon) and HsFEN1 (orange cartoon) helical channels, representative of the occurring “disordered-to-ordered” channel transition during DNA processing. Figures were generated in PyMOL™ (Schrödinger, LLC), using the deposited in PDB online database structures (HsFEN1 – Tsutakawa *et al*., (2017); T5FEN – AlMalki *et al*., (2016)).

1.3.3 Conserved *helix-turn-helix* motifs

The helix-turn-helix (HTH) motifs, consisting of two α-helices with the same length linked through a loop, consisting of 2-3 turns, are also among the structurally conserved features of FEN1 homologs and the members of 5′-nuclease superfamily (Figure 1.5). These are highly similar to the well-characterized helix-hairpin-helix (HhH) motifs, commonly found in sequence-independent DNA-interacting enzymes (Doherty *et al*., 1996). Similar to HhH, HTH motifs were shown to have indispensable roles for DNA recognition and binding. Since the very first crystal structures of FEN1:DNA complexes were published, HTH motifs were found to interact with the downstream duplex DNA forming the primary FEN1:DNA interaction interface (AlMalki *et al*., 2016; Anstey-Gilbert *et al*., 2013; Devos *et al*., 2007; Orans *et al*., 2011; Tsutakawa *et al*., 2011). Although in FEN1 enzymes derived from diverse organisms, these HTH motifs can differ in length and turn number of the loop linker bridging the two α-helices, their direct and indirect interactions with dsDNA are conserved. These interactions are mostly with the template strand of the dsDNA, or what is referred to as parental strand in DNA replication and are mediated by salt bridges with either K^+ or Na^+ ions. For example, in HsFEN1 helices α10 and α11 compose a H2TH motif, in which residues 237-244 are involved in interactions with the downstream dsDNA. Only S237, I238 and I241 interact indirectly with the dsDNA through K^+ bridges.

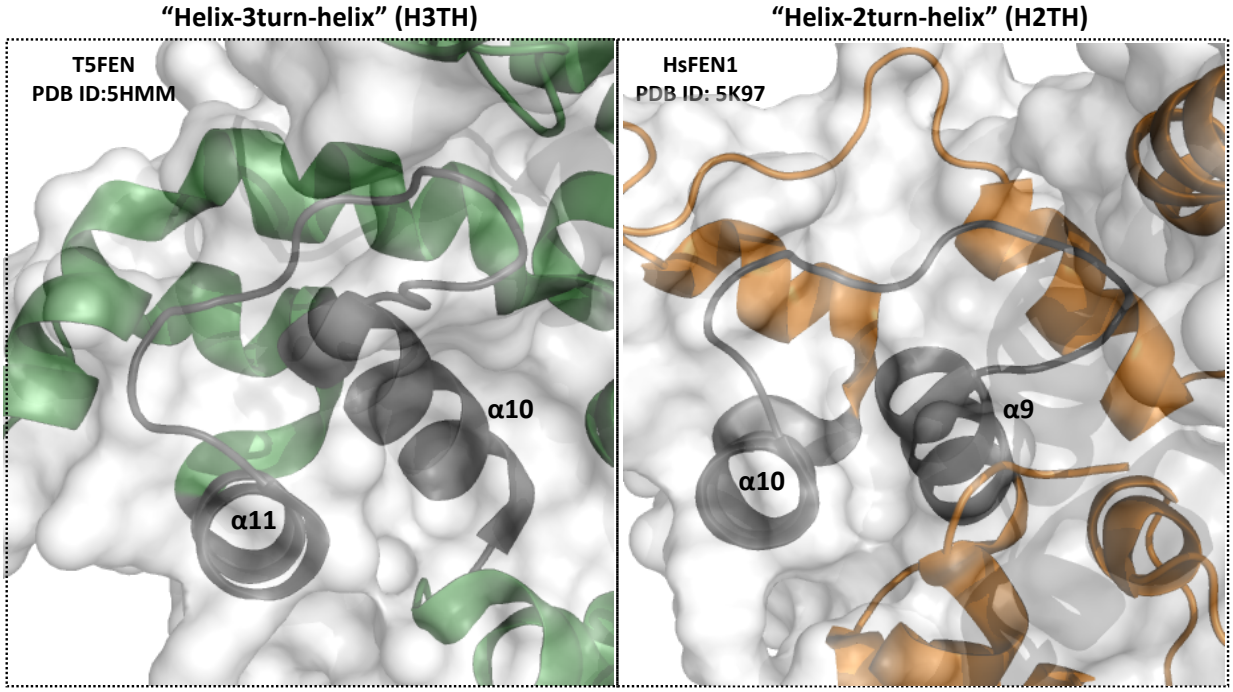


Figure 1.5 "Helix-turn-helix" motifs in HsFEN1 and T5FEN enzymes. Close-up views of the important for structure-specific DNA recognition and binding in HsFEN1 (orange cartoon) and T5FEN (green cartoon) H2TH and H3TH motif, respectively. For both proteins, the respective α-helices and various turn loops, composing the characteristic HTH motifs are shown in grey cartoon. Figures were generated in PyMOL™ (Schrödinger, LLC), using the deposited in PDB online database structures (HsFEN1 – Tsutakawa *et al*., (2017); T5FEN – AlMalki *et al*., (2016)).

1.3.4 Metallobiochemistry of FEN

Extensive work has been done to validate the metal-ion requirements for both exonucleolytic and endonucleolytic FEN catalysis *in vitro* (Feng et al., 2004b; Garforth et al., 1999). Although Mg^{2+} is usually the preferred divalent metal for optimum nuclease activity, likely due to the relatively small atomic radius and high intracellular abundance, a wide variety of metal co-factors can in fact be utilised by FEN (Mg^{2+}, Mn^{2+}, Ca^{2+}, Co^{2+}, Zn^{2+}, Ni^{2+}, etc.). Mutational analyses of conserved residues in metal-binding sites revealed differing affinities for the two viable metal co-factors (Feng *et al*., 2004b; Ghosh et al., 2020). Mutations in the binding site of the first metal ion (Me_A^{2+}), which typically binds in higher affinity than Me_B^{2+}, abolished completely flap endonuclease catalytic activity. On the other hand, mutations in Me_B^{2+} binding site retained only endonucleolytic activity.

Table 1.3 Overview of HsFEN1 alanine mutants and their exonucleolytic, endonucleolytic and DNA binding capacities [1]

Mutation	Exonuclease activity[2]	Endonuclease activity[2]	DNA binding[2]	Reference
R29A	+	+	+	(Qiu *et al*., 2002)
D34A	nd	-	+	(Shen *et al*., 1996)
R47A	-	-	-	(Qiu *et al*., 2002)
R70A	-	-	-	(Qiu *et al*., 2002)
R73A	+	+	+	(Qiu *et al*., 2002)
K80A	+	+	+	(Qiu *et al*., 2002)
D86A	nd	-	+	(Shen *et al*., 1996)
R103A	+	+	+	(Shen *et al*., 1996)
E158A	nd	-	-	(Shen *et al*., 1996)
D179A	+	+	+	(Shen *et al*., 1996)
D181A	nd	+	+	(Shen *et al*., 1996)
G231A	nd	-	-	(Shen *et al*., 1996)
D233A	nd	-	-	(Shen *et al*., 1996)

[1]The table was generated based on the information deposited in UniProt database (UniProtKB: P39748).
[2]Plus (+) and minus (-) symbols are indicative of effective and impaired/ineffective enzymatic activities, respectively, whereas nd corresponds to not determined effects.

FEN homologs are paradigms of metal ion-dependent nucleotide phosphodiesterases family, due to their ability to catalyse the breakdown of targeted phosphodiester bonds by the two-metal, in-line hydrolysis mechanism. Typically, the scissile phosphate of DNA backbone is attacked by the hydroxyl group of either a water or sugar molecule, called nucleophile, which is deprotonated (Figure 1.6). An intermediate pentacovalent product is then formed by the intervention of two metal ions, positioned on either side of scissile phosphate and joint together through the non-bridging phosphate oxygen and the carbonyl oxygens of one of the conserved active site carboxylates (Figure 1.6). The five-coordination geometry of the Me_B^{2+} entails

unfavourable entropy, which is believed to destabilize the intermediate product resulting in reprotonation of 3′-oxygen and formation of a new phosphoryl bond between 5′-phosphate group and the nucleophile (Yang *et al*., 2006) (Figure 1.6.).

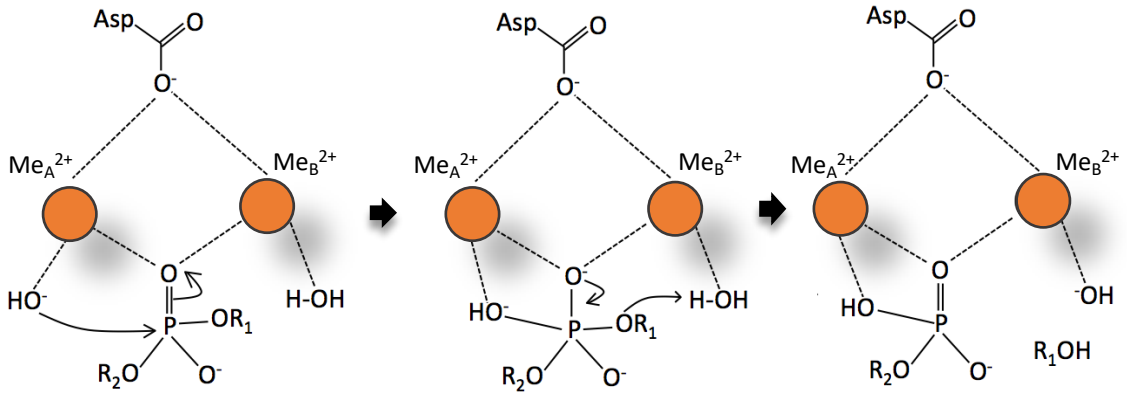


Figure 1.6 Generalised two metal ion-dependent catalysis of DNA. Schematic representation of the two metal ion-dependent phosphodiester bond hydrolysis. Figure was adapted by Roxanne A. Lau (unpublished) and modified by author based on Yang *et al*. (2006) and Dupureur, (2010).

However, in order for this two-metal, in-line hydrolysis mechanism to take place, the distance between the two metal ions has to be maintained ≤4 Å (Yang *et al*., 2006). Metal distances >4 Å are typically considered incompatible, and, given that most FEN proteins were crystallised with >4 Å distance between their two active site metals in their DNA-free structures, controversies with regards to the two-metal ion mechanism prevailed for years. Kinetic studies on the basis of T5FEN have also confirmed the requirement of a third metal ion (Me_C^{2+}) for FEN1 catalysis, in lower, however, concentrations compared to Me_A^{2+} and Me_B^{2+} (Syson *et al*., 2008). Since FEN apo structures only revealed the presence of a maximum of two divalent metals, bound to the active site, Me_c^{2+} was generally considered to possess a lower association constant for enzymes lacking substrate and only interact in the presence of DNA (Syson *et al*., 2008).

Crystallisation of T5FEN, and subsequently HsFEN1 and HsEXO1, with DNA have indeed revealed the presence of a third divalent metal (Me_c^{2+}) in the active site of an intermediate protein:DNA complex. Whether there is or not active involvement of this third metal in FEN1 catalysis and how is transiently recruited, remains unknown. Molecular dynamic simulations in HsEXO1 however, have showed a functional role for Me_c^{2+} and proposed an Me_c^{2+}-mediated mechanism, likely to be conserved among the members of 5′-nuclease superfamily. Consistent to previous studies, Me_c^{2+} is not required for processing of 5′-flap (Algasaier *et al*., 2016). Instead, is required for the expulsion of the resulting monophosphate product from the active site, following 5′-flap cleavage. Its intermitted binding/unbinding is controlled through mobile arch motions and the sophisticated translocation of a conserved glutamic acid residue (E89 in HsEXO1 and E96 in HsFEN1), who's β carboxyl group releases/occupies the respective Me_c^{2+} position within the catalytic site (Donati *et al*., 2020).

1.4 The catalytic mechanism of DNA binding and cleavage by FEN1

1.4.1 Model 1: DNA tracking by FEN1

How does FEN1 select, binds and process its 5′-flap DNA substrates has been a long-standing controversy. Paradoxically, FEN1 specificity for discontinuous ssDNA-dsDNA junctions does not match that of the wider 5′-nuclease superfamily members, which selectively target continuous DNA junctions (e.g. XPG selectivity for DNA bubbles, GEN1 selectivity for four-way Holliday junctions). How can the individual members of 5′-nuclease superfamily support such a diverse substrate specificity, despite their conserved structural fold? The answer, has, for years, been wavering over two predominant models: the proposed tracking and threading models.

Firstly proposed by (Murante *et al*., 1995), the tracking model was supported by several kinetic studies on FEN1 with DNA substrates containing chemically modified single-stranded 5′-flaps, whose cleavage was prevented. Inhibition of primer- and protein-bound 5′-flap cleavage by FEN1, suggested that FEN1 initially binds to the 5′-flap and slides along towards the ssDNA-dsDNA junction, where cleavage is expected to occur (Bornarth *et al*., 1999). To rule out the possibility of a stereochemically obstructed cleavage of these modified 5′-flaps with the excessively large adducts used (e.g. streptavidin, a 60 kDa protein), Bambara and co-workers examined the cleavage possibility of platinum-bound 5′-flaps, whose platinum adducts are typically much smaller in size (Bornarth *et al*., 1999). The inability of FEN1 to cleave these 5′-flaps, in addition to the completely inert loop and bubble DNA substrates to FEN1 cleavage, set the requirement of a free annealed 5′-end reinforcing the possibility of a tracking-based mechanism by FEN1 (Liu *et al*., 2004; Robins *et al*., 1994). These along with the assumption that the conserved helical gateway, is likely to be the main DNA-binding interface, which is of sufficient size to support tracking of the ssDNA, but not dsDNA, seemed to ensure selectivity over the continuous DNA junctions and ′shield against non-specific nucleolytic cleavage (Figure 1.7; Balakrishnan and Bambara, 2013; Balakrishnan et al., 2010; Xu et al., 2001).

If however, what was believed to be the main DNA binding interface cannot support tracking along dsDNA, then how does FEN1 exhibit its exonucleolytic activity? How does FEN1 move and cleave 1-nt onto the ssDNA-dsDNA junction? An alternative mechanism could perhaps explain these phenomena, while explaining substrate specificity and precise cleavage.

1.4.2 Model 2: DNA threading by FEN1

Confusingly, the same functional data on chemically modified 5′-flaps, supporting the FEN1 tracking model, have for years also supported an alternative threading model. Whilst some modified 5′-flaps with excessively large adducts appeared to be completely resistant to FEN1 cleavage, others with moderately large secondary structures, including the aforementioned platinum-, RNA- or protein- based adducts, appeared to be cleavable (Bornarth *et al*., 1999; Murante *et al*., 1995; Tsutakawa *et al*., 2011; Xu *et al*., 2001). Therefore, the incomparable

cleavage efficiencies of diverse 5′-flaps, including gapped 5′-flaps, could not have excluded the possibility of a FEN-threading model, in which FEN1 targets the ssDNA-dsDNA junction and the annealed 5′-flaps thread through the characteristic helical gateway for processing (Gloor *et al*., 2010). However, if threading requires some level of flexibility, which is might be required to grasp an annealed free 5′-end, then the reported inhibition of FEN1 cleavage for the removal of overly modified bulky 5′-ends, could be largely explained. Although this cannot rule out the possibility of a tracking-based FEN model, the emergence of additional functional data further supported the threading-mediated cleavage of the 5′-flaps, which was believed to be the most prevalent (Balakrishnan *et al*., 2010; Gloor *et al*., 2010).

In an effort to address the kinetics behind cleavage of 5′-flaps of diverse lengths, by a bacterial FEN homologue, DNA Polymerase I, Joyce and co-workers found inefficient processing of a 30-nt long 5′-flap compared to one with only a 10-nt long flap (Xu *et al*., 2001). This led to a rather interesting reasoning, as to why cleavage of a 10-nt long 5′-flap is significantly more efficient. Assuming that the reduced cleavage efficiency depends on the formed FEN1-DNA complex, then tracking along a 10- or 30-nt long 5′-flap should not affect the cleavage efficiency. On the other hand, threading of a 30-nt long 5′-flap with more degrees of freedom compared to a 10-nt long, with increased energy requirements and therefore, could explain varying cleavage efficiency (Balakrishnan and Bambara, 2013). In addition to this, the threading-based mechanism was further supported when the base of the 5′-flap was found to be an important determinant at the FEN1-DNA interface, with a binding affinity independent of the 5′-flap length or its modification with secondary adducts. As a result, FEN1 could catalyse its 5′-flap substrates, whilst bound to the replication toolbelt (Polδ-FEN1-PCNA complex), and maximize efficiency by binding to the flap substrate, whilst DNA intermediates, expected to block threading, are being repaired (Sakurai *et al*., 2005).

However, how are some modifications on the 5′-flaps, bulky enough not to be threaded through the helical gateway, being tolerated? That remained unclear, until the first crystal structures of HsFEN1:DNA complexes were published (Tsutakawa *et al*., 2011). Although none of these structures showed threading of the 5′-flap through the helical gateway, co-crystallization of a pre-threaded and processed flap substrate with HsFEN1 indicated the adoption of an "ordered" arrangement by the helical gateway. Base helices (α2 and α4), as well as the cap helices (α4 and α5), were found to switch over from a "disordered" to a narrower "ordered" structural arrangement upon threading (Tsutakawa *et al*., 2011. Previous structures of the bacteriophage T4 RNaseH bound to a pseudo-Y DNA substrate also supported that model, suggesting threading of 5′-flap through a completely disordered loop (Mueser *et al*., 1996). Although the model lacks six residues within the helical gateway, it appeared to allow tight interactions between the phosphate backbone of the ssDNA and the active site metals. Although the model lacks six residues within the helical gateway, it appeared to allow tight interactions between the phosphate backbone of the ssDNA and the active site metals. The existence of a disordered (absence of a threaded DNA substrate) to ordered (presence of a threaded DNA substrate) transition of the helical gateway, provided a valid reasoning as to the observed tolerance for

some substrate modifications. In addition, it seemed to explain how specificity to the ssDNA-dsDNA junctions and threading of strictly ssDNA is achieved.

After 25 years of debate, Almalki and co-workers published the first crystal structures of a bacteriophage flap endonuclease (T5 D15 exonuclease, also known as T5FEN), with an intact 5′-flap DNA strand threading through its helical gateway (AlMalki *et al*., 2016). Following a unique "metal-mimic" mutagenesis approach, a catalytically inactive T5FEN mutant (D153K), only capable of DNA binding but not catalysis, was generated for crystallization trials with substrate DNA. Structures of a pre-threaded DNA:T5FEN and a fully-threaded DNA:T5FEN, solved at 1.9 and 2.2 Å resolution, respectively, showed threading of 5′-flap through an ordered helical gateway. This supported the proposed "disordered-to-ordered" transition of the helical gateway, while it appeared to be critical for catalysis by bringing the DNA backbone in close proximity with the enzyme's active site (AlMalki *et al*., 2016).

Whether this was a conserved catalytic mechanism among the structurally similar members of 5′-superfamily, was unknown at the time. However, later structures of both HsEXO1 and HsFEN1 with intact DNA threading through, provided a clear answer (Shi et al., 2017; Tsutakawa *et al*., 2020; Tsutakawa et al., 2017).

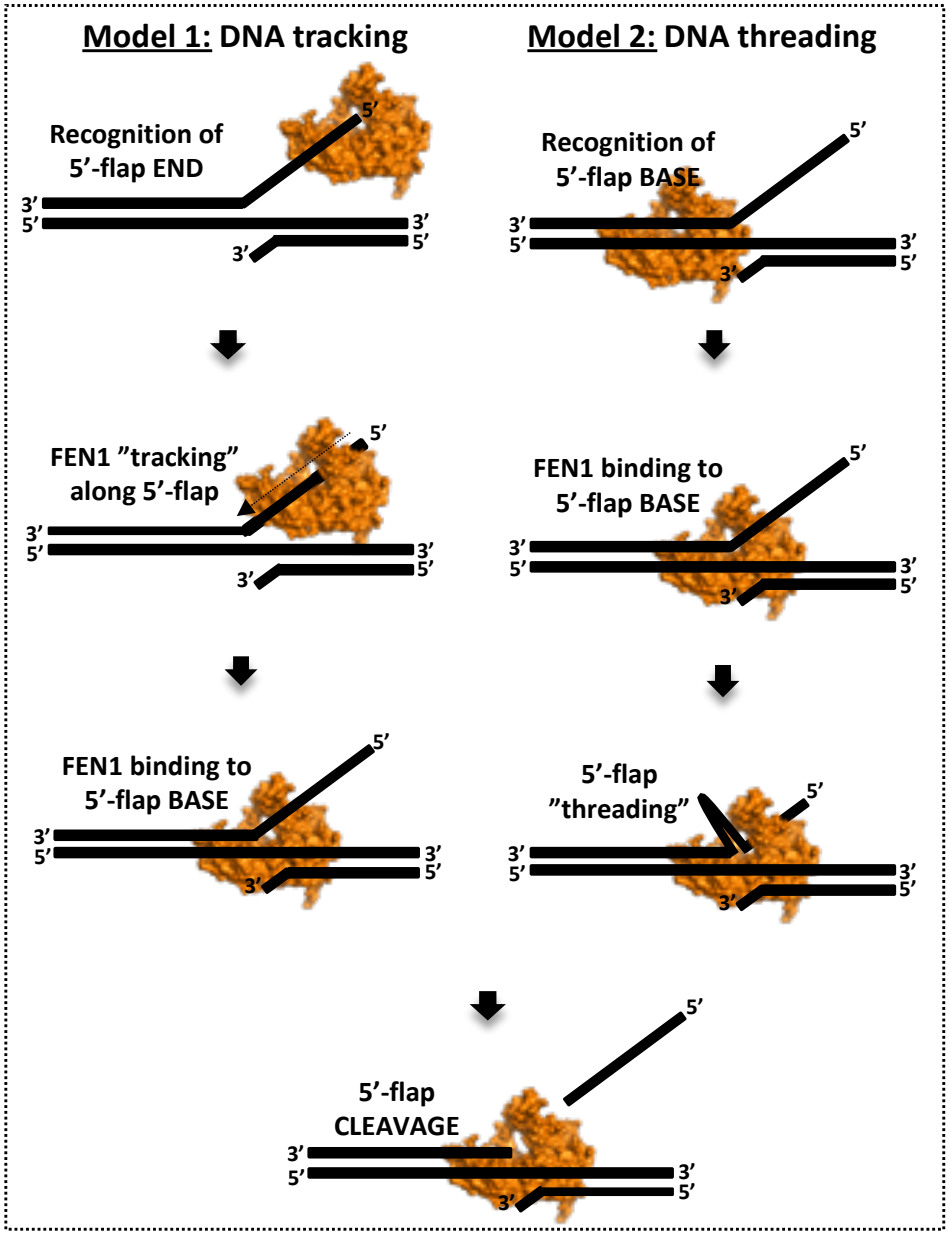


Figure 1.7 Competing DNA tracking and threading models for FEN1 catalysis. Orange surface representation of HsFEN1 macromolecule (PDB ID:3Q8L; Tsutakawa *et al*., (2011)). According to the proposed "DNA tracking" model, the 5′-flap end is recognized by FEN1, which tracks along the single-stranded 5′-flap until it reaches the ss-ds junction for cleavage. On the other hand, threading model suggests recognition and direct binding of FEN1 to the 5′-flap base from which 5'-flap threads through in a needle-like conformation. Figure was reproduced based on (Gloor *et al.*, 2010).

1.4.3 Threading-mediated catalysis by HsFEN1

Aiming to understand the inter- and intramolecular interactions driving recognition and cleavage 5′-flap structures in *Homo sapiens* (Hs)FEN1 helical gateway, Tsutakawa *et al*., (2017) published the first crystal structures of HsFEN1 with DNA threading through the helical gateway. Co-crystallization of three C-terminally tagged active site mutants of HsFEN1

(D86N, R100A and D233N) with double-flap DNA substrates of diverse length 5′-flap lengths, provided a mechanistic insight into HsFEN1 catalysis (Tsutakawa *et al*., 2017).

The primary HsFEN1-DNA interface was found to involve interactions with the dsDNA, and not the annealed 5′-flap, as was implied by the tracking mechanism. Four separate regions within HsFEN1 were shown to be involved in these interactions with the dsDNA: (a) the H2TH motif, (b) the hydrophobic wedge, (c) the active site and (d) a helix-hairpin-helix motif toward the C-terminus (Figure 1.9). Only the H2TH motif appears to interact with the downstream dsDNA, with interactions mediated by the presence of K^+ and positively charged residues (Figure 1.9). The helix-hairpin-helix motif was found to interact with the upstream dsDNA, whereas the hydrophobic wedge and the active site interact with the dsDNA just next to the bended single-stranded 5′-flap (Figure 1.9).

All three structures also supported a "one helical turn apart" spacing between the hydrophobic wedge and the H2TH motif, positioned on either side of the dual metal active site (Figure 1.3). This rather wide spacing allows selection for discontinuous ssDNA-dsDNA junctions and their extended 5′-flaps, against the continuous ssDNA junctions and the 1-nt long 3′-flaps, which are accommodated at the so called 3′-flap binding pockets. In addition, it allows shifting of the 5′-flap towards the active site, for threading through an ordered helical gateway, formed by both cap and base helices (α2, α4 and α5), consistent to previous findings (Figure 1.9).

Threading however, was shown to be mediated by steering of the single-stranded 5′-flap so that the DNA bases are oriented towards the dual metal active site (Figure 1.8). Although structures of T5FEN and HsEXO1 with DNA threading through the helical gateway support a conserved catalytic mechanism, the inverted threading of the single-stranded 5′-flap with the DNA bases oriented towards the active site metals appears to be a unique feature of HsFEN1. Inversion seems to be promoted by basic residues within the helical gateway (R129, R103, R104 and K132), whose electrostatic interactions are believed to energetically control the orientation of the phosphodiester backbone for cleavage. This was further supported by mutational analyses. Substitution of all four basic residues (R129, R103, R104 and K132) by an amino acid of repulsive charge, glutamate, led to an 18,000-fold reduction in HsFEN1 activity compared to the WT. Strikingly, the simultaneous 17-fold increase in K_d, suggests that the massively decreased HsFEN1 activity is attributed to impaired positioning of the DNA substrate and not globally deficient binding.

Shifting of the scissile phosphate bond towards the active site to maintain the distance requirements enabling catalysis, is critical. Whether this shifting is accommodated by double-, single- or no-unpairing of the surrounding bases, remains as yet enigmatic. Despite the use of the same duplex DNA template for crystallization, all three structures were found to have differences on base-pairing/unpairing, the position of scissile phosphate bond and the orientation of Y40 (Figure 1.8). Duplex DNA was found to widen in all structures, while approaching the active site in catalytic distance. However, in D86N HsFEN1-DNA complex the bases surrounding the scissile phosphate remained paired, whereas in R100A HsFEN1- and

D233N HsFEN1-DNA complexes were found to be partially and fully unpaired (Figure 1.8). Although this could be attributed to the induced mutations, it is also likely to be a result of the 5′-flap and the scissile phosphate.

Threading-mediated catalysis appears once again to be the predominant mechanism among the members of 5′-nuclease superfamily. Cleavage of single-stranded 5′-flaps by HsFEN1 occurs through a unique inverted threading, not required for T5FEN and EXO1 (Figure 1.9; AlMalki *et al*., 2016; Shi *et al*., 2017; Tsutakawa *et al*., 2020). Despite the improved understanding of HsFEN1 catalysis, provided by the structures of Tsutakawa *et al*., (2017), distinct differences in double base unpairing as duplex DNA reaches the active site set the requirement for future re-evaluation.

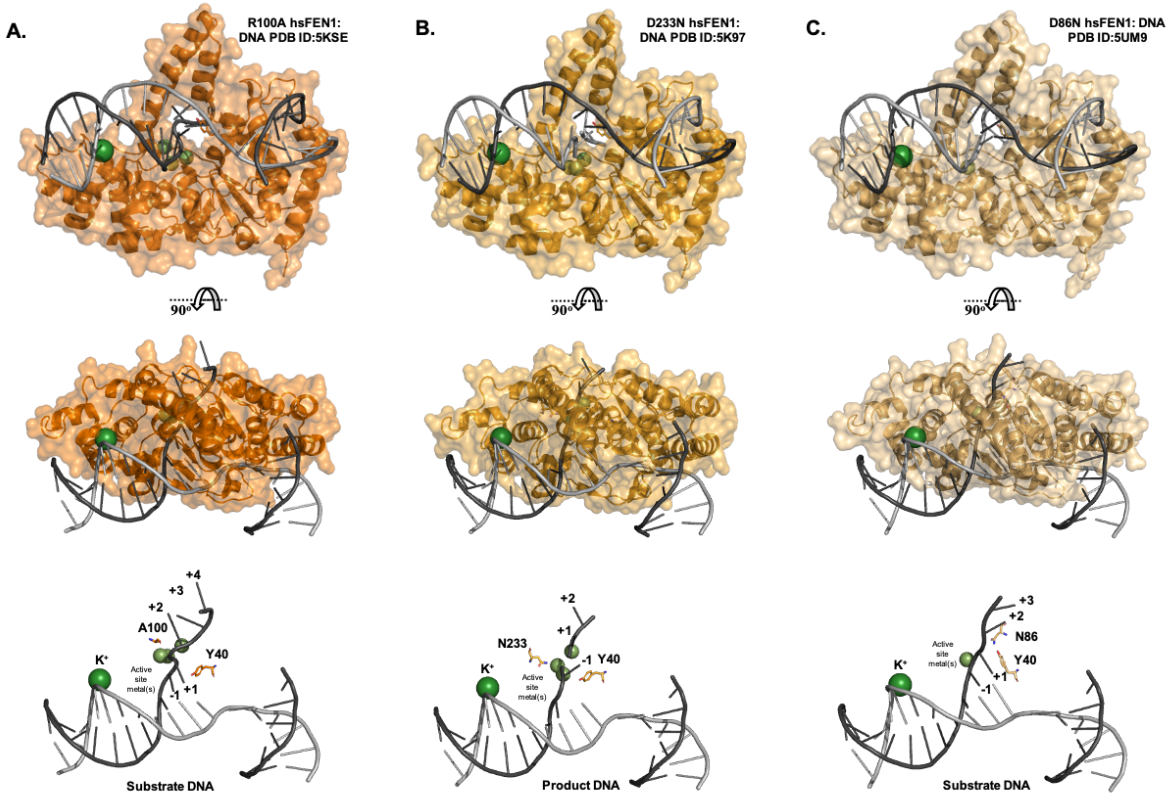


Figure 1.8 Threading-mediated catalysis by HsFEN1. Threading of 5′-flap through the helical gateways of three HsFEN1 mutants, all published by Tsutakawa *et al*., (2017): (A) R100A HsFEN1 (orange; PDB ID: 5KS3), (B) D233N HsFEN1 (yellow; PDB ID: 5K97) and (C) D86N HsFEN1 (sand; PDB ID: 6UM9). Each complex is shown in front and top views, highlighting threading of 5′-flap and its respective orientation with regards to the active site metals. The crystallized active site Sm^{3+} (pale green) and K^{+} (green) ions are shown in spheres. In all cases, mutated residues and Y40, indicating the differing rotamer orientations tracking the incised DNA movement, are shown in sticks. Isolated DNAs are representative of the

diverse base pairing/ unpairing accommodating threading of 5′-flaps in the individual mutant HsFEN1:DNA structures. Structures (3q8k etc) were rendered using PyMOL™ .

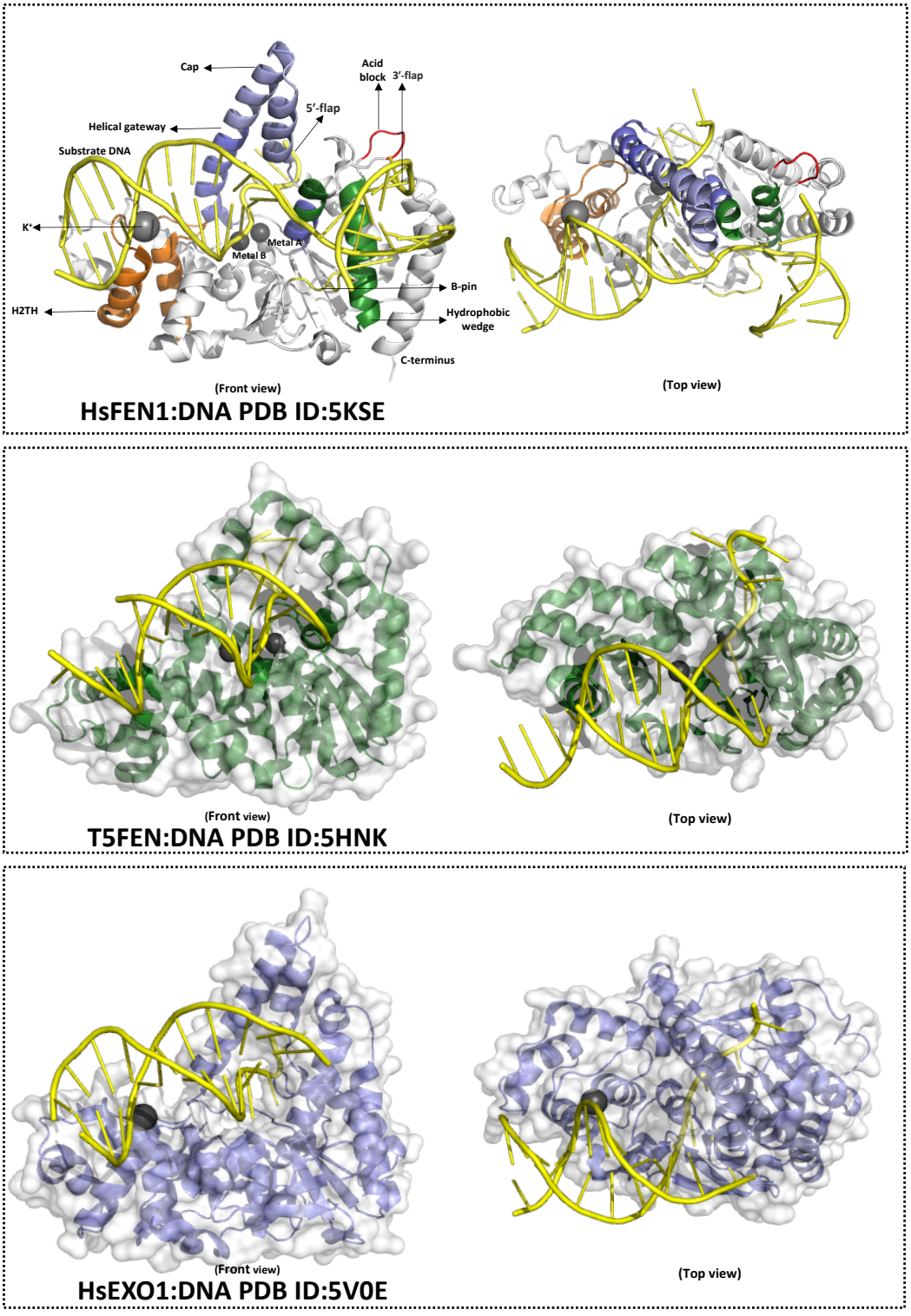


Figure 1.9 Conserved threading-mediated catalysis by FEN1 homologs. Structures of catalytically inactive HsFEN1, T5FEN, and HsEXO1 mutants indicating threading of single-stranded 5'-flaps through an ordered helical gateway for processing. Grey spheres are representative of the co-crystallized, in the corresponding complex structures, potassium and the active site metal ions mediating protein-DNA interactions. Figures were generated in PyMOL™ .

1.5 Modulation of FEN1 activity

DNA replication and repair proteins act synergistically to minimise the overall error rates during DNA synthesis, maintaining genome fidelity. Intermolecular protein-protein interactions are well-known for sequential timing and regulation of the vast majority of these processes. Several FEN1-protein interaction partners, including post-translational modifiers, have been reported to tightly control FEN1 activity. FEN1-protein interactions pairs for which, with the only exception of FEN1-PCNA, little is known about their specific intermolecular interactions from a mechanistic point of view.

1.5.1 Post-translational modification of FEN1

Protein fold, stability and localization are well-known to be controlled by post-translational modifications (PTMs), often acting as a sequential on/off switch of proteins to ensure mechanistic precision and genomic integrity are maintained. Several PTMs have been described to regulate the function of FEN1. Modification-specific proteomic studies and functional studies on specific post-translationally modified FEN1, revealed an array of modified residues that either activate or degrade FEN1 in a cell cycle-dependent manner (Table 1.4).

Table 1.4 Overview of known[1] PTMs for modulation of FEN1 function

Post-translational modification	Modification site	Effect	Reference
Methylation	R19	No impact on FEN activity & FEN1-PCNA interaction	(Guo *et al.*, 2010)
Acetylation	K24	↓ DNA binding & FEN activity	(Xu *et al.*, 2016)
Acetylation	K80	nd	(Choudhary *et al.*, 2009)
Methylation	R100	No impact on FEN activity & FEN1-PCNA interaction	(Guo *et al.*, 2010)
Methylation	R104	No impact on FEN activity & FEN1-PCNA interaction	(Guo *et al.*, 2010)
Acetylation	K125	↓ DNA binding & FEN activity	(Xu *et al.*, 2016)
SUMOylation	K168	FEN1 degradation	(Guo *et al.*, 2012)
Phosphorylation	S187	Abolished PCNA binding	(Guo *et al.*, 2010; Henneke *et al.*, 2003)
Methylation	R192	No impact on FEN activity & FEN1-PCNA interaction	(Guo *et al.*, 2010)
Phosphorylation	S197	FEN1 inactivation	(Olsen *et al.*, 2010)
Succinylation	K200	No impact on FEN activity & FEN1-PCNA interaction	(Shi *et al.*, 2020; Xu *et al.*, 2016)
Acetylation	K252	↓ DNA binding & FEN activity	(Xu *et al.*, 2016)
Acetylation	K254	↓ DNA binding & FEN activity	(Xu *et al.*, 2016)
Phosphorylation	S255	nd	(Zhou *et al.*, 2013)
Phosphorylation	S293	nd	(Zhou *et al.*, 2013)
Acetylation	K314	↓ DNA binding & FEN activity	(Xu *et al.*, 2016)
Phosphorylation	T336	nd	(Zhou *et al.*, 2013)
Acetylation/ SUMOylation/ Ubiquitination	K354	↓ DNA binding & FEN activity	(Guo *et al.*, 2012; Hasan *et al.*, 2001)
Phosphorylation	T364	↓ DNA binding & FEN activity	(Olsen *et al.*, 2010)
Acetylation	K375	↓ DNA binding & FEN activity	(Hasan *et al.*, 2001)
Acetylation	K377	↓ DNA binding & FEN activity	(Hasan *et al.*, 2001; Thandapani *et al.*, 2017)
Acetylation	K380	↓ DNA binding & FEN activity	(Hasan *et al.*, 2001)

[1]PTMs modulating FEN1 function, published to date. Modification sites derive from modification-specific proteomic studies and/or functional studies of post-translationally modified FEN1. Data refer to *in vitro* and/or *in vivo* assays. Not all modification sites stated above have been characterized for their functional effect on FEN1 cleavage and binding capacity.

The first PTM proposed to regulate FEN1 function was acetylation by the transcriptional co-activator p300. Through its histone acetyl transferase activity, p300 catalyses acetylation of FEN1 at the C-terminus (K354, K375, K377 and K380), leading to a significant reduction of both exo- and endo-nucleolytic capacity of FEN1 (Hasan *et al*., 2001). Despite the reduced activity, PCNA binding of acetylated FEN1 remained unaffected suggesting defects in DNA binding, rather than interactions with PCNA and other FEN1-interacting proteins (Hasan *et al*., 2001). As for its biological importance, inactivation of FEN1 and simultaneous activation of acetylated Dna2 was proposed to promote longer flap formation ensuring precise removal of mismatched nucleotides during the error-prone synthesis by DNA Polα (Balakrishnan *et al*., 2010).

FEN1 could also serve as a direct substrate for cyclin-dependent kinases (Cdk1 and Cdk2), when in complex with their regulatory subunit cyclin A. Accumulation of a defective to phosphorylation FEN1 mutant, S187A, during the late S phase suggested that phosphorylation by cdk1/2-cyclin A occurs at the late S-phase, while S187 was recognized as the primary phosphorylation site within FEN1 among those tested (Henneke *et al*., 2003). Although phosphorylated FEN1 was found to have decreased both its exo- and endonucleolytic activities, its DNA binding capability was not affected significantly. However, binding to PCNA, and perhaps other known FEN1-interacting proteins, was completely abolished. Given that S187 is located internally at the FEN1 nuclease domain, inhibition of FEN1 activity due to post-translational phosphorylation has been attributed to conformational changes likely to be taking place.

Interestingly, later studies showed that FEN1 degradation is highly dependent on FEN1 phosphorylation (Guo *et al*., 2012). SUMOylation of phosphorylated FEN1, by SUMO, and subsequent stimulation of ubiquitination, by UBE1/UBE2M/PRP19 complex, were found to lead to FEN1 degradation. This ubiquitin-proteasome pathway is well-known for degradation of several endogenous proteins at the late S-phase. Typically, expression of endogenous FEN1 increases at the G1-phase, reaches its peak at the S-phase and dramatically decreases from the late S-phase to G2/M-phase, where DNA replication is completed and removal of excess FEN1 is required. Inhibition of this ubiquitin-proteasome pathway both through pathway inhibitors and mutant FEN1, defective to phosphorylation, led to cell cycle delays and characteristic genomic instability phenotypes (e.g. chromosomal segregation and polyploidy). These data suggest that ubiquitin-proteasome pathway is the primary clearance mechanism of phosphorylated FEN1, whose dysfunction is likely to be associated with FEN1 overexpression in cancer and linked to cancer etiology.

Stimulation of FEN1 SUMOylation was recently shown to be highly dependent on succinylation of phosphorylated FEN1 with critical roles in DNA repair (Shi *et al*., 2020). Replication fork stalling, either using UV radiation or well-established chemical agents, was found to stimulate succinylation of phosphorylated FEN1 mainly during G1 phase after its demethylation in S phase. Succinylation promotes in turn SUMOylation of FEN1, which

stimulates the GEN activity of FEN1, while it significantly enhances its interactions with some of its known DNA repair interaction partners.

Among the PTMs regulating FEN1 function, phosphorylation and methylation were found to antagonistically compete each other (Guo *et al*., 2010). Methylation occurs mainly at R192, by arginine methyltransferase 5 (PRMT5), and unlike acetylation and phosphorylation, it does not negatively impact the activity of FEN1 and its binding capacity to DNA and FEN1-interacting proteins. In fact, methylation of FEN1 was found to facilitate binding to PCNA with affinities similar to those of un-methylated FEN1. Methylation of R192 was shown to be critical for FEN1 localization at the DNA replication loci with a highly phosphorylated methylation-defective FEN1 mutant found to be accumulated in the S-phase of the cell cycle.

However, the conformational changes likely to arise as a result of these PTMs and how these changes might affect DNA binding and cleavage, from a mechanistic point of view, are largely unknown. Xu *et al*., (2018) published the first crystal structure of a mutant analogous to previously proposed R192 methylated FEN1. The primary FEN1 methylation site, R192, was substituted by phenylalanine (F) believed to mimic the β-pin hydrophobicity induced by di-methylation of R192. Helices β8 and β9, composing the β-pin region, were found to undergo a unique loop-to-helix transition, not seen in previous FEN1 structures (Figure 1.10). Given that β-pin is involved in interactions with the upstream dsDNA, this loop-to-helix transition is proposed to be critical for DNA-binding and association/dissociation of FEN1 to its interacting protein pairs.

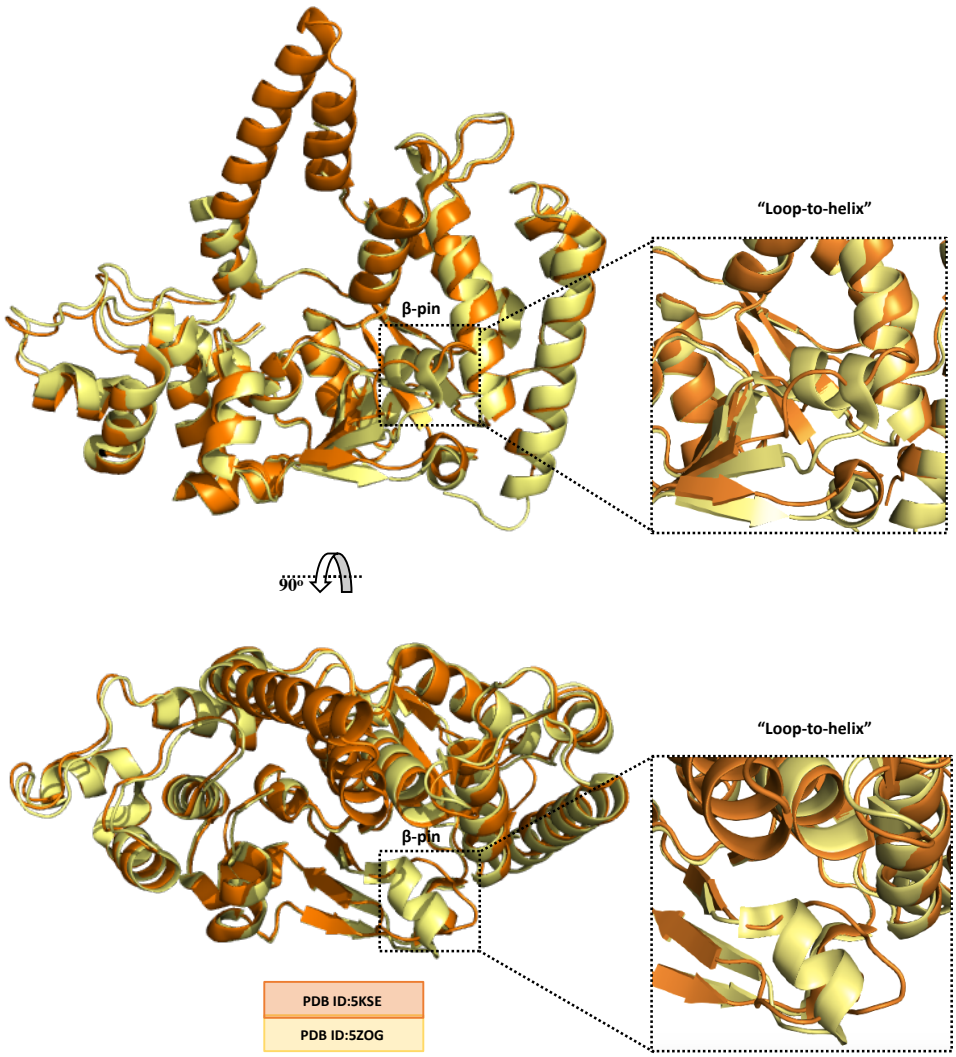


Figure 1.10 "Loop-to-helix" transition of the methylated-mimic R192F FEN1 mutant. Structure of R192F FEN1 mutant (PDB ID: 5ZOG; (Xu et al., 2018), with the "loop-to-helix" transited β-pin region, superimposed to the structure of R100A FEN1 (PDB ID: 5KSE; (Tsutakawa *et al.*, 2017) for direct comparison and rendered using PyMOL™ .

1.5.2 FEN1-protein interaction pairs

1.5.2.1 FEN1-PCNA interaction pair

Among the most widely studied FEN1-protein interaction pairs is that of FEN1 with PCNA, occurring during lagging strand replication. PCNA's role in FEN1 recruitment and coordination of its synergistic action with Polδ and DNA ligase I for efficient processing and maturation of Okazaki fragments has been supported from early on. Following the findings of Wu *et al.*, (1996), who showed that physical recruitment of FEN1 by PCNA stimulates both its endo- and exonucleolytic activity, several structural studies aiming to address the intermolecular interactions driving FEN1 activation and regulation have been performed

(Table 1.5). These initially involved co-crystallisation of PCNA with only small FEN1 C-terminal peptides, which, although they suggested coordination of FEN1 mainly through C-terminal-mediated interactions, they did not shed light on the precise molecular mechanism underlining cooperation between these two proteins (Bruning and Shamoo, 2004; Chapados *et al*., 2004; Duffy *et al*., 2016)

Table 1.5 Overview of published PCNA-FEN1 structures to date

PDB ID	Organism	FEN1 length	Method	Resolution (Å)	Reference
6TNZ	*Homo Sapiens*	Residues 1-352	Electron microscopy (EM)	4.05	(Lancey *et al*., 2020)
5E0V	*Homo Sapiens*	Residues 335-348	X-ray	2.1	(Duffy *et al*., 2016)
1UL1	*Homo Sapiens*	Residues 1-359	X-ray	2.9	(Sakurai *et al*., 2005)
1RXZ	*Archaeoglobus fulgidus*	Residues 326-336	X-ray	2	(Chapados *et al*., 2004)
1U7B	*Homo Sapiens*	Residues 331-350	X-ray	1.88	(Bruning and Shamoo, 2004)

All published structures of PCNA bound to FEN1, or part of it, shown here, are based on the available information deposited in PDB database. Structures are shown in a chronological order, starting with the most recent.

The longest FEN1 fragment (residues 1-359 out of 380 residues, which compose the native full-length HsFEN1) was firstly crystallized with PCNA by Sakurai *et al*., (2005) at a resolution of 2.9 Å (Figure 1.11). Generally, the basic C-terminal residues 360-380 are considered dispensable for PCNA binding, although a defect in DNA binding has been shown in affinity and electrophoretic mobility shift assays (EMSA; (Stucki *et al*., 2001). Three FEN1 molecules were found bound to each of the three monomeric subunits of PCNA, through C-terminal-mediated interactions (Figure 1.11). Despite the diverse conformations of PCNA-bound FEN1 molecules adopted due to differential twisting of the C-terminal residues 333-336 (also referred to as "hinge linker"), an antiparallel interaction between two β-strands in FEN1 and PCNA C-terminus, respectively, was shown to be conserved. Interestingly, a well-established PCNA-binding motif (QXX(I/L/M)XXF(F/Y)), formed by residues 337-344 in HsFEN1, was shown to mediate this antiparallel interaction between the two molecules, consistent with previous studies (Frank *et al*., 2001; Stucki *et al*., 2001). The "hinge linker", as well as the PCNA-binding motif are conserved among eukaryotic FEN1 enzymes suggesting a common interaction mechanism. On the other hand, the non-conserved flanking PCNA-binding motif, X338-339 and X341-342, of PCNA-binding motif, which do not appear to interact directly with PCNA, are thought to be involved in interactions with the negatively-charged phosphate backbone of FEN1 DNA substrate (Stucki *et al*., 2001). This also applies to the basic C-terminal residues 360-380.

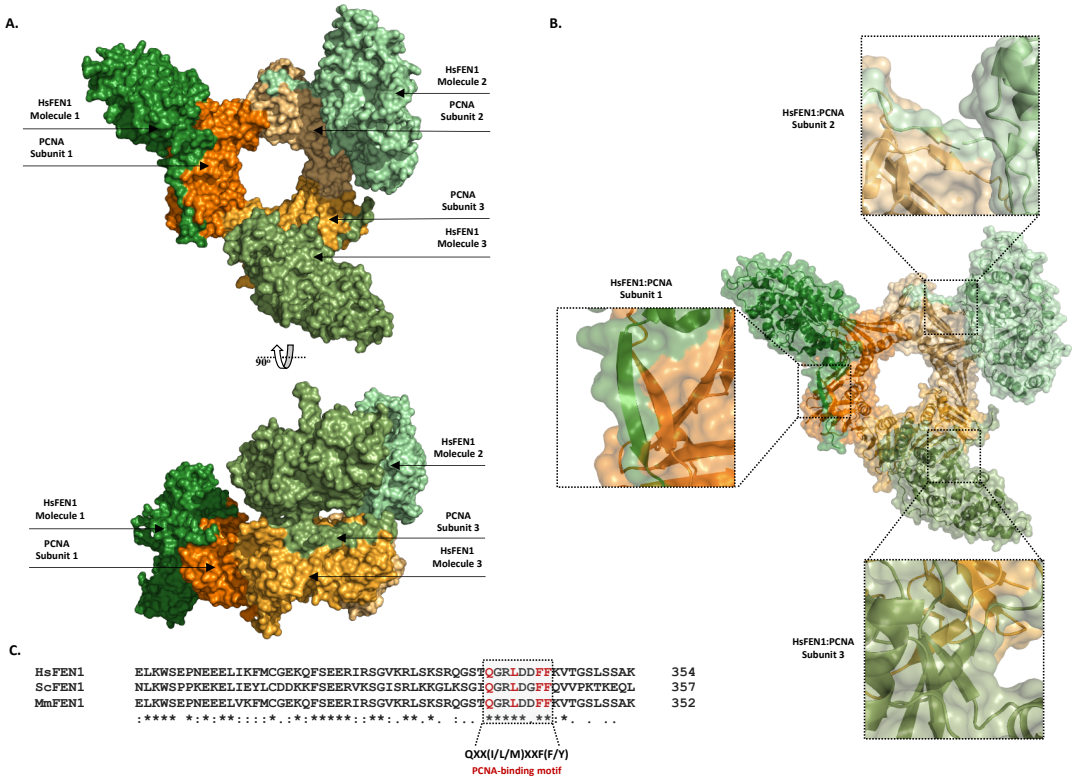


Figure 1.11 Structure of PCNA-FEN1 complex and close-views of selected intermolecular interactions within FEN1 and PCNA C-terminus. (A) Structure of the homotrimeric PCNA-FEN1 complex, as this have been initially published by (Sakurai *et al*., 2005)). Each of the three monomeric PCNA subunits are coloured in shades of orange, and their respective FEN1 molecules bound, are coloured in shades of green. (B) Close-view of the antiparallel interactions between FEN1 and PCNA C-terminus. (C) Sequence alignment of homo sapiens FEN1 (HsFEN1; NP_004102.1), Saccharomyces cerevisiae FEN1 (ScFEN1; GFP67859.1) and Mus musculus FEN1 (MmFEN1; NP_001258544.1) C-terminus, indicating the conserved PCNA-binding motif (QXX(I/L/M)XXF(F/Y)). Structure of PCNA-FEN1 complex (PDB ID: 1UL1; (Sakurai *et al*., 2005) were rendered using PyMOL™ .

Aiming to address the catalytic mechanism underlining the processive 5′-flap cleavage by FEN1 in synergy with Polδ and PCNA, and with the potential to unravel FEN1-DNA interactions upon binding to PCNA and Polδ at the replication fork, Lancey *et al*., (2020) published 15-years later the cryo-EM structure of the entire Polδ-FEN1-PCNA toolbelt complexed with DNA (Figure 1.12). Despite the novel mechanistic insights in the context of Polδ catalysis provided, the absence of electron density maps of FEN1 helical gateway as well as the use of a truncated FEN1 mutant lacking C-terminal residues 337-380 did not reveal any novel interactions of FEN1 C-termini with DNA. Two of the three PCNA monomeric subunits were found to be occupied by Polδ, leaving the last PCNA monomer free for binding to other PCNA interaction partners (Figure 1.12). These include FEN1, which was found bound to PCNA in a seemingly non-analogous way to that observed in the PCNA-FEN1 structure solved by Sakurai *et al*., (2005). Twisting through the flexible "hinge linker" was shown however to result in the same "upright" configuration of FEN1 molecule. Adoption of such configuration

seemed to allow FEN1 to orient for binding to the downstream dsDNA, while the single-stranded DNA template bended 90° towards FEN1 binding site for threading through the helical gateway, consistent to previous studies (Tsutakawa *et al*., 2011).

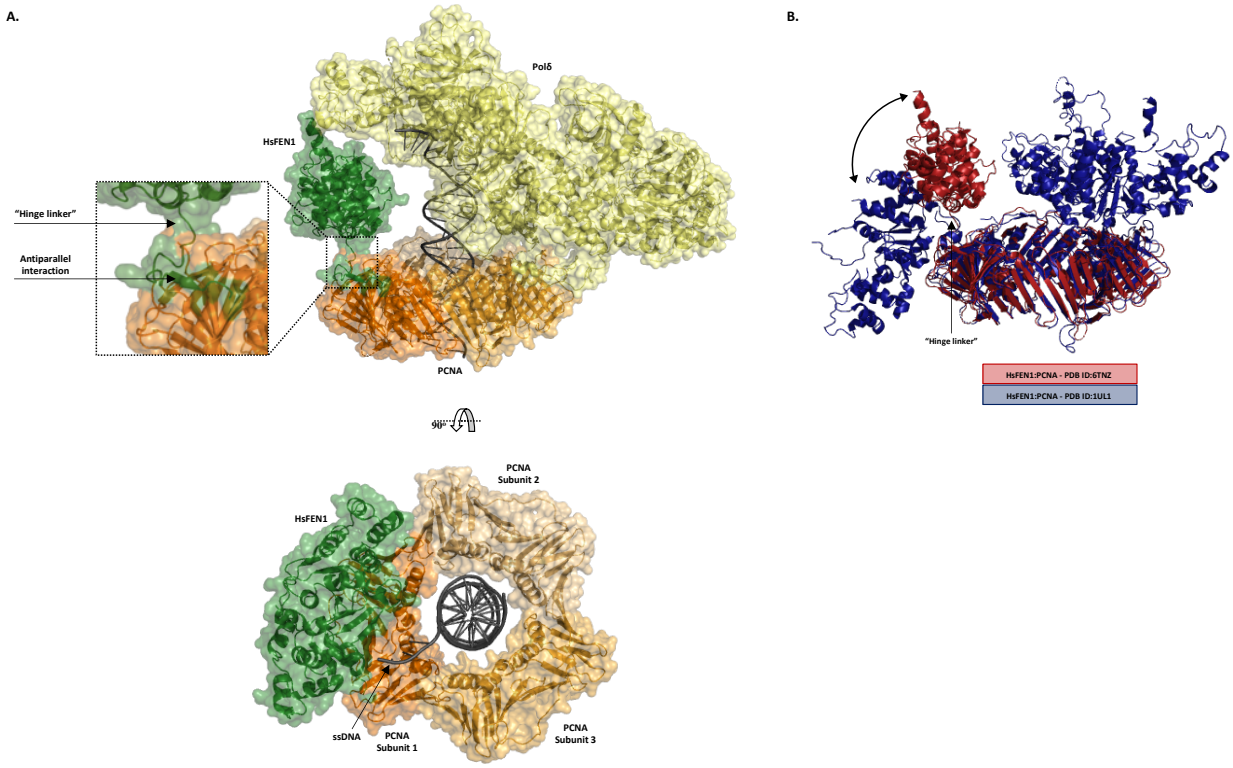


Figure 1.12 Structure of Polδ-FEN1-PCNA toolbelt. (A) Structure of the homotrimeric PCNA (shades of orange) bound to FEN1 (green), Polδ (yellow) and dsDNA (grey), indicative of the DNA replication toolbelt. Close-view highlights the antiparallel interaction between FEN1 and PCNA C-terminus, as well as the flexible "hinge linker". (B) Superimposition of the two available PCNA-FEN1 complex structures, indicating the different FEN1 orientations adopted in the presence and absence of DNA as a result of hinge linker's flexibility. Structures of Polδ-FEN1-PCNA toolbelt (PDB ID: 1UL1; (Lancey *et al*., 2020) and PCNA-FEN1 complex (PDB ID: 1UL1; (Sakurai *et al*., 2005) were rendered using PyMOL™ .

Despite the improved understanding of FEN1 recruitment by Polδ-PCNA complex in Okazaki fragment maturation through a toolbelt mechanism, the unclear involvement of the remaining C-terminal residues in interactions with PCNA and/or the downstream dsDNA set the requirement for further research.

1.5.2.2 FEN1-RecQ helicase family members interaction pairs

Due to its role in processing flap substrates arising as intermediates, not only in DNA replication, but also in multiple DNA repair pathways, FEN1-interacting partners include a range of DNA repair proteins. Among these, are Werner syndrome (WRN) and Bloom syndrome (BLM) proteins, members of RecQ helicase family, both shown to stimulate the

nucleolytic activity of FEN1 (Sharma *et al*., 2004). Overexpression of both WRN and BLM in yeast *dna2* mutants, defective in flap removal during DNA replication, were shown to rescue impaired growth phenotypes upon replication stress (Sharma *et al*., 2004). Biochemical analyses focusing on the intermolecular interactions between FEN1 and WRN showed that cleavage by FEN1 is not only enhanced for double-flap substrates, the optimal DNA substrates of FEN1, but also for alternative single-flap, Holliday junctions and bubble DNA structures, upon WRN interaction (Sharma *et al*., 2004; Zheng *et al*., 2005). Following a chemically-induced DNA damage, FEN1 and WRN were shown to co-localize at stalled replication forks *in vivo* (Sharma *et al*., 2004; Sharma *et al*., 2005). Although there are no crystal structures of FEN1 in complex with WRN to date, functional studies using GST-tagged fragments of FEN1 showed that binding activity of WRN was only retained with FEN1 fragments containing the C-terminal residues 363-380 (Sharma *et al*., 2005). Similarly, a BLM fragment analogous to that of WRN found to interact with FEN1 was also found to interact with FEN1 C-terminal domain with similar affinity (Sharma *et al*., 2005). This C-terminal FEN1 region is adjacent to the C-terminal PCNA binding region of FEN1, suggesting the existence of a complex through which co-ordination of the catalytic activities of these proteins occurs.

1.6 Targeting *Homo sapiens* Flap Endonuclease 1 (HsFEN1) to impede cancer

1.6.1 Role of HsFEN1 in cancer development and progression

Growing evidence has accumulated over the last two decades implicating impaired enzymatic activity and dysregulated expression of HsFEN1 with increased cancer incidence and metastatic disease progression (Table 1.6). *Fen1* knockout homozygosity (*Fen1*$^{-/-}$) in murine embryos resulted in early *in utero* lethality, whereas mice with *Fen1* heterozygosity (*Fen1*$^{-/+}$) mated with those with the adenomatous polyposis coli gene mutation, *Apc*^{1638N}, resulted in offspring with rapid gastrointestinal tract tumorigenesis and reduced median survival (Kucherlapati *et al*., 2002). Subsequently, DNA sequencing of the *Fen1* coding region in human tumor samples revealed a number of somatic mutations. Interestingly, some of them retain the activity of HsFEN1 and abolishing those of GEN and EXO also preserving genome stability (Zheng *et al*., 2007). Mice homozygous for one such mutation, the E160D, sustain rapid tumor development whereas both E160D homozygous and heterozygous mice were predisposed to cancer, deriving highly from the aggregation of apoptotic DNA and confined immune responses (Larsen *et al*., 2008; Zheng *et al*., 2007). In summary, due to the crucial role of HsFEN1 in DNA replication and repair processes, cancer development is predisposed by eukaryotic genome instability resulting from HsFEN1 enzymatic malfunction, whereas commonly overexpression of HsFEN1 in several tumor types promotes cancer cell proliferation likely driving tumorigenesis (Table 1.6).

Table 1.6 Functional deficiency of HsFEN1 in tumoragenesis

Cancer syndrome	Experimental methods	Findings	Reference(s)
Brain cancer	cDNA microarrays, RT-PCR & Western blotting	↑* Protein & mRNA expression	(Krause *et al*., 2005; Nikolova *et al*., 2009)
Breast cancer	Cancer profiling array & Immunohictochemistry	↑ mRNA expression	(Singh *et al*., 2008)
Colon cancer	Cancer profiling array & Immunohictochemistry	↑ mRNA expression	(Singh *et al*., 2008)
Gastric cancer	cDNA microarrays, RT-PCR, Cancer profiling array & Immunohictochemistry	↑ Protein & mRNA expression	(Kim *et al*., 2005; Singh *et al*., 2008)
Kidney cancer	Cancer profiling array & Immunohictochemistry	↑ mRNA expression	(Singh *et al*., 2008)
Lung cancer	Western blot, Cancer profiling array & Immunohictochemistry	↑ Protein & mRNA expression	(Nikolova *et al*., 2009; Sato *et al*., 2003; Singh *et al*., 2008)
Ovarian cancer	Cancer profiling array & Immunohictochemistry	↑ mRNA expression	(Singh *et al*., 2008)
Pancreatic cancer	cDNA microarrays	↑ mRNA expression	(Iacobuzio-Donahue *et al*., 2003)
Prostate cancer	cDNA microarrays, RT-PCR & Immunohistochemistry	↑ Protein & mRNA expression	(LaTulippe *et al*., 2002)
Testicular cancer	Western blot	↑ Protein expression	(Nikolova *et al*., 2009)
Uterine cancer	Cancer profiling array & Immunohictochemistry	↑ mRNA expression	(Singh *et al*., 2008)

*↑ indicates overexpression of either mRNA or protein levels.

Using transcription profiling analyses in nonrecurrent, primary and metastatic prostate carcinomas, LaTulippe *et al*. (2002), identified upregulated expression of *FEN1* in tumor tissues among hundreds of higher or less expressed genes involved. Increased expression of *FEN1* was also detected at both mRNA and protein levels in lung cancer cell lines, including small and non-small (Nikolova *et al*., 2009; Sato *et al*., 2003; Singh *et al*., 2008). Western blotting, performed also by Nikolova *et al*. (2009), showed an increased expression of HsFEN1 in testicular and brain tumors, including glioblastoma multiforme and astrocytoma tissue samples. Consistently, cancer profiling arrays and immunohistochemical analyses revealed overexpression of *FEN1* in the vast majority of breast cancer tissue samples, compared to normal tissues, and other cancer syndromes such as colon, gastric, kidney, lung, ovarian and uterine (Singh *et al*., 2008). Further, in their drive to elucidate the mechanism of *FEN1* dysregulated transcription and based on previous reports regarding the implication of reduced methylation in several cancers, they performed DNA methylation analysis showing

hypomethylation of the CpG promoter regions of *FEN1*, largely explaining its elevated expression levels.

1.6.2 Chemotherapeutic potential of selective HsFEN1 inhibition

In the light of the above, and given its crucial role in the maintenance of genomic fidelity, it is not surprisingly that inhibition of HsFEN1 has been substantially posited as a potential "Achilles' heel" for the development of complementary and personalised cancer therapies. Down-regulation of HsFEN1 either by small-molecule inhibitors or siRNA induced silencing was shown to increase sensitivity to DNA alkylating agents such as temozolomide (TMZ), methyl methanesulfonate (MMS), cisplatin and nitrogen mustard (He *et al*., 2016; Nikolova *et al*., 2009; Ward *et al*., 2017; Xie *et al*., 2016), whilst it significantly attenuated proliferation of cancer cells. Similarly, recent studies, have shown that the role HsFEN1 inhibition as a sensitizer is not restricted to alkylating agents, but it also increases sensitivity to ionizing radiation and different classes of chemo drugs (Ward *et al*., 2017). These include trastuzumab, an anti-tumor antibiotic, and olaparib, a poly (ADP-ribose) polymerase inhibitor (PARP) (Ward *et al*., 2017; Zeng et al., 2017) These data suggest that HsFEN1 inhibition could possibly be used in combination with the conventional chemotherapeutic drugs, lowering the minimum effective concentrations and minimising potentially the extensive DNA damage caused in response to the toxic effect of the current drugs.

Since then, experimental validation studies of targeted inhibitors are continuously expanding focusing on enduring chemotherapeutic resistance or blocking synthetic lethal interaction networks. Van Pel (2013) identified *FEN1* as the synthetic lethal interacting partner of *CDC4* and *MRE11A,* which are recurrently mutated in colorectal and other cancer types. Using either siRNA-induced silencing of *Fen1* or inhibition with small molecular weight molecules in *CDC4*$^{-/-}$ and *MRE11A*$^{-/-}$ HCT-116 human colorectal cancer cells, they found inhibition of tumor cell proliferation possibly deriving from extensive internal DNA damage.

Using high-throughput screening procedures, Bennani and co-workers identified two distinct classes of HsFEN1 inhibitors. The series of 2,4-diketobutyric acid inhibitors showed decreased potency and selectivity over the homologous endonuclease XPG, compared to the *N*-hydroxyurea inhibitors (Table 1.7; Tumey *et al*., 2005; Tumey *et al*., 2004). The second N-hydroxyurea class of molecules exhibited a more selective inhibitory activity with some cases of chemically modified molecules via alkylation with an additional benzyl group, possessing 240-fold selectivity over XPG. Even though in most cases the proposed N-hydroxyurea inhibitors overcome the XPG homology, about 10 years later, Exell *et al*. (2016) reports the simultaneous inhibition of HsEXO1 towards HsFEN1, through fluorescence anisotropy and FRET binding assay. And since the resulting phenotypes in response to HsFEN1 and/or HsEXO1 inhibition could not be distinguished, it is possible that cellular studies may be misinterpreted if simultaneous HsEXO1 inhibition is not considered.

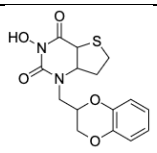
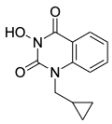
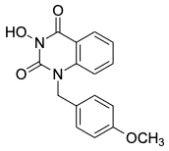
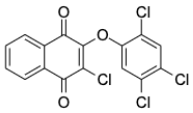
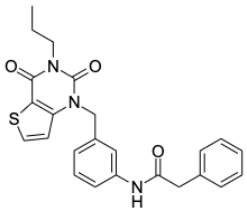
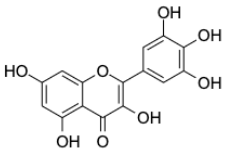
The therapeutic potential of three N-hydroxyurea inhibitors was validated further by Ward, McHugh and Durant (2017). This study focused on whether increased sensitivity to treatment in response to HsFEN1 inhibition by the *N*-hydroxyurea inhibitors is tissue and/or genetic background dependent. About 280 human cancer cell lines, covering a wide range of cancer types, were screened using a high-throughput cell-killing assay. A number of hypermutable colorectal and gastric cancer cells, a condition known as microsatellite instability (MSI), showed an increased sensitivity to the *N*-hydroxyurea inhibitors compared to Olaparib (Lynparza; AstraZeneca, UK) and other DNA damage response inhibitors. These data suggested that HsFEN1 inhibition for cancer therapy is, indeed, a possible approach. The efficacy of different DNA damage response inhibitors in different cancer types, varies depending on the genetic defects.

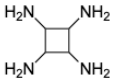
Although for years N-hydroxyurea inhibitors almost monopolised the interest for HsFEN1 inhibition, a few studies later on revealed distinct classes of HsFEN1 inhibitors. *In silico* screening performed by Panda *et al,* (2009), revealed NSC-281680 as a low molecular weight compound capable of interacting with Asp181 in the active site of HsFEN1 (Table 1.7). Its inhibitory potency was validated on its ability to block the LP-BER pathway. Indeed, NSC-281680 enhanced the cytotoxic effect of TMZ in HCT-116 human colon cancer cells and decreased the concentration of TMZ, required for half-maximal responses. Given that HsFEN1 is a primarily expressed protein in DNA repair pathways, the effect of HsFEN1 inhibition in noncancerous cells needs to be further examined. It is suspected that, following HsFEN1 inhibition, healthy cells will be also targeted by TMZ. This is a constant challenge in DNA repair pathways inhibition for the development of cancer therapies.

Similarly, two more compounds have been identified *in vitro* with inhibitory effects over HsFEN1 in DLD-1, colon cancer cells and MCF7, breast cancer cells (Deshmukh *et al*., 2017; He *et al*., 2016) (Table 1.7). Further studies determining the clinical potential of these inhibitors are needed, whereas it is uncertain whether the reported inhibition rates have been misled by a simultaneous HsEXO1 inhibition.

Although the chemotherapeutic potential of HsFEN1 inhibition has been largely validated *in vitro*, the identification of small molecule inhibitors that selectively target HsFEN1 in cancerous cells is limited. The genetic complexity of cancer itself in conjunction with the deficiency of current computational screening procedures confine the coherent design of mechanism-based inhibitors. In the lights of these, the importance for a comprehensive dissection of the molecular and biochemical differences between normal and malignant cells as well as HsFEN1 catalytic mechanism is underlined.

Table 1.7 Structures of published HsFEN1 inhibitors to date[1]

	Inhibitor	IC_{50}[2] (μM)	Reference(s)
1	 (Compound 1)	0.0464	
2	 (Compound 2)	0.0301	(Exell *et al*., 2016; Tumey *et al*., 2005)
3	 (Compound 4)	0.0169	
4	 (JFD00950)	5.5	(Deshmukh *et al*., 2017)
5	 (SC13)	0.0042	(He *et al*., 2016)
6	 (Myricetin)	0.7-12	(Ma *et al*., 2019; van Pel *et al*., 2013)
7	 (NSC281679)	29.3	(Panda *et al*., 2009)

8	 (NSC337807)	16.8
9	 (NSC281680)	1.2

[1]Structural formulas were rendered using ChemDraw Prime 16.0, whilst compound names refer to those used in the respective published papers.

[2]IC_{50} values refer to the half-maximal responses determined *in vitro* using a FRET-based assay.

1.7 Project aims

This project focuses on the identification of HsFEN1 inhibitors for the eventual development of novel anti-cancer monotherapies and/or combination therapies with the current conventional chemotherapeutic agents. Given the crucial roles of HsFEN1 in cellular DNA replication and repair processes, and its overexpression in a plethora of human cancers, which has been reported extensively, modulation of DNA repair pathways through HsFEN1 inhibition is a promising course for improving the conventional cancer therapeutic approaches. Especially, since HsFEN1 inhibition was shown to confer resistance to a range of alkylating agents in different cancer types.

1. **Development of a robust and reproducible protocol for production and purification of highly purified full-length WT and truncated D179K HsFEN1 proteins.** Aiming at the production of HsFEN1 proteins suitable for further biochemical and structural analyses with respects to the wider aims of this project, HsFEN1 were overexpressed, extracted and purified on a three-step process, combining affinity, ion-exchange and size exclusion chromatography.
2. **Development of an effective *in silico* screening strategy with strong predictive power for the identification of potent HsFEN1 inhibitors at an early stage.** Aiming at the identification of novel HsFEN1 inhibitors, which often requires screening of thousands or even millions of potential small-molecule inhibitors, an *in silico* screening strategy was developed allowing rapid and cost-effective identification of the potential HsFEN1 inhibitors at an early pre-clinical stage.
3. **The identification of HsFEN1 inhibitors, following a HTS approach and a pilot hit expansion (HE) study,** through an internal collaboration with Dr Gian Marco Ghiandoni and Professor Valerie J. Gillet (Information School, University of Sheffield).
4. **Crystallisation of HsFEN1-DNA complexes, providing mechanistic insights into HsFEN1 catalysis.** Understanding the catalytic mechanism of HsFEN1 and the interactions mediating threading and processing of the single-stranded 5′ flap, could potentially facilitate future experiments on designing and developing novel target-specific HsFEN1 inhibitors. Despite the improved understanding on the DNA threading mechanism of HsFEN1 over the last years,

questions on the specific base pairing/unpairing accommodating DNA incision and cleavage, as well as the potential interactions of the full-length C-terminus, remain unanswered.

5. **Evaluation of the *in vitro* haemolysis and cytotoxicity induced, not only by the identified HsFEN1 inhibitors, but also inhibitors of microbial FEN1 from parallel FEN1 research projects.** Inhibitors of HsFEN1 and inhibitors of microbial FEN1 enzymes, identified by DeFENition Ltd, were evaluated for their haemolytic and cytotoxic effects, aiming to determine their suitability for future drug design and development experiments. DeFENition Ltd, focuses on the development of a novel class of antibiotics, whose inhibition efficiencies over a range of microbial FEN1 enzymes offer a unique potential for treatment of microbial infections and overcoming antibiotic-resistant bacterial infections. Grounds supporting the potential of FEN1 inhibitors as anti-microbial drugs will be discussed in Chapter 6 (Section 6.1.2).

Chapter 2 - Materials and Methods

2.1 General chemicals, reagents and buffers

Unless stated otherwise, general laboratory chemicals were purchased mainly from ThermoFischer Scientific and Sigma-Aldrich.

Aqueous buffer solutions were prepared in 18 MΩ.cm deionised water (Milli-Q®) according to the standard laboratory protocols (Gorrec, 2009; Sambrook et al., 1989). All buffers were autoclaved and/or filtered sterilised as appropriate using Thermo Fischer Scientific filter units of 0.45 μm pore size.

2.2 Bacterial strains

For long-term preservation recombinant plasmids were stored in XL1-Blue and DH5α Escherichia coli (*E. coli*) cells at -80°C. Recombinant proteins were expressed in BL21(DE3) *E. coli* cells (Bullock et al., 1987; Grant et al., 1990; Studier and Moffatt, 1986). All transformations where performed in calcium chloride ($CaCl_2$) treated competent cells (Bergmans et al., 1981). Details of the genotypes and characteristics of *E. coli* host strains used predominantly in this study can be found in Table 2.1.

Table 2.1 Genotypes of *E. coli* strains used for plasmid storage and protein expression.

Strain	Genotype	Reference
DH5α	*F^- endA1 glnV44 thi-1 recA1 relA1 gyrA96 deoR nupG purB20 ϕ80dlacZΔM15 Δ (lacZYA-argF)U169, hsdR17 ($r_K^-m_K^+$), λ^-*	Invitrogen (UK)
XL1-Blue	*recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac [F´ proAB $lacI^qZ\Delta M15$ Tn10 (Tet^r)]*	Stratagene (UK)
BL21(DE3)	*B F^- ompT gal dcm lon $hsdS_B(r_B^-m_B^-)$ λ(DE3 [lacI lacUV5-T7 gene 1 ind1 sam7 nin5]) $[malB^+]_{K-12}(\lambda^S)$*	Pharmacia (UK)

2.3 Plasmids

The pET-21a(+) plasmid vector, purchased by Novagen, was used for overexpression of all HsFEN1 proteins (Mierendorf et al., 1998). Plasmid DNA was isolated and purified using Monarch® Nucleic Acid Miniprep and Purification kits from New England Biolabs. Long-term storage of plasmid DNA was performed in 6 mM Tris pH 8, 6 mM NaCl and 0.2 mM $EDTANa_2$ at -20°C. Vector map and gene insert sequences are provided in Supplementary Information (Figures SI.1;SI.2; SI2.1).

2.4 Transformation and preparation of *E. coli* culture glycerol stocks

The pET-21a(+) plasmid vectors with the genes encoding the full-length WT HsFEN1 and the truncated D179K HsFEN1 proteins were transformed into *E. coli* BL21(DE3) strain, suitable for subsequent large-scale protein expression by auto-induction. XL1-Blue and DH5α *E. coli* strains were used for long-term preservation of the recombinant plasmids. Generally, 1-5 µl (100 ng) of DNA were transformed into 100 µl of $CaCl_2$ treated competent cells. Once incubated on ice for 30 min, cells were heat shocked for 1 min at 42°C. Then, they were transferred back on ice for 2 min and after the addition of 400 µl LB media, 100 uL of that suspension were spread on LB agar plates with 100 µg/ml carbenicillin. Untransformed cells were also plated onto an LB agar plate with 100 µg/ml carbenicillin, used as negative controls. The plates were incubated overnight at 37°C and selected formed colonies were grown at 37°C to generate 20% glycerol stocks, typically stored at -80°C.

2.5 Expression and isolation of recombinant proteins

2.5.1 Large-scale protein expression by auto-induction

Expression of targeted proteins was carried out routinely by auto-induction according to William Studier's protocol for T7 promoter systems (Studier, 2005; Studier, 2014). All proteins were expressed using pET21a(+) derivative in *E. coli* strain BL21(DE3). Bacterial growth was initially carried out overnight in 3 ml of non-inducing MDG media (2 mM $MgSO_4$, 0.2x trace metals, 0.5% glucose, 18.8 mM sodium aspartate, 25 mM Na_2HPO_4, 25 mM KH_2PO_4, 50 mM NH_4Cl and 5 mM Na_2SO_4) and 100 µg/ml carbenicillin at 37°C. Protein expression was then induced by allolactose in 2 L sterile baffled flasks containing 400 ml of 2YT or Super broth auto-inducing media (FORMEDIUM, UK), 100 µg/ml ampicillin and 0.01% (v/v) antifoam 204 (Sigma-Aldrich, UK) at 22°C and 250 rpm.

2.5.2 Following bacterial growth

Bacterial cell density was measured spectrophotometrically at 600 nm ($A_{600\ nm}$) using Eppendorf BioSpectrometer® (Eppendorf, UK). Samples from the bacterial cultures were taken regularly, generally every 90 min intervals, and diluted appropriately to obtain $A_{600\ nm}$ reading within the instrument's linear range.

2.5.3 Harvesting of *E. coli* cells

Bacterial cultures were harvested by centrifugation using the Avanti J-26XP centrifuge (Beckman Coulter, UK). With the Beckman Coulter JLA-16.250 fixed angle rotor, *E. coli* cells were spun down at 21,600 × *g* for 20 min and at an operating temperature of 4°C. In all cases the supernatant was discarded and the wet cell pastes were weighed and stored at -80°C.

2.5.4 Cell lysis

Frozen cells were thawed and resuspended in 50 mM Tris/HCl pH 8, 120 mM NaCl, 5% (v/v) glycerol and 2 mM EDTA pH 8 at a volume of 10 ml per gram of cells. Lysozyme (Sigma-Aldrich, UK) was then added to 400 μg/ml final concentration, from a freshly prepared 2 mg/mL stock. The suspension was incubated at 4°C with gentle agitation for 20 min. A 23 mg/mL stock phenylmethanesulfonyl fluoride (PMSF) was also prepared in absolute ethanol (stock at ~ 23 mg/mL), which was added to the resuspension to give 23 μg/ml final concentration. The solution was incubated at 4°C with gentle agitation for 40-60 min, until the suspension turned viscous.

Sodium deoxycholate was added to 0.5 mg/mL final concentration, which was followed by the addition of dithiothreitol (DTT; Sigma-Aldrich, UK) at 2 mM final concentration. Incubation continued for another 20 min under the same conditions. The solution was sonicated at maximum amplitude for 30 s bursts, using Soniprep 150 Plus sonicator (MSE) and a tip diameter probe of 9.5 mm. Typically, 3-5 bursts with 1 min cooling intervals were performed to reduce viscosity. Insoluble material such as cellular debris and unbroken bacteria were removed at a final stage by centrifugation of the sonicated solution at 43,600 × *g* and 4°C for 30 min. Purification was monitored by SDS-PAGE.

2.5.5 Poly(ethyleneimine) precipitation

Removal of nucleic acids was performed using the minimum amount of 5% poly(ethyleneimine) (PEI; Sigma-Aldrich) stock solution adjusted to ~pH 7.9 with concentrated HCl (Cordes et al., 1990). PEI was added to the total lysate at 4°C while stirring for 15 min, after the addition of 500 mM ammonium sulphate. The precipitated nucleic acids were removed by centrifugation at 43,600 × *g* and 4°C for 15 min.

2.5.6 Ammonium sulphate precipitation

Soluble proteins, including HsFEN1 and variants, were isolated from PEI by $(NH_4)_2SO_4$ precipitation. The relevant amount of solid ammonium sulphate required for a theoretical final concentration of 4 M was added slowly to prevent foaming of the solution, which can cause protein denaturation. In practise, this will be diluted down to approximately 3 M. Once the ammonium sulphate was fully dissolved, the precipitated solution was centrifuged at 43,600 × *g* and 4°C for 20 min. The supernatant was discarded and the pellet was resuspended in the requisite buffer, typically using 5 ml of buffer per g of original cell paste weight.

2.6 Dialysis

Before proceeding to the subsequent steps of protein purification process, ammonium sulphate was removed by dialysis. An 8,000 molecular weight cut off (MWCO) permeable membrane, purchased by Spectrum Labs, was used. The membrane was prepared by boiling in 200 ml of 1 mM $EDTANa_2$ and 0.25 M of sodium hydrogen carbonate for the removal of preservative

agents, usually sodium azide and glycerine. The membrane was then boiled solely in 200 ml of 1 mM $EDTANa_2$ and washed twice with 1 mM $EDTANa_2$ to chelate any heavy metals that are likely to be presented. The membrane was allowed to cool in a minimum amount of 1 mM $EDTANa_2$ at 4°C. Protein solution was finally loaded in the dialysis bag and dialysed overnight against 1 L of the requisite low salt purification buffer with two buffer changes. The procedure was performed at 4°C while stirring.

2.7 Protein purification

Following expression of FEN1 proteins, the enzymes were purified in a 3-step process including ion exchange, affinity and size exclusion chromatography. Purification was carried out at 4°C throughout the process.

2.7.1 Purification of full-length WT HsFEN1

All steps were performed using an Äktaprime plus liquid chromatography system. The dialysed protein solution (Section 2.6) was initially loaded onto a 5/20 ml HiTrap™ Heparin (GE Healthcare Life Sciences) column, equilibrated with 10 column volumes (CV) of low ionic strength buffer (Buffer CA; 25 mM PO_4^{3-}, 25 mM NaCl, 2 mM EDTA, 1 mM DTT and 5% glycerol pH 7). This was performed using an XK 16/20 (GE Healthcare Life Sciences) column, which was packed with 20 ml of sulphopropyl (SP) Sepharose fast flow resin (GE Healthcare Life Sciences) according to the manufacturer's protocol. Unbound proteins or proteins with low net charge were washed from the column with the low ionic strength buffer, whereas strongly retained proteins were eluted with a linear gradient of 0-100% Buffer CB (Buffer CB; 1 M NaCl in Buffer CA) and 20 CV gradient volume. Fractions of 10 ml were collected at a constant flow rate of 5 ml min^{-1}. Purity was estimated by analysis on 10% SDS-PAGE gels. Fractions determined to contain the full-length WT HsFEN1 at highest purity were pooled and dialysed for further processing.

The dialysed protein pool was loaded onto a prepacked 5 ml HiTrap™ Q (GE Healthcare Life Sciences) coupled to a 5 ml HiTrap™ sulfopropyl (SP) Sepharose column (GE Healthcare Life Sciences), equilibrated with low ionic strength buffer (Buffer AA; 25 mM Tris, 25 mM NaCl, 2 mM $EDTANa_2$, 1 mM DTT and 5% glycerol pH 8). Unbound proteins or proteins with low net charge were washed off with the low ionic strength buffer, whereas the full-length WT HsFEN1 and strongly retained contaminants were eluted from heparin column with a linear gradient of 0-100% Buffer AB (Buffer AB; 1 M NaCl in Buffer AA) and 100 ml gradient volume. Fractions of 5 ml were collected at a constant flow rate of 2.5 ml min^{-1} and analysed for their purity by SDS-PAGE (Section 2.7.1.1). Fractions determined to contain the full-length WT HsFEN1 in high purity were pooled for a final purification step (Section 2.6.2).

2.7.2 Purification of truncated D179K HsFEN1

All steps were performed on Àktaprime plus. The dialysed protein solution was initially loaded onto a 5/20 ml HiTrap™ Heparin (GE Healthcare Life Sciences) column, equilibrated with low

ionic strength buffer (Buffer CA: 25 mM PO_4^{3-}, 25 mM NaCl, 2 mM $EDTANa_2$, 1 mM DTT and 5% glycerol pH 7). Unbound proteins or weakly bound proteins were washed off with the low ionic strength buffer, whereas the to be purified protein and strongly retained contaminants were eluted from Heparin column with a linear gradient of 0-100% Buffer CB (Buffer CB; 1 M NaCl in Buffer CA) and 100 ml gradient volume. Fractions of 5 ml were collected at a constant flow rate of 2.5 ml/min. Purity was estimated by SDS-PAGE (Section 2.7.1.1). Fractions determined to contain the truncated D179K HsFEN1 at high purity were pooled and dialysed for further processing.

The dialysed protein pool was loaded onto a 2nd 5 ml HiTrap™ sulfopropyl (SP) Sepharose column coupled to 5 ml HiTrap™ Q (GE Healthcare Life Sciences), equilibrated with 10 CV of low ionic strength buffer (Buffer CAb: 25 mM Tris, 25 mM NaCl, 2 mM $EDTANa_2$, 1 mM DTT and 5% glycerol pH 8). Unbound or weakly bound proteins were washed off, whereas negatively charged contaminants retained on the column were eluted with a linear gradient of 0-100% Buffer CBb (Buffer CBb; 1 M NaCl in Buffer Cab) and 20 CV gradient volume. Fractions of 5 ml were collected at a constant flow rate of 2.5 ml min^{-1}. Fractions determined to contain truncated D179K HsFEN1 at high purity were pooled and underwent gel filtration (Section 2.6.3.).

2.7.3 Size-exclusion chromatography

Pooled fractions were concentrated to a maximum of 500 µl and loaded on 0.1 x 30 cm Superdex™ 200 HR (separation range 10-60 kDa, GE Healthcare Life Sciences). A buffer consisting of 20 mM Tris, 150 mM NaCl and 1 mM DTT at pH 8 was used for both equilibration and elution with 20 CV using an Äkta fast protein liquid chromatography (FPLC) system. Fractions of 0.5 ml were collected at a constant flow rate of 1 ml/min and analysed for their purity by SDS-PAGE. Peak fractions containing the highly purified protein of interest were pooled and prepared for subsequent biochemical and structural characterisation. The temperature was maintained at 4°C throughout the purification process.

2.8 Analytical techniques

2.8.1 Qualitative analysis

2.8.1.1 Agarose gel electrophoresis

DNA electrophoresis was performed mainly for qualitative purposes using 1% (w/v) agarose gels. 1% (w/v) agarose gels were prepared by melting powdered agarose in 1x TAE buffer (40 mM Tris pH 8, 20 mM acetic acid, 1 mM EDTA) supplemented with 0.1 µg/ml ethidium bromide and allowed to set at room temperature. The set gel was then transferred in the electrophoresis tank and DNA samples, prepared at 1:1 volume ratio with Quick-Load® purple DNA loading dye (New England BioLabs, UK), were loaded. The gel was run at 7 V/cm in 1x TAE buffer until full separation of the 1 kb DNA ladder (New England BioLabs, UK) was achieved (typically around 30 min). Gels were finally visualised under UV light and imaged using Gel Doc™ EZ imaging system (Bio-Rad).

2.8.1.2 SDS-PAGE electrophoresis

Protein expression and purity were routinely examined in denaturing SDS-PAGE electrophoresis. Protein analysis was performed in 10% (37.5:1 acrylamide:bisacrylamide, National diagnostics) polyacrylamide gel, prepared in 100 mM Tris/bicine pH 8.3, 0.1% SDS 0.5 mg/mL ammonium persulphate (APS) 0.15% (v/v) and N,N,N',N'-tetramethylethylenediamine (TEMED). Gels were 86 x 67 x 1 mm with a large-pore stacking gel on the top (120 mM Tris/HCl pH 6.9, 10% acrylamide 30:1 acrylamide:bisacrylamide 0.1% SDS 0.5 mg/mL APS 0.003% (v/v) TEMED).

Samples to be analysed were incubated at 95°C for 5 min with 2x loading dye (0.2% SDS, 2.5% glycerol, 75mM Tris/HCl pH 6.8, 0.5% bromophenol blue, 5 mM $EDTANa_2$ and 1 mM DTT) in 1:1 (v/v) ratio. All samples were run at a constant current of 35 mA per gel, in 50 mM Tris/bicine pH 8.3 and 0.1% SDS buffer, alongside with the prestained Precision Protein Plus™ ladder (10-250 kD; Bio-Rad, US). Visualisation of protein separation pattern was carried out using an anionic protein dye, Coomassie Brilliant Blue R-250. Proteins were fixed in an acidic solution containing 40% methanol/ 10% acetic acid and 2 mg/mL of the dye for 30-60 min. The excess of the dye was, then, removed and the gels were destained in 20% methanol and 10% acetic acid, until the background was fully destained and protein bands became visible. Imaging of the gels was carried out using Gel Doc™ EZ imaging system (Bio-Rad).

2.8.1.3 Zymogram PAGE electrophoresis

Also known as zymography (Vandooren et al., 2013). In DNA substrate PAGE electrophoresis, FENs and any contaminating hydrolytic enzymes are detected due to their ability to degrade a DNA substrate that has been copolymerised with acrylamide in the resolving SDS-PAGE gel. Sample preparation, gels preparation and electrophoresis were carried out as described standard SDS-PAGE electrophoresis (Section 2.7.1.1.). The only difference is the addition of 800 µg Type XIV DNA (Sigma-Aldrich), previously prepared and stored according to the manufacturer's protocol, in the resolving gel. After sample separation, the gels were washed with TBG buffer (100 mM Tris/HCl pH 8, 100 mM bicine and 10%(v/v) glycerol) thrice. Each wash was accommodated by a 15 min incubation on the same buffer, to allow protein refolding *in situ* to recover their potential nuclease activity. Proteins were allowed to react by an overnight incubation in a buffer containing 100 mM Tris/HCl pH 8, 100 mM bicine, 50 mM NaCl, 1mM DTT, 10 mM $MgCl_2$, 100 mM KCl and 10%(v/v) glycerol. A final incubation in 20 ml TBG and 0.5 µg/ml ethidium bromide (Fisher Scientific) was performed for 20 min. Imaging of the gels was carried out using Gel Doc™ EZ imaging system (Bio-Rad).

2.8.1.4 Mass spectrometry

Proteins determined to be ~90% pure by SDS-PAGE, were analysed for their molecular weight using Bruker Reflex III mass spectrometer by the Facility of Mass Spectrometry (Department of Chemistry, University of Sheffield).

2.8.2 Quantitative analysis

2.8.2.1 Spectrophotometrically

Protein quantification was performed using NanoDrop ND-100 (Thermo Fischer Scientific). The instrument was initially blanked with a reference solution, which was exactly the same with the solution that the protein was suspended in in the sample. About 2 μl sample were pipetted onto the instrument's lower pedestal, to ensure accuracy and reproducibility. Protein's absorbance was measured at 280 nm ($A_{280\ nm}$) in triplicate, based on the total concentration of aromatic amino acids. These include tryptophan (Trp, W) and tyrosine (Tyr, Y), but not phenylalanine which absorbs at around 240-265 nm. Protein concentration was calculated from the mean A_{280} value and its molar extinction coefficient ($M^{-1}cm^{-1}$), assuming that all cysteines do not form disulfide bonds due to the presence of DTT.

2.8.2.2 Bradford assay

Protein quantification was also performed using Bradford protein assay (Bradford, 1976), due to the ability of Coomassie Brilliant Blue G-250 to bind proteins under acidic conditions. The dye binds to proteins primarily through aromatic (phenylalanine, tyrosine and tryptophan) and basic (arginine, lysine, histidine) amino acids. When the dye binds to proteins, it is converted from its protonated cationic form with a maximum absorbance (A_{max}) at 465 nm to its unprotonated anionic form with A_{max}=595 nm. This is accompanied by a red to blue colour change. Protein assay dye was prepared with 5-fold dilution of a concentrated dye reagent (Bio-Rad). For quantification, 20 μl of the protein sample were added to 1 ml of the protein assay dye and A_{max} readings were taken in triplicate. Protein concentration (mg/mL) was determined by the linear equation of a standard curve with known concentrations of bovine serum albumin (BSA) on x-axis, and their absorbance on y-axis.

2.9 Protein concentration and storage

High purity proteins were concentrated to >200 μM using Vivaspin 2 columns (Sartorius, UK) of 10,000 Da MWCO for maximum recovery. Protein solution was loaded in the column, whose polyethersulfone (PES) membrane has been previously washed with the requisite buffer to remove trace amounts of preservative agents, such as glycerine and sodium azide. Concentration of proteins was performed by centrifugal diafiltration at 4°C and 3,000 x *g* and at a spinning time varied depending on the overall flow rate. This was performed by refilling the concentrator with the requisite buffer and centrifuging under the same conditions. About 5 cycles were performed for each protein to ensure sufficient buffer exchange and removal of the

initial content. The concentrated sample was collected into a recovery cap by reverse spinning and protein concentration was determined as described in Section 2.8.2.1. Finally, proteins were flash-frozen in liquid nitrogen and stored in 20 mM Tris, 150 mM NaCl and 1 mM DTT buffer at pH 8 at –80°C.

2.10 Oligonucleotide synthesis

The DNA oligonucleotides used for crystallisation were purchased by Eurofins and prepared in 20 mM Tris, 150 mM NaCl and 1 mM DTT buffer at pH 8, according to the manufacturer's oligonucleotide synthesis report (Table 2.2).

The oligonucleotide substrate used to detect endonuclease activity was designed as a double flap substrate, consisting of two individual oligonucleotides (1:1) supplied by Eurofins. These include a dual labelled oligonucleotide with a carboxyfluorescein (FAM) fluorophore attached at the 5′ overhang (dT) and a cyanine 3 (Cy3) fluorophore attached at the 5′ end (dT), and a complementary oligonucleotide to form the double-stranded region (Section 2.12; Figure 2.1). Stock oligonucleotide (Assay3/Assay2Invader; Table 2.2) were prepared in 25 mM HEPES, 0.5 mM $EDTANa_2$ and 100 mM NaCl at pH 7 and annealed at 5 µM concentration. Annealing was performed at 95°C for 5 min. The oligonucleotides were left to cool at ambient temperature for 30 min prior use and stored at -20°C.

Table 2.2 Nucleotide sequences of DNA substrates used for crystallisation and FEN1 activity assays

Oligonucleotide	Sequence[1]
5ov4	5′-AAAAGCGTACGC-3′
5ov4ex	5′-AAAAGGTGCGTACGCACC-5′
5oCv5	5′-AAAAAGCGTCACGC-5′
JT1	5′-ACCGTCC-3′
JT3	5′-ACTCTGCCTCAAGACGGT-3′
JT2	5′-TTGAGGCAGAGT-3′
JT2+1	5′-ATTGAGGCAGAGT-3′
JT2+2	5′-AATTGAGGCAGAGT-3′
JT2+2_5P	5′-P-AATTGAGGCAGAGT-3′
JT2+4	5′-AAAATTGAGGCAGAGT-3′
JT2+5	5′-AAAAATTGAGGCAGAGT-3′
Assay3	5′-Cy3-TTTTCGCTGTCTCGCTGAGT-FAM-3′
Assay2Invader	5′-ACTCAGCGAGACAGCGCCGGAACACACGCTGCGTGTGTTCCGGT-3′

[1]Red, is indicative of complementary self-annealing sequence regions within a single oligonucleotide and green, is indicative of complementary sequence regions between individual oligonucleotides. The emerging single-stranded 5′- and 5′-flaps of various lengths are shown in blue. "P" in JT2+2_5P oligonucleotide stands for the phosphate group attached to the respective 5′-flap end.

2.11 Protein crystallization by X-ray crystallography

2.11.1 Co-crystallization of truncated D179K HsFEN1 with substrate DNA

2.11.2 Screening for initial crystallization conditions

Initial crystallisation screening was carried out using the Mosquito® Crystal robot (TTP Labtech, UK) and a number of commercially available screens including PACT, JCSG+, Proplex, Morpheus, $(NH_4)_2SO_4$, Natrix and Silver Bullets (Molecular Dimensions) available in the X-ray crystallography research facilities (Department of Molecular Biology and Biotechnology, University of Sheffield). The 96 conditions presented in each of the above screens were tested using the standard Swissci 96-well 2-drop plates (Molecular Dimensions). Two drop chambers were available per reservoir well, containing about 40-50 µl of mother liquors presented on the screens. Crystallisation experiments were set up on the raised wide 2 drop wells under controlled humidity levels (~70%), using sitting-drop vapour diffusion. Crystallisation drops were typically set up using a 1:1 ratio (200 nL : 200 nL) of protein to mother liquor ratio. However, occasionally, different protein to mother liquor ratios were used. The plates were sealed with transparent clear sealing sheets (TTP Labtech) and stored under controlled stable temperature, usually 17°C, where they were checked periodically (2-days, 1-week, 2-weeks, 1-month, 2-months, etc.) for crystal formation using Leica M26 microscope.

2.11.3 Optimization of initial crystallization conditions

Initial hits were identified and optimised further in order to increase their current diffraction potential and quality. Optimisations were traditionally carried out in Swissci 96-well 2-drop plates (Molecular Dimensions). Initial optimisation approach involved varying the pH and the precipitant's concentration, around the original crystallisation condition. However, alternative optimisation methodologies involving differing protein concentrations and/or additive mixtures were also used (described further in Chapter 5). Stock solutions of the initial crystallisation conditions were always made according to the manufacturer's protocol. These included mainly a precipitant, an additive and a predefined buffer system, which were combined in fixed ratios for setting up a grid of new conditions. Preparation of these solutions was performed either manually, using Corning® 1 ml deep-well plates (CORNING), or using FORMULATOR® screen builder robot (FORMULATRIX®). Optimisation plates were set using Mosquito® Crystal robot (TTP Labtech) following the previously described methodology in section X. The plates were sealed with transparent clear sealing sheets (TTP Labtech) and stored under controlled stable temperature, usually 17°C, where they were checked periodically (2-days, 1-week, 2-weeks, 1-month, 2-months, etc.) for crystal formation using Leica M26 microscope.

2.11.4 Seeding stock preparation for optimisation of crystallisation conditions

Seeding was used for optimisation of hit crystallisation conditions, aiming to improve crystal quality. Typically, a crystal stabilizing solution analogous to the crystallisation condition the

crystal of interest primarily grew in, was prepared and aliquoted in microcentrifuge tubes containing seed beads (Aliquot 0: 50, 1: 99, 2: 99, 3: 99, 4: 99 µl). Crystals of interested were then crushed, using a glass crystal crusher probe, and transferred in solution 0 by pipetting. This was followed by vortex for 1-2 min and 1 µl of solution 0 was then transferred in solution 1. Serial dilutions were carried out following the same procedure, up to solution number 4. Seeding stocks were flash-frozen in liquid nitrogen and stored at -80°C.

2.11.5 Mounting crystals

Crystals were harvested from the crystallisation solution using cryogenic polyimide resin litholoops (Hampton Research, USA) with loop diameter range from 0.025 mm to 1 mm, depending on the size of the selected crystal. Looped crystals were cryoprotected by soaking for 5 seconds into a relevant cryo-solution and then flash-frozen in liquid nitrogen to prevent formation of ice crystals, and thus expansion, which will damage the crystals. Cryo-solutions were prepared based on the composition of the original crystallisation condition with an additional 10 - 30 % (v/v) of a cryprotectant (DMSO, glycerol, ethylene glycol and saturated paraffin oil). Often, certain conditions on crystallisation screens (e.g. Morpheus; (Gorrec, 2009) are composed of precipitant mixtures providing cryoprotection. Crystals grown in such conditions were looped and frozen directly (Pflugrath, 2004). The looped crystals were finally being attached – mounted, using the magnetic ring at one end, to the magnetic goniometer base centred on the X-ray beamline for analysis. Cryoprotection and mounting of crystals for data collections was performed by Dr Domen Zafred (Marie Skłodowska-Curie action Fellow, University of Sheffield).

2.11.6 Data collection and structure determination

The X-ray diffraction data for all crystals were collected using I03/I04-1 beamlines at the Diamond Light Source, UK. Initial data collection was performed by collecting 3 images at a 45° angle and, once the protein nature of the crystal was confirmed, a full dataset of 3600 images at 0.1° angle was collected. Auto-processing diffraction integration (indexing, integrating and scaling) was carried out using X-ray Detector Software (XDS) in Diamond's ISpyB interface (Evans, 2006; Kabsch, 2010a). Among the datasets collected from the autoprocessing pipelines (autoPROC, fast DP, xia2 DIALS), the best dataset, defined by its completeness, multiplicity and I over sigma (I/σ_I), was selected for structure determination. Structure determination was performed by molecular replacement using PHASER-MR software and refined automatically using Refmac5 (McCoy et al., 2007). The structure of truncated R192F HsFEN1 (PDB ID: 5ZOG, Resolution: 2.3Å), available currently in the literature, was used as a search model (Tsutakawa *et al*., 2011). Remaining residues, both protein and DNA, were build manually in COOT and refined again using Refmac5. Final structure validation was performed in COOT, using also MolProbity as part of it (Williams et al., 2018). Structure figures, included in the present PhD thesis, were generated in PyMol (Schrödinger, LLC).

2.12 Förster Resonance Energy Transfer (FRET) Assay

Detection of structure-specific endonuclease activity and rapid screening of potential low molecular weight inhibitors, was performed using FRET assay in black Costar® 96-well plates supplied by CORNING, UK (Heyduk et al., 1996). For FRET analysis, 0.5 µg/ml of the WT HsFEN1 and 2 µM of each test compound were incubated in 25 mM HEPES pH 7.3, 0.5 mM EDTA, 100 mM NaCl, 10 mM $MgCl_2$, 100 mM KCl, 2 mM DTT and 0.1 mg/mL acetylated BSA and room temperature for 20 min.

A double flap DNA substrate, with Cy3 positioned on the 5′ flap end and FAM on the 5′ end (Table 2.3; Figure 2.1), was added to 200 nM final concentration and the reaction rates were monitored by using Varioskan Flash (Thermo Fischer Scientific). Once the 5′ flap attached with the acceptor Cy3 is cleaved, endonuclease activity is detected as an enhanced fluorescence intensity of the donor, FAM. The fluorescent emission spectra data were collected at 30 second intervals for 20 minutes at 496/519 nm ($\lambda_{ex}/\lambda_{em}$) wavelengths. Mean values were plotted against a negative control, containing no enzyme, and a positive control, containing no inhibitor. Z' values, reflecting the assay quality, were calculated using the formula $Z'=1-[3(SD_{POS}+SD_{NEG})/(|\bar{x}_{POS}-\bar{x}_{NEG}|)]$ (Zhang et al., 1999). SD_{POS} and SD_{NEG} are the standard deviations of positive and negative controls, respectively, whereas $\bar{x}_{POS}$ and $\bar{x}_{NEG}$ are the mean values for the positive and negative controls, respectively. The minimum acceptable value of Z' for a reliable and interpretable assay is 0.5. Generally, assays with $Z'>0.5$ were considered of sufficient quality.

Table 2.3 Fluorophore excitation and emission rates

Dye	Absorbance Wavelength (nm)	Emission Wavelength (nm)
FAM	495	516
Cy3	550	570

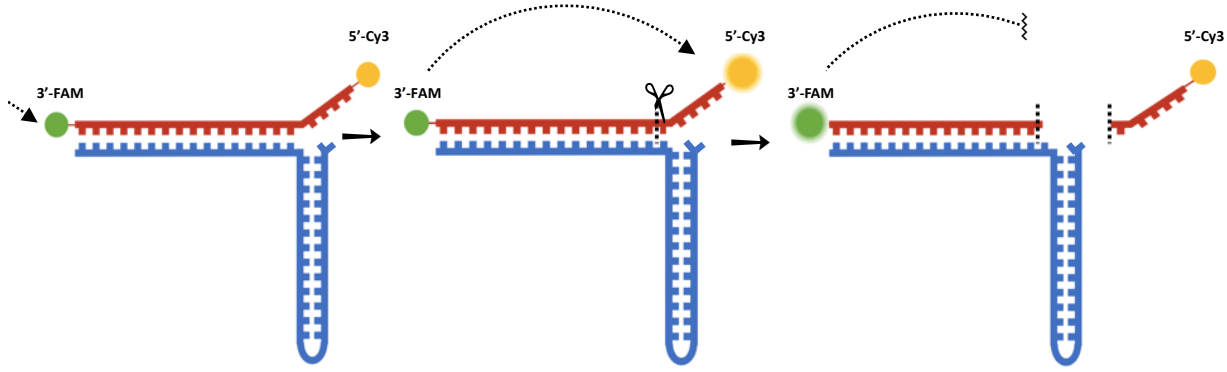


Figure 2.1 FRET principle. Dual-labelled fluorescent double-flap DNA substrate used for detection HsFEN1 structure-specific endonuclease activity. In the absence of HsFEN1, and upon FAM (Donor; Green) excitation, the close proximity of the two dyes enables efficient energy transfer to Cy3 (Acceptor; Yellow). In the presence of HsFEN1, and upon 5′-flap cleavage, energy transfer is prohibited and the respective structure-specific endonuclease activity is detected and quantified from the enhanced fluorescence intensity of the donor. Blue and red are indicative of Assay3Invader and Assay2 oligonucleotides, respectively, annealed to generate the presented double-flap DNA substrate. Sequences of Assay3Invader and Assay2 are shown in Table 2.2.

2.13 *In silico* Drug Screening

The highest resolution X-ray crystal structure of HsFEN1 (PDB ID: 5K97, Resolution: 2.1Å), available in the current literature (Tsutakawa *et al*., 2011), was simplified and optimised for use in fragment-based screening. The product substrate DNA was removed and the mutated codon for asparagine (Asn233), was reverted back to the native aspartic acid using PyMOL™ version 1.7.4 Schrödinger, LLC. HsFEN1 was defined as the target macromolecule using Autodock Vina version 1.1.2. Gasteiger charges and polar hydrogens were added to the protein, whereas the co-ordinates (x,y and z) of the gridbox defining the search area for ligand docking were adjusted to fit the active site and the conserved arch residues. The co-ordinates were defined at -73.9, -78.2, 118.2 and the gridbox at 58, 64, 77 for x, y and z, respectively, with 0.375 Å spacing. A set of 1000 compounds from Maybridge "Rule of Three" (Ro3) library (Congreve et al., 2003; Lipinski et al., 2001; Veber et al., 2002) was screened with PyRx version 0.8 through an inbuilt Autodock vina wizard. "Rule of Three" standards are summarized on Table 2.4. Default docking parameters were used and minimised energy conformations were output for the top 10 poses. Docking results were analysed using DataWarrior V4.6.1 software. The tested compounds were classified according to the predicted binding affinity (kcal/mol), root-mean-square deviation (rms) and similarity between high scoring compounds.

Table 2.4 "Rule of Three" standards applied on Maybridge Ro3 Fragment Library

Criteria[1]	
Molecular weight	≤ 300
CLogP[2]	≤ 3
Donors/Acceptors hydrogen bonds	≤ 3
Number of rotatable bonds	≤ 3
Polar surface area	≤ 60 Å^2

[1]All compounds in Maybridge Ro3 library are ≥ 95% pure and have been selected to be primarily non-reactive.

[2]Partition coefficient, ClogP, is a measure used to describe the lypophilicity of a neutral compound. It is defined as ratio of the concentrations of the solute compound in a two-compartment system, usually octanol-water, under equilibrium conditions (Mannhold et al., 2009). It is calculated in the logarithmic form using the formula, $logP_{oct/wat}=log([solute]_{octanol}/[solute]_{water})$.

2.14 Preliminary Toxicity assays

2.14.1 Haemolysis assay

Relevant concentrations of the test compounds, prepared in DMSO, were incubated in 2% Red Blood Cells (RBCs) suspension and whole horse blood for 1 hr at 37°C (1% v/v final DMSO concentration). Defibrinated horse blood was purchased from Thermo Scientific (Ref #SR0050C) and used for experiments only within the first two weeks of its arrival. For the

generation of RBCs suspension, whole blood was initially centrifuged at 1,000 rpm for 5 min and the supernatant was aspirated. RBCs were washed three times with 1x PBS and each time centrifugation was carried out at 1,000 rpm for 5 min. The relevant RBCs suspension was then generated in 1x PBS. Experiments were carried out in Falcon® U-bottom 96-well plates (Corning, US). After incubating for 1 hr at 37°C, the plates were centrifuged at 1,000 rpm for 10 min and the supernatant was transferred in Costar® flat-bottom 96-well plates (Corning) for measurement. An additional centrifugation at 1,000 rpm for 10 min was incorporated for the experiments carried out in whole horse blood to fully pellet RBCs. For whole blood experiments, measurements were performed using a 1:20 dilution of the supernatant in 1x PBS. Released haemoglobin levels were evaluated by measuring the absorbance at 540 nm using Varioskan Flash plate reader (Thermo Fischer Scientific). Each value represents the normalised mean of three technical replicates. 10% (v/v) Triton X-100 was used as an indicator of the 100% Haemolysis control and 1% (v/v) DMSO as the negative/vehicle control (n=6). *Z′* values, reflecting the assay quality, were calculated using the formula $Z'=1-[3(SD_{(+)}+SD_{(-)})/(|Mean_{(+)}-Mean_{(-)}|)]$. $SD_{(+)}$ and $SD_{(-)}$ are the standard deviations of positive and negative controls, respectively, whereas $Mean_{(+)}$ and $Mean_{(-)}$ are the mean values for the positive and negative controls, respectively. The minimum acceptable value of *Z′* for a reliable and interpretable assay is 0.5.

2.14.2 Resazurin assay

Drug-induced toxicity was estimated using Resazurin cell viability assay. Once the relevant cell suspension was generated following the procedure described in Section 2.15, 89 ul of the cell suspension (2,500 cells) were transferred into each well of the Corning™ Costar™ tissue culture-treated 96-well plates. Cells were allowed to adhere for 24 hr at 37°C, in 95% relative humidity and 5% CO_2 atmosphere. The cells were then treated with the relevant test compound (top concentration 1000 μM) for another 24 hr at 37°C, in a humidified/ 5% CO_2 atmosphere (1% v/v final DMSO concentration). Ten μl of 0.1 mg/mL filter-sterilised resazurin reagent or 7-hydroxy-3H-phenoxazin-3-one 10-oxide, prepared in 1x PBS, were added in each of the 96 wells and incubated for 2 hr at 37°C. Cell viability is proportional to the reduction of blue resazurin, whose bioconversion product is the pink fluorescent resorufin. The relative fluorescence of resorufin was measured at 540/590 nm (excitation/emission) wavelengths using Varioskan Flash plate reader (Thermo Fischer Scientific). Each value represents the normalised mean of four technical repeats. 1% (v/v) Triton X-100 was used as an indicator of the 100% cell death control and 1% (v/v) DMSO as the vehicle control (n=8). Data were analysed using nonlinear regression [Inhibitor] vs. response -Variable slope (four parameters) model in GraphPad Prism version 8, 2020. Z′ values, indicating the assay quality, were also calculated as previously described in section 2.12. The minimum acceptable value of Z′ for a reliable and interpretable assay is 0.5.

2.15 Cell culture and counting

Frozen stocks of human colorectal cancer HCT-116 (ATTC® CCL-247™) and human kidney HEK-293 cell lines (ATTC® CRL-1573™) prepared in 5% DMSO were obtained internally to use for preliminary toxicity assays. Both cell lines were cultured using phenol red-free Dulbecco's Modified Eagle medium (DMEM; Lonza) supplemented with 10% Fetal Bovine Serum (FBS) according to American Type Culture Collection's (ATCC; www.atcc.org) standard tissue culture protocols. Typically, 80-90% confluent flasks were washed twice with 1x phosphate buffer saline (PBS) and trypsinized with 1 ml of Trypsin-EDTA (Gibco; Catalog #11590626) in 37°C for 2-3 minutes. Neutralisation was then carried out with 9 ml of complete growth medium and cells were spun down by centrifugation for 5 min at 1,000 rpm. Supernatant was finally aspirated and cells were subcultured in Corning® T-75 flasks (Corning, US) at 37°C with 95% relative humidity and 5% CO_2. Subculturing was carried out 2-3 times a week at 1:5 and 1:10 ratios depending on confluency. Cultured cells were tested for mycoplasma contamination every 1-2 months depending on the on the departmental schedule of mycoplasma testing.

Prior to any cell assay, cells were counted using TC20™ automated cell counter (Bio-Rad; Catalog #145-0101) in a range of $5x10^4$ – $1x10^7$ cells/ml to ensure accuracy and reproducibility. Cell suspension was supplemented with filter-sterilized trypan blue at 1:1 volume ratio and about 10 µl were loaded into a disposable dual-chamber cell counting slide (Bio-Rad; Catalog #1450015). Cells were counted on the basis of trypan blue detection and, based on that, the original cell suspension was diluted with complete growth medium to allow plating at the desired cell densities.

Chapter 3 – Development of an *in silico* screening strategy for the identification of novel HsFEN1 inhibitors

3.1 Introduction

3.1.1 Pre-clinical drug design and development: Lead identification

The process of drug design and development through to approval, manufacture and marketing is undoubtedly a long, complex and costly process. Up until the late 1990s, pharmaceutical and biotechnological industries were at the forefront of drug discovery and development, with academia being a crucial but yet background player (Flier, 2019; Frearson and Wyatt, 2010). With the academic drug discovery consortium (aD_2C) counting more than 150 registered centres to date, joint academic-industrial drug development collaborations are certainly a popular trend and one could argue that this shift in academic expertise and external multidisciplinary collaborations may be associated with the higher average numbers of FDA approved drugs recorded lately (Mullard, 2020; Smietana *et al*., 2016). Although the biggest class of FDA approved drugs fell back in 1996, the average numbers recorded between 2017 to 2019 are the highest ever seen, indicating an increasing momentum in drug approvals (Mullard, 2020).

Successful drug development is a process often achieved after many intermediate failures. The largest study performed to date regarding the success rates in drug development, showed that only 10% of the drug candidates entering Phase I are likely to progress to further stage clinical trials and FDA approval (Smietana *et al*., 2016). This has set the requirement for more thorough pre-clinical analyses at the early stages of drug development, which have led to an expanded pre-clinical pipeline and increasing costs.

Given an extensively validated and established therapeutic target, the first step towards the development of a novel drug involves the identification of a lead molecular entity. Such entities are usually small-molecules with a desirable effect on the biological function of a target protein, whose chemical structure could serve as the starting building block for the development of compounds with improved potency. Among the available computational and experimental strategies deployed for the identification of lead molecules, high-throughput screening (HTS) and structure-based drug discovery (SBDD) are currently popular paradigms at both industrial and academic level. Each of these approaches aims to identify leads with improved ligand potencies and "druggable" characteristics in a significantly shorter period of time and lower cost than the traditional approaches.

High-throughput screening (HTS) involves the *in vitro* validation of thousands, even millions, of compounds against a target protein using well-established physical assay. Such assays can rely on different principles, and their suitability depends on the requirements set by the target protein and the intrinsic physicochemical properties of the compound library. Fundamentally, HTS relies on a single experimental replicate, for which reproducibility and robustness of HTS

has been repeatedly criticised. Despite successful drug discovery stories emerging from HTS and its early popularity, HTS is also associated with low success rates and poor data quality (Payne *et al*., 2007).

Alternatively, compound libraries can be screened computationally following the SBDD approach, also known as *in silico* screening or virtual HTS. Although successful drug discovery stories have been underlined by both HTS and SBDD, history has shown that SBDD has been successfully applied to complex therapeutic targets, where traditional approaches (e.g. quantitative structure-activity relationship – QSAR) and even HTS have struggled (Batool *et al*., 2019). The availability of a three-dimensional (3D) structure of the target protein is the starting point for successful SBDD. Thorough understanding of the intra- and intermolecular interactions driving substrate recognition and catalysis is key for the design of potent leads. Such information, including binding sites, allosteric pockets, cavities and any potential conformational changes can only be studied with accuracy through 3D protein structures. Information, which is crucial to target the appropriate protein region(s) for subsequent drug development. Upon determination of the druggable pocket, compound libraries are then screened *in silico*, in a process known as molecular docking. Molecular docking not only allows the potential identification of small-molecule drug candidates at an early stage, but it also offers the unique advantage of modelling the protein-inhibitor interactions, likely to drive the desired effect on the target's function (Anderson, 2003) . These then require some sort of physical screening in a biochemical or cell assay, as well as further validation and optimisation towards the development of an improved drug candidate, often on an iterative basis.

Lead identification at the early pre-clinical stages of any drug discovery strategy is critical. It is typically starts by the identification of individual "hit" compounds with modest binding affinities and low initial potencies in the micromolar range. Due to their low MW, optimisation and subsequent expansion of the lead molecules can readily lead to improved potencies. Both HTS and SBDD have underlined numerous success drug discovery stories over the last decades, starting with the identification of an appropriate lead (Meng *et al*., 2011). Whether these relate to industry or academia, there is no doubt that advances in HTS and SBDD have revolutionised the world of drug discovery. However, fundamental issues related to reproducibility in HTS and inaccurate predictions of the *in silico* models have been major drawbacks (Macarron *et al*., 2011; Maia *et al*., 2020). Despite the average annual numbers of FDA approved drugs being the highest seen since 1993, improvements providing consistent solutions and fulfilling the early expectations are needed.

3.1.2 Aims

Aiming at the identification of "hit" compounds with significant inhibitory activity over HsFEN1, initial experiments focused on the development of an *in silico* screening strategy with strong prediction power and the potential to facilitate downstream HTS and SBDD approaches. Such strategy, will ideally possess a strong predictive power allowing rapid and cost-effective identification of potential HsFEN1 inhibitors at an early stage. This was approached by

combining virtual and physical screening tools. The unbiased commercially available "Maybridge Rule of 3" fragment library (MaybridgeRO3) was screened against the highly purified WT HsFEN1. Using a FRET-based HTS assay, previously optimized for FENs, *in vitro* inhibition of each individual compound was determined. Experimental inhibitions were correlated to their respective binding energies, as these were calculated during multiple docking runs using AutoDock and AutoDock Vina software.

For the development of a sensitive and accurate docking model, differing docking parameters focused on the active site metals and manipulation of their partial charge. Numerous biochemical and structural studies have repeatedly supported the requirement of two active site magnesium ions for optimal nucleolytic activity of FENs, including HsFEN1 (Feng *et al*., 2004b; Garforth *et al*., 1999; Tsutakawa *et al*., 2011; Tsutakawa *et al*., 2017). Similar to other alkaline earth metals, magnesium is a positively charged ion in aqueous solution (Mg^{2+}). Typically, its presence in the enzyme catalytic site strongly influences the active site's net charge and leads to multiple coordination geometries, challenging the *in silico* prediction of inhibitor/ligand-metal interactions (Chen *et al*., 2007a; Haas and Franz, 2009). Here, following analogous approaches for the optimization of active site metal ion parameters for docking, it has been hypothesized that manipulation of the virtual environment by considering the active site metal charges can lead to a docking model with increased predictive power for the identification of HsFEN1 inhibitors (Chen *et al*., 2007a; Riccardi *et al*., 2018).

In a similar manner, the likelihood of inhibition mechanisms involving chelating of both active site metals (M_A^{2+} and M_B^{2+}), one (M_A^{2+}/M_B^{2+}) or even none was explored. In all cases, the estimated binding energies for the docking runs exploring these mechanisms were correlated to the experimentally determined inhibition of each individual compound observed, *in vitro*, in the biochemical assay. Compounds with high inhibitory potency over HsFEN1 underwent preliminary similarity analysis for the identification of conserved structural motifs, likely to be critical for HsFEN1 inhibition.

3.2 RESULTS & DISCUSSION: Screening of MaybridgeRO3 fragment library using a FRET-based HTS assay

3.2.1 Assay rationale and key features of MaybridgeRO3 fragment library

As a first step towards the development of an improved *in silico* screening strategy and the identification of novel HsFEN1 inhibitors, a core set of fragments was used for the collection of both *in silico* and experimental data, that will allow multiple parallel comparisons. MaybridgeRO3 is a chemically diverse fragment library with compounds characterized by high pharmacophore content and 100% compliance to Christopher's A. Lipinski "Rule of Three", where top scored hits are traditionally found to lie in (Figure 3.1) (Lipinski and Hopkins, 2004; Lipinski *et al*., 2001). "Rule of Three" guidelines, an extension of, and perhaps a nod to, "Rule of Five" predicts that chemical compounds with ≤3 H-bond donors/acceptors, ≤3 rotatable bonds, ≤300 Da MW, hydrophilicity (ClogP) ≤3 and a calculated PSA ≤60 $Å^2$ are more likely

to deliver optimized candidates for downstream drug development, or the so called lead-like molecules. Often, these physicochemical properties, in addition to the biological properties of a drug candidate are more likely to be effectively translated into Phase I clinical trials, and are relevant to the increased success rates of that phase (Lipinski *et al.*, 2001).

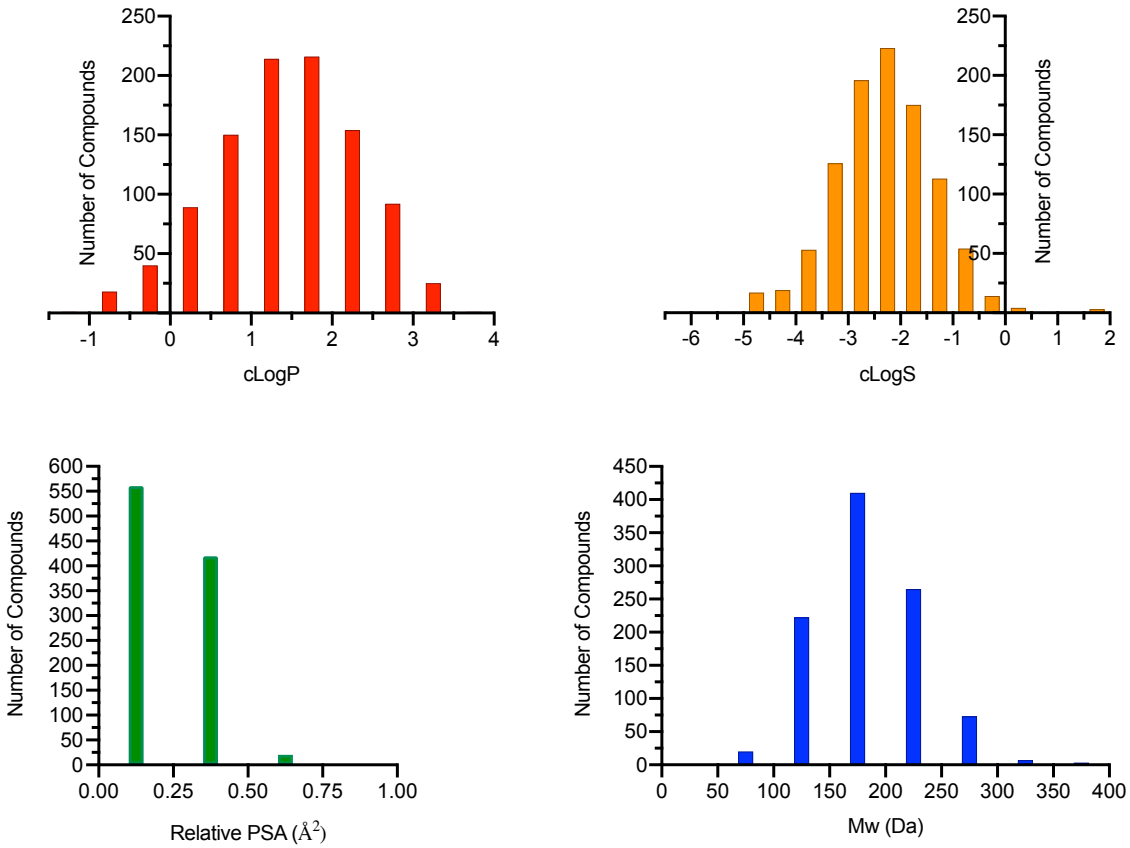


Figure 3.1 Physicochemical properties of MaybridgeRO3 fragment library. The number of corresponding chemical compounds is plotted against the (a) Molecular weight (MW), (b) calculated lypophilicity (cLogP), (c) calculated solubility (cLogS) and (d) relative polar surface area (PSA). Physicochemical parameters were calculated using DataWarrior 5.2 and plotted using GraphPad Prism 8.4.1.

In this study, 1,000 core compounds from the commercially available MaybridgeRO3 library were screened, encompassing the diversity of the entire library, which is drawn from >53,000 molecules today (Thermofisher Scientific, Catalog #R3W035). Using an optimized FRET-based HTS assay, the MaybridgeRO3 library was screened against full-length WT HsFEN1. Inhibition of HsFEN1 endonuclease activity was expressed as percentage inhibition (%), with a threshold of at least 60% inhibition for the most potent HsFEN1 inhibitors.

A FRET-based assay was originally developed for the evaluation of FEN1 activity by Dr Jing Zhang and subsequently modified for use as a HTS assay by Dr Sarbendra L. Pradhananga in Sayers' laboratory, University of Sheffield (AlMalki *et al.*, 2016). Typically, enzymatic reactions are carried out using a dual-labelled fluorescent double flap substrate, where

quenching of the donor (FAM) by the acceptor (Cy3) occurs as the two dyes are in close proximity (<10 nm) in the substrate. In the presence of a hydrolytic enzyme, such as HsFEN1, DNA cleavage, or removal of ss-flap, prevents energy transfer between the two dyes and enzymatic activity is detected as an enhanced florescence intensity of the donor.

Each reaction is also carried out under an established buffer system, containing optimum concentrations of additives and co-factors, shown to be indispensable for HsFEN1 catalytic activity (Tsutakawa *et al*., 2011). In particular, reactions were carried out in the presence of 10 mM $MgCl_2$, 100 mM KCl and 2 mM DTT, maximising the catalytic efficiency of HsFEN1 according to previous studies from both Sayers' and Tainer's laboratories (Dr Sarah L. Oates, unpublished; Tsutakawa *et al*., 2011). Although optimum pH conditions for FEN1 enzymes vary among diverse organisms, and should be taken into account when screening for the identification of inhibitors against bacterial and viral FENs, pH 7.3 used for HTS against HsFEN1 mimics the intracellular pH (pHi) of the human host. Maintaining the pH of a healthy human host, the expected drug-acceptor, which ranges between 7-7.4 without compromising the function of the enzymatic target is critical for a potentially efficient translation into the later stages of drug discovery.

3.2.2 Optimization of HsFEN1 concentration for maximum HTS performance

Although FRET-based HTS has been extensively optimized before and several studies from both Sayers' and Tainer's laboratories have set the requirements for maximum HsFEN1 performance, optimization of the protein concentration is critical (AlMalki *et al*., 2016; Exell *et al*., 2016). Prior to any inhibitor screening, it is crucial to ensure that the concentration of HsFEN1 used, produces a detectable and reproducible signal, while it maintains the cost-effectiveness of the assay. Three different concentrations of full-length WT HsFEN1 (1.18 nM, 11.8 nM and 118 nM) were examined for their respective FRET signals over a 20 min time-course and compared to a negative/no enzyme and positive controls. Full-length WT HsFEN1 produced previously by Dr Sarah L. Oates was used at 11.8 nM as a positive control, allowing direct comparisons of the catalytic activity and the observable reaction rates between two different production batches of the same protein.

Typically, any increase in FRET signal is analogous to the catalytic activity of HsFEN1 and cleavage of ss-flap. Detectable signals were observed for all three concentrations tested. The activity of full-length WT HsFEN1 at 1.18 nM was followed by a relatively slow increase in the fluorescence intensity, indicating that a concentration of 1.18 nM for HsFEN1 is much too low to provide the high sensitivity required in the subsequent HTS inhibitor assays (Figure 3.2). The same applies to the highest concentration tested (118 nM), where the relative fluorescence intensity was rapidly increased and reaction saturation occurred within the first 10 min of the assay (Figure 3.2).

At 11.8 nM the activity of the protein mirrored that of WT HsFEN1, derived from an older batch produced by Dr Sarah L. Oates (Figure 3.2). The concentration of 11.8 nM was selected as the optimal concentration of full-length WT HsFEN1 to use for HTS inhibitor assays.

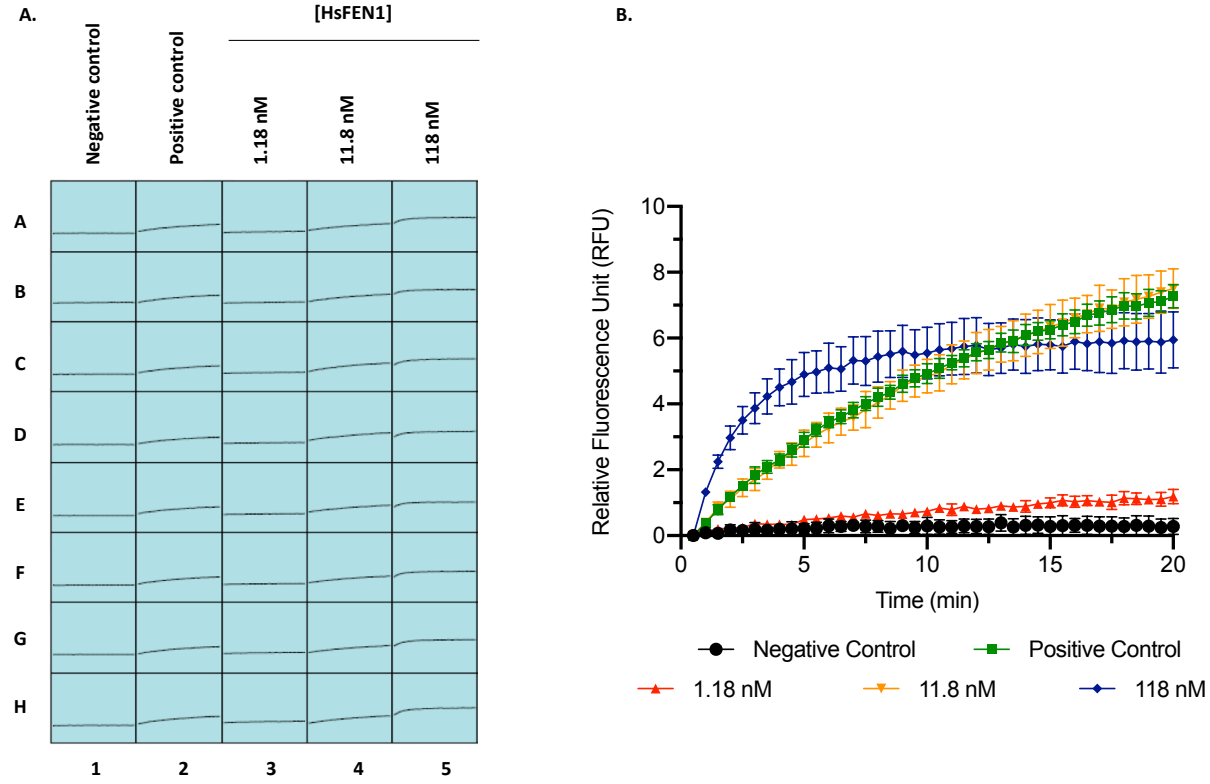


Figure 3.2 Structure-specific endonuclease activity of full-length WT HsFEN1 assessed by FRET. (a) Fluorometric report representative of the full emission spectra recorded over a period of 20 min. Columns 1-5 represent the negative/no enzyme control, positive control and protein concentrations from 1.18 nM, 11.8 nM and 118 nM, respectively. Rows A-H are representative of the 8 technical replicates. (b) Typical fluorescence emission data for full-length WT HsFEN1 binding A3Invader DNA substrate at 1.18 nM, 11.8 nM and 118 nM. Relative fluorescence was plotted against time and analyzed by non-linear regression following Michaelis-Menton equation (GraphPad Prism 8.2). Error bars represent the SD of the normalized mean, calculated from 8 technical replicates performed in a single experiment (i.e. $N = 1, n = 8, Z' = 0.75$).

The Z' value calculated for 11.8 nM protein concentration, further supported that choice. Z-factor or Z' is an established "screening window coefficient" indicative of the assay quality and suitability for subsequent large-scale screening. HTS has by nature been designed for rapid and cost-effective screening of chemical libraries. Thousands of compounds are only validated in singlets for the identification of hits, that show a desired biological effect over the protein of interest (e.g. activation/inhibition). Use of a statistical parameter that eliminates signal variability and maintains robustness and reproducibility is crucial. Z' value is calculated based on the means and standard deviations of the positive and negative controls, assuming normal distribution and a chance of >99% for the true activity of a compound lying within 3 SDs from initial estimated activity. Hence, any signal variability due to human/instrumental errors will

have an effect on Z' and assay quality. Typically, a high quality or "excellent" assay will have a Z' value between 0.5-1, with the ideal assay having a Z' value equal to 1. Table 3.1. shows the calculated Z' values for each different concentration of HsFEN1 tested. Although Z' values were found to be >0.5 in all cases, higher protein concentrations (11.8 nM and 118 nM) showed higher Z' than that seen at 1.18 nM (Table 3.1). Despite the fact that Z' was found to be 0.72 for both 11.8 nM and 118 nM concentrations, a concentration of 118 nM for HsFEN1 deemed less suitable for HTS due to rapid reaction saturation described above (Table 3.1). Therefore, the calculated Z' values support further our decision on using 11.8 nM as the optimal concentration of full-length WT HsFEN1 for subsequent HTS inhibitor assays.

Table 3.1 Calculation of Z-factor[1] (Z') and categorization of assay quality[2] for the different protein concentrations tested

Protein concentration (nM)	Mean	Standard deviation (SD)	Z'
1.18	1.2	0.22	0.51
11.8	7.5	0.6	0.72
118	5.9	0.85	0.72

[1]Z' values for the different protein concentrations were calculated using the formula $Z' = 1-[3(SD_{POS}+SD_{NEG})/(|\bar{x}_{POS}-\bar{x}_{NEG}|)]$ (Zhang *et al.*, 1999). SD_{NEG} and $\bar{x}_{NEG}$ represent the standard deviation and the mean values for the negative control, respectively. SD_{POS} and $\bar{x}_{POS}$ represent the standard deviation and the mean values for the positive control, respectively, which are calculated from the corresponding fluorescence at each protein concentration tested.
[2]Assay quality is determined as "marginal", "ideal" and "excellent" based on $Z' < 0.5$, $1 > Z' \geq 0.5$ and $= 1$, respectively.

3.2.3 FRET-based screening of MaybridgeRO3 library against HsFEN1

Following the optimized FRET-based assay protocol for HTS, 1,000 core compounds from the entire MaybridgeRO3 fragment library were evaluated for their inhibitory potencies over 11.8 nM HsFEN1. Each compound was tested at a concentration of 2 mM in a 96-well format alongside a negative/no enzyme ($n = 8$) and positive/no inhibitor controls ($n = 8$). A total of 13 assays needed to be performed to obtain a single measurement on each of the 1,000 MaybridgeRO3 compounds, given the plate layouts of stock MaybridgeRO3 plates and following the methodology described in Figure 3.3. Since one of our aims involved the development of an *in silico* screening strategy with improved prediction power, obtaining reproducible and robust experimental data to correlate with the following docking runs is critical. Screening of the 1,000 MaybridgeRO3 molecules was carried out three times over three independent experiments, resulting in triplicate measurements for each compound and bringing the total number of 96-well plates assayed performed to 39 (Supplementary Information; Figure SI.4).

Assay quality was determined by the Z', based on the means and SD values of the positive and negative controls from each individual assay. The calculated Z' and SD values of every assay are shown in Table 3.2. For the majority of the assays, the individual Z' values were maintained above 0.5 with an overall Z' equal to 0.64, indicative of an "excellent" assay according to the classification rules of Zhang, Chung and Oldenburg (1999). However, 7 out of the total 39 assays performed were found to have $Z' < 0.5$, indicative of marginal but still acceptable assay quality. These assays of compromised quality mostly appear during the 3rd replicate. Generally, the individual Z' values as well as the mean Z' (Mean $Z' = 0.44$) of the 3rd replicate were lower than the ones observed during the 1st (Mean $Z' = 0.64$) and 2nd (Mean $Z' = 0.67$) repeats (Table 3.2). This is likely due to random errors, which were expected to be minimised by the three independent HTS experiments carried out.

Table 3.2 Quality of individual FRET-based assays during HTS of MaybridgeRO3 fragment library

MaybridgeRO3 stock plate (384-well plate)	Assay number (96-well plate)	Ligands	Replicate 1		Replicate 2		Replicate 3	
			Z′	SD	*Z′*	*SD*	*Z′*	SD
1	A	1-80	0.82	2.7	0.61	9.8	0.51	11.5
	B	81-160	0.51	10.8	0.62	9.5	0.39	16.3
	C	161-240	0.84	2.5	0.79	5.2	0.51	12.1
	D	241-320	0.56	12.2	0.74	6.5	0.5	3.7
2	A	321-400	0.54	10.4	0.64	10	0.51	8.2
	B	401-480	0.75	7.1	0.71	7.9	0.31	14.9
	C	481-560	0.58	10.5	0.86	2.9	0.32	14.3
	D	561-640	0.62	8.6	0.55	8.4	0.48	9.2
3	A	641-720	0.67	7.2	0.69	7.6	0.32	7.9
	B	721-800	0.64	7.4	0.44	14.5	0.56	7.5
	C	801-880	0.5	10.7	0.66	9.4	0.29	11.6
	D	881-960	0.61	9.8	0.77	5.1	0.53	5.8
4	A	961-1000	0.7	6.8	0.68	7.2	0.43	11.5
Mean			0.64	8.2	0.67	8	0.44	10.4

[1]The overall assay quality was determined by the means of the calculated $Z'$from each of the three assay repeats. $Z' < 0.5$ and $1 > Z' \geq 0.5$, correspond to "marginal" and "excellent" assay quality, respectively.

Triple screening of the entire MaybridgeRO3 library, was carried out gradually over time in three independent experiments. Fragments from each stock plate, corresponding to four or less assays, were screened within a day. Optimum conditions, involving reagent, substrate and protein concentrations were maintained throughout the three repeats. However, positive controls seemed to have an endpoint activity lower in the 3rd repeat than in the 1st and 2nd replicates (Figure 3.3). In addition, reaction saturation occurred much faster, with the majority of the linear reaction likely to have occurred at an earlier non-captured stage (Figure 3.3). Since 3rd repeat was the last one performed, with detection rates that were less of a problem, the exact reasons for this variation were not investigated further (Section 3.2.4).

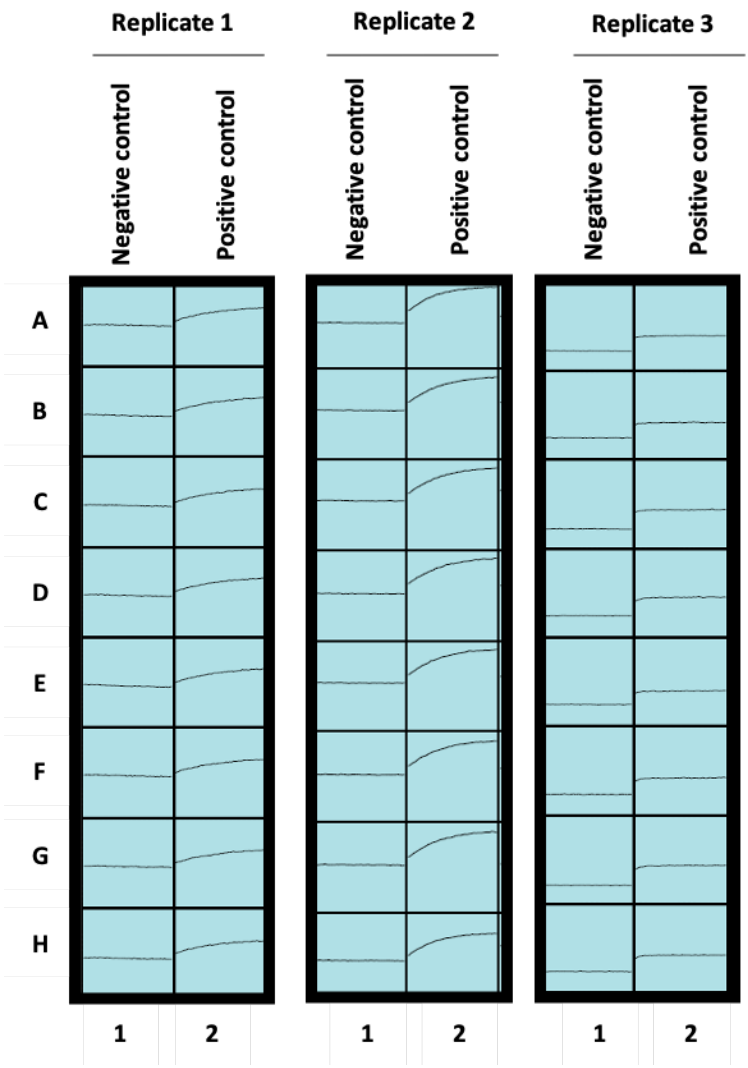


Figure 3.3 Representative examples of positive and negative controls for each of the three independent experimental repeats of MaybridgeRO3 fragment library. Columns 1 and 2, represent the negative/no enzyme and positive control, respectively, in each of the three experimental repeats. Rows A-H are representative of the 8 technical replicates performed.

Positive control caveats of the 3rd FRET-based assay repeat are likely to be related to the enzyme *per se*, substrate and/or the wider environment, such as temperature. Although it has been aforementioned that reactions were performed under the exact same conditions and concentrations, one could argue that protein and/or substrate batch to batch variations are likely to be responsible for that change in the observed activity of HsFEN1. Despite the fact that all assays were performed with the same batch of protein, validation of HsFEN1 activity from different batches resulted in remarkably similar activities (Section 3.2.2). Substrate on the other hand, is prepared as a communal reagent due to the large number of FRET assays carried out routinely in Sayers' laboratory. The protocol for substrate preparation is well-established and each batch produced is validated multiple times in the parallel FRET assays carried out against diverse FENs. Therefore, if the issue was related to a particular batch of DNA substrate, then it was going to be apparent in other FRET assays carried out in the same period. Unless the issue was related to substrate contamination, occurred during the final dilution of the

concentrated substrate stock, just before loading the substrate into the 96-well plates. Finally, similar to most chemical reactions, enzymatic reactions are also temperature-dependent. So, any variations in the temperature are likely to affect the activity of the enzyme. Typically, preparation of assay plates and fluorescence measurements were taken place in different laboratories. The temperature was normally maintained at 22°C but occasionally it rises up to 27°C. Thus, if the 3rd repeat was carried out at one of the two extremes, that might explain the observed difference between the positive controls.

Ideally, a 4th repeat under a temperature-controlled environment could potentially clear up the reasons led to compromised assay quality during the 3rd repeat. In addition, performing a complete screening of all 1,000 ligands within a day, instead of a 3-4 days period that was used so far (4 assays per day), could minimise possible variations derived from time-sensitive factors. However, the feasibility of that depends highly on the availability of the plate reader. Generally, the mean Z' values for each assay were maintained above the desired 0.5. This, in addition to the respective detection rates presented in section 3.2.4, supported replicability of the results, at least for the identified hits.

3.2.4 Identification of primary "hits" with inhibitory activity against HsFEN1

For the identification of the most potent HsFEN1 inhibitors tested, an arbitrary threshold of at least 60% inhibition was used based on the rule of 3SDs (Zhang *et al*., 1999). Active compounds with significant inhibitory effect over HsFEN1 (≥ 60%), larger than the calculated percentage of 3SDs for their respective assay, are unlikely to have occurred randomly and were shortlisted for further analysis. Despite the fact that Z' values were maintained in their majority above 0.5, different degrees of data variability were observed. However, fluorescence measurements for fragments with inhibition ≥60%, identified as "hits", were found to be more consistent compared to non-inhibitors (Figure 3.4).

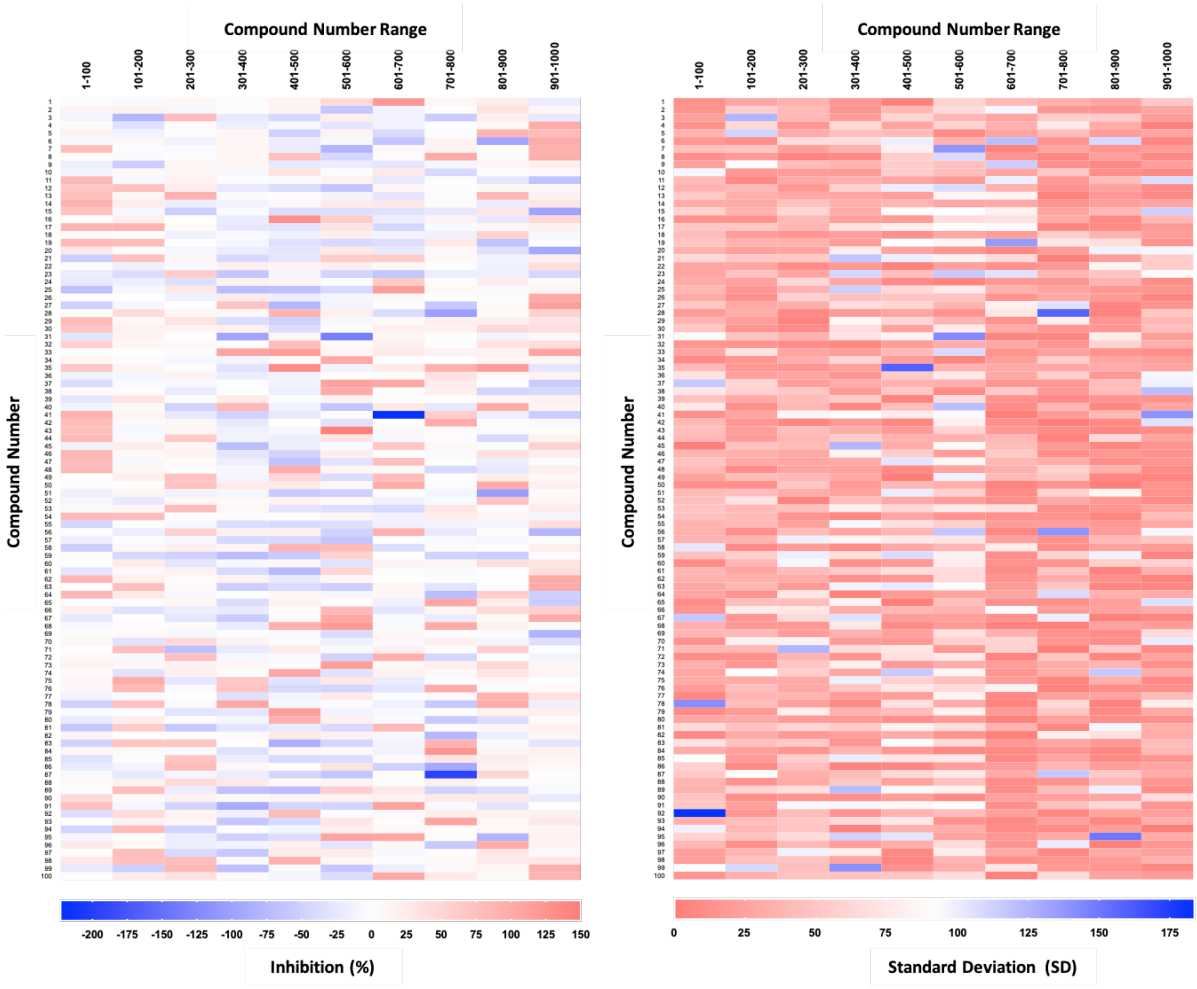


Figure 3.4 High-throughput screening of 1,000 molecules in the MaybridgeRO3 library for their inhibitory activity against HsFEN1. Heat maps are representative of the mean percentage inhibition (Left) and the standard deviation (right) for each of the 1,000 MaybridgeRO3 ligands, calculated from three independent singlet measurements (i.e. $N = 3$, $n = 3$). In both cases, each column corresponds to 100 MaybridgeRO3 ligands (1-100, 101-200, 201-300, etc.) and each row to one MaybridgeRO3 ligand starting from ligand 1 at the top left corners. Red is indicative of the desired outcome, involving high HsFEN1 inhibition (%) and low SD. Percentage inhibition was calculated based on the formula Inhibition (%) = 100-[$(RFU_{sample}-\bar{x}_{NEG})/(\bar{x}_{POS}-\bar{x}_{NEG})$]*100, where $\bar{x}_{POS}$ and $\bar{x}_{NEG}$ are the mean RFU of the positive and negative controls, respectively.

HTS has been repeatedly criticised for frequent issues related to reproducibility and data quality. Most commercially available compound libraries, such as MaybridgeRO3, have been designed for the very early stages of small-molecule hit identification with a desired biological effect over a range of structurally diverse proteins. Suitable screening methods can also rely on different principles. Therefore, there is no ideal library to be screened for a particular family of proteins, unless it has been built for that purpose, and of course, no ideal method to cover the diverse properties of the individual ligands. Results and even reproducibility capacity are highly dependent on the intrinsic chemical properties of the compounds composing a fragment library. For example, DNA intercalation, auto-fluorescence and quenching are commonly

associated with increased numbers of false-positives and/or false-negatives (Murray and Rees, 2009). These, in addition to random errors, are likely to be responsible for the observed difference in consistency between inhibitors and non-inhibitors, as well as the apparent increases in activity seen for some ligands (Figure 3.4). Potential problems associated with the intrinsic properties of the ligands will be discussed in more details in section 3.4.3.

Overall, screening of MaybridgeRO3 fragment library resulted in 127 primary hits with percentage inhibition ≥ 60%, larger than the calculated 3SDs for their respective assay (Figure 3.5). Since MaybridgeRO3 is an unbiased library composed of structurally diverse ligands, which have been randomly distributed across the four plates, an even distribution of the identified "hits" was expected (Figure 3.5). Despite the different levels of variability observed among the three repeats, compounds with significant inhibitory effect over HsFEN1 were found indeed to be distributed evenly across the four stock plates of MaybridgeRO3 fragment library (Table 3.3; Supplementary Information – Figure SI.4). The total number of the identified hits constitute the 12.7% of the entire library screened, out of which 5% were found in plate 1, 3.6% in plates 2 and 3 respectively, and an additional 0.5% in plate 4. Although the apparent percentage of hits in plate 4 seems to be much lower, when comparing the percentages of the identified hits calculated with respect to the total number of compounds presented per plate, it is found to be analogous to those seen in plates 1, 2 and 3 (Table 3.3). This is due to a lower number of fragments presented in plate 4 than in plates 1, 2 and 3.

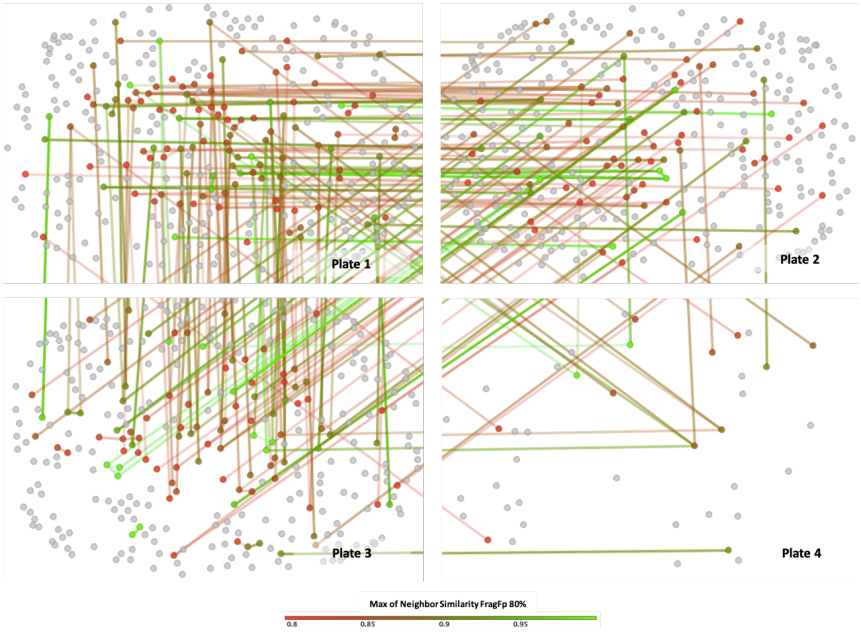


Figure 3.5 Similarity analysis and distribution of 1,000 ligands from MaybridgeRO3 fragment library across four plates. Molecules were randomly distributed across the four stock MaybridgeRO3 plates, resulting in structurally similar compounds, of different degrees, assigned in all four plates. Connecting lines are indicative of structurally similar compounds with calculated similarity ≥ 80%. The bottom spectra bar shows the level of structural similarity between two molecules, starting from red (Tanimoto score=0.8) to green (Tanimoto score=1.0). The closest the Tanimoto score of 1 is, the highest the level of similarity between two molecules. Similarity or activity cliffs analysis was performed using default settings based on a binary fingerprint descriptor (FragFp) in DataWarrior 5.2.

Table 3.3 Distribution of primary "hits" across plates according to their inhibitory effect against HsFEN1

MaybridgeRO3 stock plate	Number of "hits" ("Hit" rate, %)[1]						
	Replicate 1	Replicate 2	Replicate 3	Total[2]			
	≥ 60%	≥ 60%	≥ 60%	≥ 60%	≥ 70%	≥ 80%	≥ 90%
1	7 (2.2%)	54 (16.9%)	79 (24.7%)	50 (15.6%)	44 (13.8%)	19 (5.9%)	4 (1.25%)
2	29 (9.1%)	38 (11.9%)	66 (20.6%)	36 (11.2%)	32 (10%)	27 (8.4%)	23 (7.2%)
3	43 (13.4%)	41 (12.8%)	60 (18.8%)	36 (11.2%)	35 (10.9%)	31 (9.7%)	27 (8.4%)
4	6 (15%)	5 (12.5%)	11 (27.5%)	5 (12.5%)	0 (0%)	0 (0%)	5 (100%)
Sum	85 (8.5%)	138 (13.8%)	216 (21.6%)	127 (12.7%)	111 (11.1%)	77 (7.7%)	59 (5.9%)

[1]Hit rate indicates the percentage of "hits" identified per plate. Plates 1, 2 and 3 were composed from a total of 320 ligands, while 40 ligands constituted plate 4.
[2]Total number of "hits" and their breakdown, according to their % inhibition, is based on the calculated mean % inhibition for each individual compound.

Detection rates also varied across the three replicates. Replicate 3 appeared to be more sensitive with a total of 216 compounds showing % inhibition ≥ 60% (Table 3.3). This number corresponds to the 21.6% of the entire library, which is suspiciously high compared to 8.5% and 13.8% identified in repeats 1 and 2, respectively (Table 3.3). As aforementioned, 3rd repeat showed generally lower *Z′* values, which have likely resulted in a high number of false positives. However, since 3rd repeat was the last one performed, % inhibition and therefore, the total number of identified hits was brought down by repeats 1 and 2. In this way, false positives that appear to have % inhibition ≥ 60%, only in the 3rd repeat, were ruled out from subsequent analysis. Finally, this resulted in a shortlist of top hits, which were found to inhibit HsFEN1 more consistently across the three repeats. Subsequent similarity analysis can potentially unravel conserved structural patterns, likely to be key for HsFEN1 inhibition.

3.2.5 Preliminary analysis of HsFEN1 inhibitors identified from MaybridgeRO3 HTS screening

Shortlisted hits derived from the triplicate screening of MaybridgeRO3 fragment library, were analysed further for the identification of any conserved structural motifs, likely to be responsible for the observed inhibitory effect over HsFEN1. Structural similarity, also referred to as activity cliffs analysis was performed using DataWarrior 5.2., based on a default binary fingerprint descriptor (FragFp). About 102 out of the 127 identified hits, appeared to have one carboxyl group in their structures, suggesting an important role for HsFEN1 inhibition. These

constitute the 80.3% of the identified top hits, as well as the 58% of the total number of fragments (= 175) with a carboxyl group in their structures, presented in MaybridgeRO3 library. Among the compounds with inhibition ≥ 60%, thirteen clusters were identified in total with 6 or less structurally similar ligands each (Table 3.4).

Only compounds with a carboxyl group in their structures appeared to be clustered, suggesting a similar inhibition mechanism likely through chelation of the two active site metals. Cluster 1 represents the largest set of structurally similar compounds, consisting of six fragments. Pyridinecarboxylic acid, composed from a pyridine ring with one carboxyl group attached to it, represents the common structural core between compounds of the 1st cluster (Table 3.4). Despite the common core, the observed % inhibitions seemed to vary. The addition of extra groups on the pyridine ring, such as chlorine (Cl-, Fragment 695), fluorine (F-, Fragment 700), methyl (CH_3-, Fragment 213) and benzene (Fragment 189) led to increased % inhibitions, compared to pyridine-4-carboxylic acid (Fragment 12). Generally, larger molecules are likely to exhibit a greater inhibition potential, due to their ability to form multiple interactions with the target protein. Fragment 516 constitutes the only exemption from cluster 1, in which the addition of an ortho-fused benzene ring to pyridine led to a flat polyaromatic molecule with lower inhibition potency just above the 60% threshold (60.4%). Flat polyaromatic systems have generally restricted conformational flexibility, compared to small-molecules with branched functional groups and/or hydrocarbon chains. Therefore, their ability to form multiple interactions and, in turn, inhibit the target protein is highly dependent on the presence of additional functional groups (e.g. carboxyl group) and not the hydrocarbon rings *per se*.

Similar to pyridine-4-carboxylid acid (Fragment 12), pyrrole-2-carboxylic acid (Fragment 203) also consists from a carboxyl group attached to pyrrole ring, instead of pyridine (Table 3.4). The difference in the observed inhibition potencies among the two ligands, suggests an important role of the nitrogen's position on the aromatic ring. In pyridine-4-carboxylid acid, nitrogen is positioned opposite the carboxyl group on benzene ring, whereas in pyrrole-2-carboxylic acid is positioned next to the carboxyl group. Since the active site of HsFEN1 is rich in negatively charged residues, close-proximity of fragment's 203 nitrogen to the active site is likely to be the reason for the increased inhibition potency observed. Of course, assuming that inhibition of HsFEN1 occurs through chelation of the active site metals by the carboxyl group, so that the carboxyl group faces the active site.

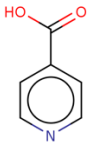
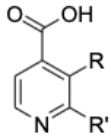
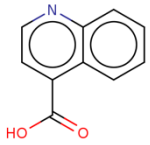
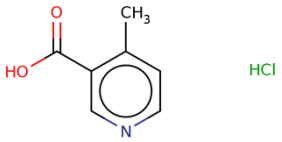
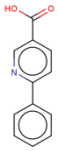
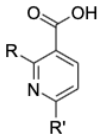
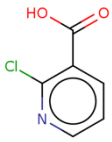
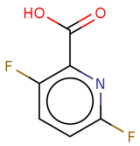
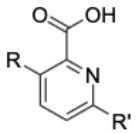
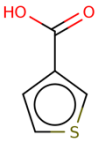
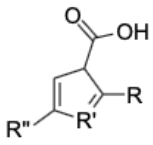
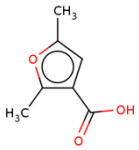
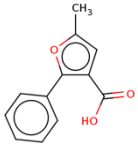
Pyryliumcarboxylic acid (Cluster 2, 3 and 9), thiophenecarboxylic acid (Cluster 2, 3, 5, 6 and 9) and benzoic acid (Clusters 7 and 10) have also been identified as alternative common structural cores (Table 3.4). Despite their overall structural similarity to pyridinecarboxylic acid and pyrrolecarboxylic acid, the addition of atoms with differing electronegativity has led to diverse inhibition potencies above the 60% threshold. Typically, the polarity or strength of a chemical bond, which is critical for the inhibition potency of a ligand, increases with more electronegative atoms. Therefore, substitution of one or more carbons (C) on the aromatic rings by more electronegative atoms, such as sulphur (S), nitrogen (N), and oxygen (O) can potentially justify the increased inhibition efficiency of some fragments. In addition, based on

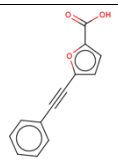
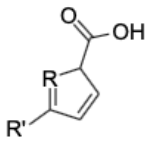
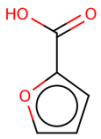
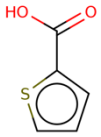
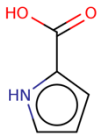
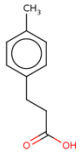
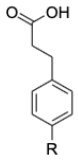
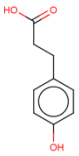
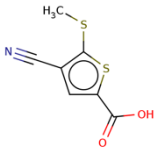
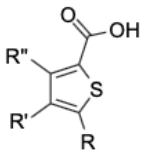
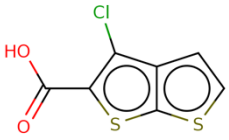
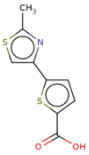
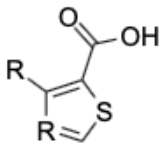
the same principle, the presence of chlorine (Cl) and/or fluorine (F) also resulted in increased inhibition potencies in the majority of cases. However, their respective position on the molecule and the overall stereochemistry is another critical parameter determining their ability to interact with the target protein. Therefore, exceptions such as fragment 171 (Cluster 6) are expected.

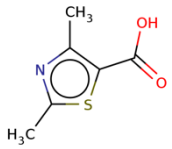
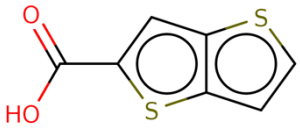
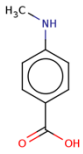
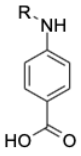
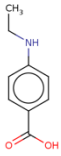
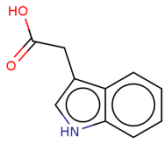
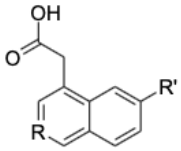
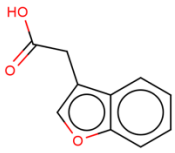
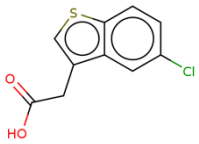
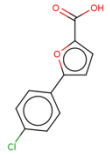
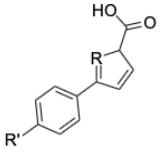
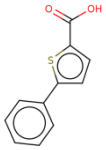
Clusters 4 and 13 represent a different example of highly similar, but alternative structural cores to what discussed so far. Although the generalized structures of both clusters consist of a carboxylic acid attached to a benzene ring, the length of these acids is found to have an effect on the observed % inhibition (Table 3.4). Despite their similarity, fragments from cluster 4 with a one carbon shorter side chain (propanoic acid) were found to have higher % inhibition compared to fragments from cluster 13, in which butanoic acid is attached to benzene. These data suggest that fragments 650 and 184 (Cluster 4) are able to inhibit HsFEN1 through a binding position, which seems to be not possible for fragments 178 and 183 (Cluster 13). Although in all cases inhibition is likely to occur through chelation of the active site metals by -COOH, the ability of these molecules to rotate might allow the adaptation of conformations, in which the longer fragments from cluster 13 are likely to form steric clashes.

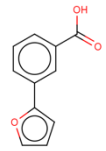
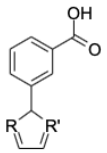
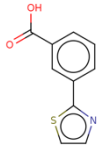
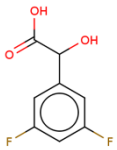
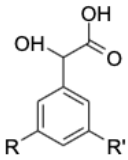
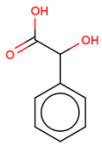
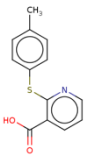
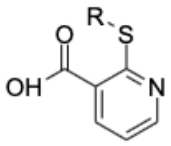
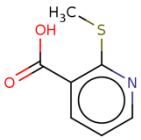
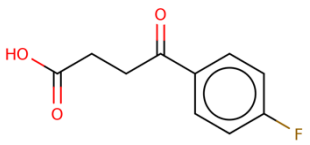
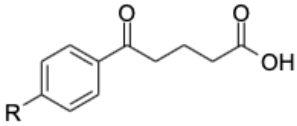
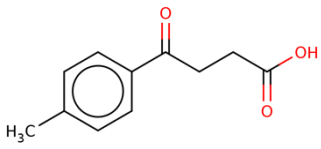
Despite the observed variability among the FRET repeats, similarity analysis revealed some common structural cores with conserved features. Interestingly, even the identified structural cores seemed to possess some level of similarity between them. In this section, comparisons focused on the common structural cores identified from each cluster, and not the individual ligands and their likely intermolecular interactions. Subsequent docking studies and prediction of the top scoring binding poses will allow visualization of protein-ligand complexes and direct comparisons of the interactions, likely to be responsible for inhibition. Further analysis of these molecules can potentially reveal an appropriate lead molecule with increased inhibition potency against HsFEN1.

Table 3.4 Clusters of HsFEN1 inhibitors identified from MaybridgeRO3 fragment library based on their structural similarities

[1]Cluster	Compound	Inhibition (%)	[2]Molecule structure	[3]Core structure
1	12	66.1		
	516	60.4		
	213	93.1		
	189	77.4		
	695	104.6		
	700	99.6		
2	595	106.8		
	283	69.8		
	543	148.8		

3	416	130.9		
	11	84.8		
	15	70.9		
	203	78		
4	650	97		
	194	82.8		
5	927	112.9		
	117	89.9		
6	17	89.1		

	171	77.1		
	573	115.5		
7	468	93.8		
	163	77.7		
8	480	92.9		
	537	110.4		
	62	88.1		
9	835	116.7		
	13	75.1		

10	784	124		
	601	119.4		
11	154	77.2		
	783	88.9		
12	175	98.8		
	176	81.8		
13	178	74		
	183	72.7		

[1]Similarity or activity cliffs analysis was performed using a binary fingerprint descriptor (FragFp) in DataWarrior 5.2.

[2]2D structures of the corresponding ligands were generated using the online Smi2Depict tool from ChemDB Chemoinformatics Portal (Chen et al., 2007b).

[3]Core structures were generated through ChemDraw version 19 online tool.

3.3 *In silico* docking of MaybridgeRO3 fragment library against HsFEN1: evaluation of the prediction accuracy

3.3.1 Preparation of target macromolecule and search space selection

Aiming to develop an effective molecular docking strategy, which will ideally allow rapid and cost-effective identification of potential HsFEN1 inhibitors, MaybridgeRO3 fragment library was also screened *in silico*. Multiple docking runs were performed under differing conditions, to determine the parameters which gave the best correlation between the experimental % inhibitions and the estimated binding energies is obtained.

Among the available crystal structures of HsFEN1 in the PDB database, a 2.1 Å structure published by Tainer and co-workers was selected as the target macromolecule for docking (PDB ID: 5K97) (Tsutakawa *et al*., 2017). Selection criteria involved mainly resolution and completeness. Although 5K97 represents an active site mutant of HsFEN1 complexed with DNA, it is the highest resolution structure of HsFEN1 available to date with a complete helical gateway. Helical gateway is a flexible structurally conserved region in FENs, which exists in two different conformations (ordered/disordered), depending on DNA binding and threading (AlMalki *et al*., 2016; Tsutakawa *et al*., 2017). Structural heterogeneity and regional plasticity pose a substantial challenge in protein crystallization. Flexible regions are often difficult to capture and in many cases their excision is preferred to facilitate crystallization. Therefore, although there are incomplete crystal structures of HsFEN1 with resolution as high as 1.9 Å, 5K97 was selected as the most appropriate highest resolution structure of HsFEN1 for docking studies (Xu *et al*., 2018).

Prior to molecular docking however, a number of pre-processing steps were still required to convert the HsFEN1 protein, from its raw state in 5K97, into the appropriate state for binding free energy calculations. Substrate DNA, water molecules and co-factors, such as metals, were all removed with the only exception of the two active site metals, required for optimal nucleolytic activity. Since in 5K97 structure, HsFEN1 was crystallised in the presence of Sm^{3+} ions instead of Mg^{2+}, the corresponding Sm^{3+} ions were replaced by Mg^{2+}, to resemble better the biologically active and stable state of the protein. The mutated residue (Asn233) was reverted back to the native aspartic acid and polar hydrogen atoms were added using AutoDock. Structures determined by X-ray crystallography, typically lack hydrogen atoms due to their respective crystallisation environment and resolution, which often makes it difficult to detect. Maintaining these hydrogens is critical to correctly estimate the binding affinity of a ligand due to their ability to form non-covalent interactions and stabilise the protein-ligand complexes, whether that refers to a DNA substrate and/or inhibitor in the case of HsFEN1 (Lippert and Rarey, 2009).

Next, a docking search space was defined. By definition, docking search space represents the space in which each ligand will be docked in every possible pose for the identification of those orientations and conformations that are most energetically favourable (Feinstein and Brylinski,

2015). Ideally, the docking box must be large enough to generate sufficient conformations of the largest ligand docked, but small enough to avoid high numbers of biologically irrelevant conformations. Therefore, a confined gridbox with 58 x 64 x 77 Å size, centred at -73.904, -78.164, 118.268, x, y and z coordinates respectively, was defined as the docking search space for the identification of potential HsFEN1 inhibitors (Figure 3.6). The size and orientation of this gridbox was adjusted to allow docking within the HsFEN1 active site and part of the helical gateway (Figure 3.6). Neighbouring residues R100 and Y40, located at the base of the arch, were selectively assigned as flexible during docking (Figure 3.6). Similar to the other members of RAD2/XPG family, processing of the ss 5′-flap is mediated by the two-metal mechanism and their interactions with the carboxylate-rich active site. Despite the fact that both the active site and helical gateway are highly conserved, targeting the helical gateway can potentially lead to some level of specificity. Unlike the active site, the helical gateway possesses a relatively low sequence similarity driving substrate specificity. Therefore, despite the low MW of MaybridgeRO3 fragments, the potential HsFEN1 inhibitors identified at this stage should ideally reveal critical interactions with the helical arch, likely to drive specificity.

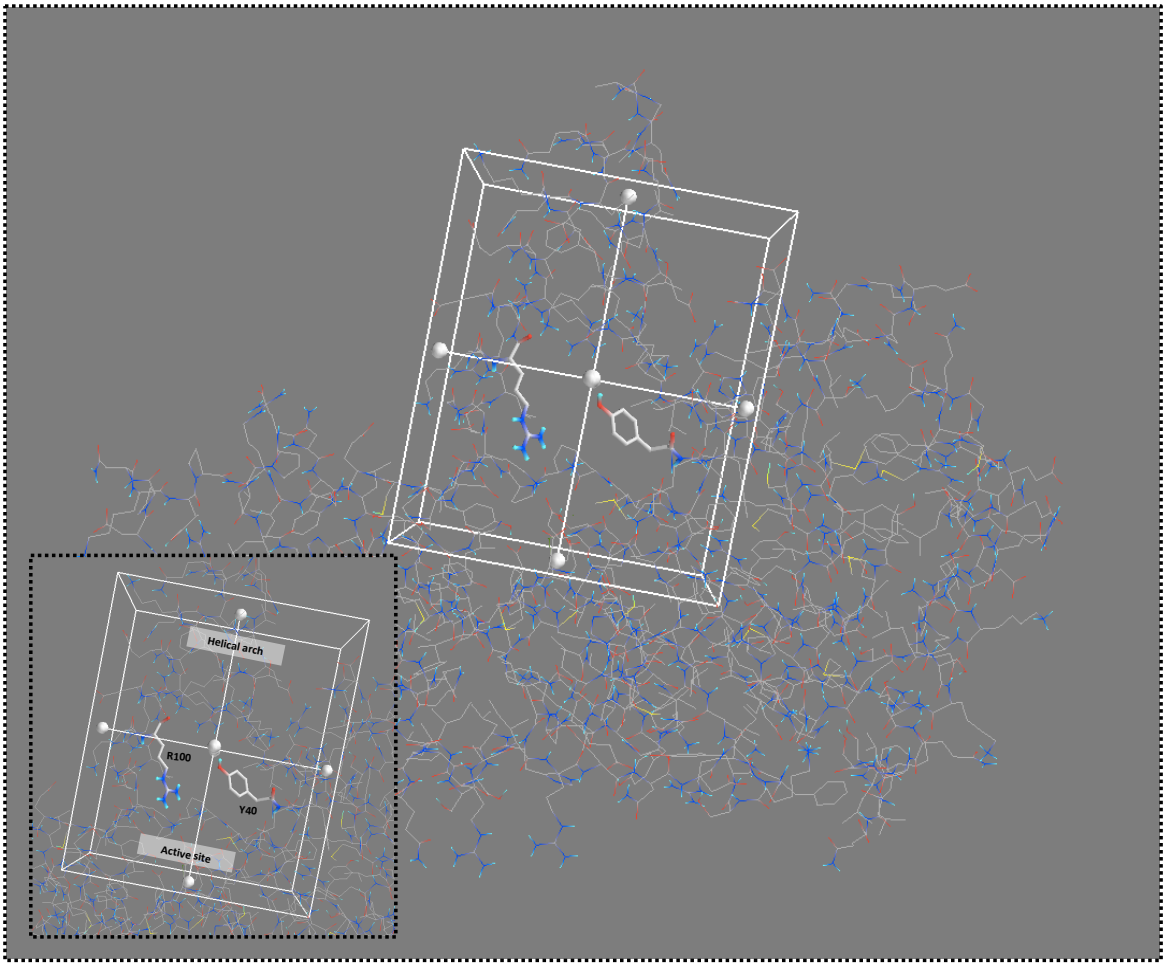


Figure 3.6 Defined docking search space for the identification of HsFEN1 inhibitors from MaybridgeRO3 fragment library. Open- and close-up views of the orientation and size set for the defined gridbox on the processed HsFEN1 structure, covering the active site and helical arch. Bottom left box highlights the helical gateway, HsFEN1 active site and the flexible R100 and Y40 residues (sticks).

3.3.2 Molecular docking of MaybridgeRO3 fragment library using AutoDock Vina

The 1,000 virtual ligands from MaybridgeRO3 library were initially docked against HsFEN1 using AutoDock Vina 1.1.2 (Trott and Olson, 2010). Default docking parameters were used and the top 10 poses for each individual ligand were output based on their predicted binding energies (kcal/mol). Typically, the lower the predicted binding energy is, the higher the *in vitro* binding affinity, which translates in more potent HsFEN1 inhibitors.

Average binding energies between -2.87 and -6.69 kcal/mol were predicted, with only 32 compounds having an average binding energy < -6 kcal/mol (Figure 3.7). Taking into account that some ligands completely abolished HsFEN1 activity *in vitro*, the predicted binding energies for the top scoring ligands are higher than expected. Typically, a binding energy threshold of -7 kcal/mol and even lower is used to differentiate strong binders with increased *in vitro* inhibition potential from non-specific/weak binders (Chang *et al*., 2007).

Bravais-Pearson analysis was carried out to correlate the predicted binding energies for all 1,000 MaybridgeRO3 ligands with the experimental % inhibitions from FRET-based HTS screening. No correlation between the two datasets was obtained, even for the top scoring compounds (Figure 3.7). In fact, none of the top 5 fragments with the lowest binding energies predicted was found to inhibit HsFEN1 *in vitro* (Table 3.5). Only compounds 333 and 568 from the 32 ligands with the highest predicted binding energies (< -6 kcal/mol) were shown to inhibit HsFEN1 *in vitro*. However, none of these was classified in any of the identified clusters during the previous similarity analysis (Section 3.2.5).

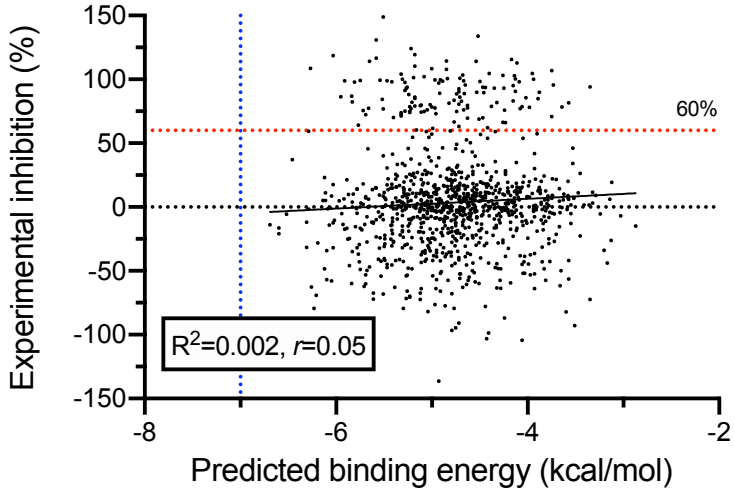
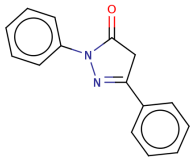
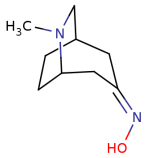
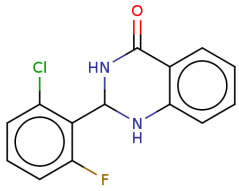
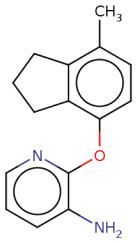
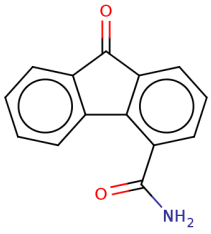


Figure 3.7 Relationship between experimental % inhibition and predicted binding energy for MaybridgeRO3 ligands calculated by AutoDock Vina. The plotted binding energy values represent the mean binding energy of the top 10 poses with minimized energy conformations for each individual ligand. Red and blue dotted lines are indicative of the 60% and -7 kcal/mol thresholds used for the identification of top scoring HsFEN1 inhibitors, respectively. R^2 ($1 \geq R^2 \geq 0$), representative of the strength, and r ($1 \geq r \geq -1$) representative of the direction of the linear correlation between the two datasets were calculated using Bravais-Pearson analysis in GraphPad Prism 8.2. Data points corresponding to six compounds with extremely negative % inhibitions *in vitro,* were excluded from analysis to allow plotting the correlation data within a sensible axis range.

Table 3.5 Top scoring compounds based on the predicted binding energies by AutoDock Vina

Compound	Predicted binding energy (kcal/mol)	Experimental inhibition (%)	Molecule structure
831	-6.69	-14	
392	-6.6	-16.4	
901	-6.6	-20.8	
791	-6.52	-5.6	
806	-6.47	-97.1	

2D structures of the corresponding ligands were generated using the online Smi2Depict tool from ChemDB Chemoinformatics Portal (Chen *et al*., 2007b).

The 3-hydroxy-1H-quinazoline-2,4-dione, belonging to the most extensively validated and established class of HsFEN1 inhibitors, was also docked against HsFEN1 as a reference ligand (control). This control docking run was performed aiming to understand the reasons behind the relatively high binding energies predicted by AutoDock Vina. Although 3-hydroxy-1H-quinazoline-2,4-dione is known to inhibit HsFEN1, the average binding energy of the top 10 conformations was also greater than -6 kcal/mol (Mean PBE: -5.6 kcal/mol) (Exell *et al*., 2016).

Further analysis of the predicted binding poses revealed binding modes inconsistent to literature. Structural studies on *N*-hydroxyurea class of inhibitors, including 3-hydroxy-1H-quinazoline-2,4-dione, showed that inhibition of HsFEN1 involves chelation of the active site metals (Exell *et al*., 2016). Contrary to that, the predicted binding modes of the reference ligand by AutoDock Vina involved in their majority interactions with the helical arch (Figure 3.8). The reference ligand seemed to be completely repelled from the carboxylate-rich active site, which is only justified by the absence or ignorance of the active site metals.

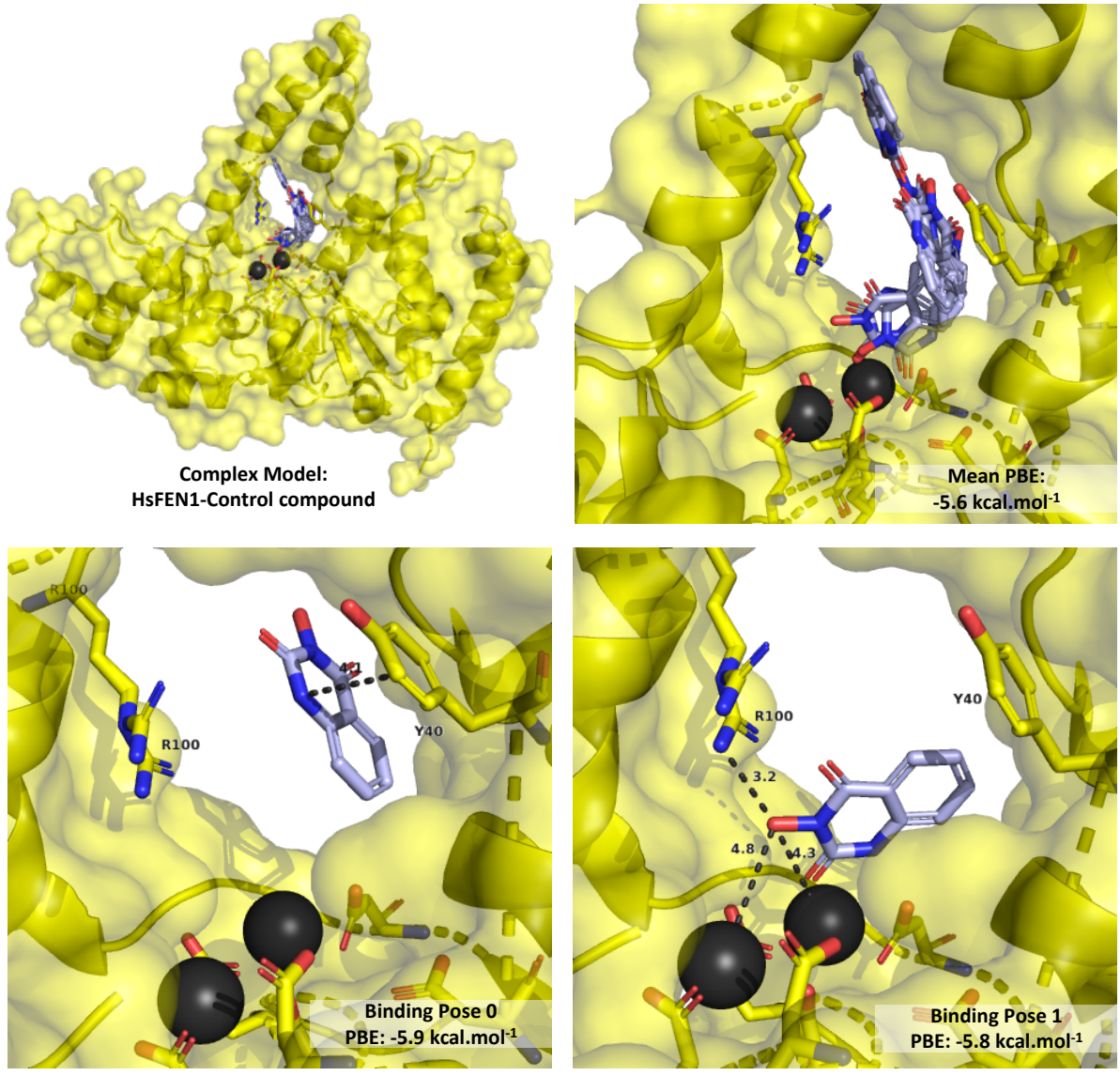


Figure 3.8 Models of HsFEN1-Control inhibitor (3-hydroxy-1H-quinazoline-2,4-dione) complex. (top left, top right) Open- and close-up views of top 10 binding poses distribution of control inhibitor along with their respective mean predicted binding energy (PBE). (bottom left, bottom right) Models of HsFEN1 active site and control inhibitor linkage at the top two binding poses with the lowest PBE, as these have been predicted by AutoDock Vina. Structure of HsFEN1 is shown in yellow cartoon, with selected key residues shown in sticks. Active site metals are shown as dark grey spheres, while control inhibitor is shown in light blue. Figures have been generated using the modified 5K97 PDB file, used as the target macromolecule for molecular docking, and rendered by PyMOL™ version 2.3.5 Schrödinger, LLC.

Grid maps and charges are pre-calculated by AutoDock Vina using a combination of a knowledge-based and empirical scoring function (Trott and Olson, 2010). Therefore, supplied atomic charges are being ignored, which is a critical parameter for the prediction accuracy of a docking software. Especially, in the case of metalloenzymes, such as HsFEN1, where the prediction accuracy of AutoDock and AutoDock Vina was found to vary across different families of metalloproteins (Vieira and Sousa, 2019). The relatively high binding affinities predicted, as well as the poor correlation with the experimental data, suggest that AutoDock Vina is not suitable for HsFEN1 docking studies and very likely nor for the wider FEN family of metallonucleases.

3.3.3 Molecular docking of MaybridgeRO3 fragment library using AutoDock and varying active site metal charges

The 1,000 ligands from MaybridgeRO3 library were docked against HsFEN1 using AutoDock 4.2 (Morris *et al*., 2009). Unlike AutoDock Vina, a combination of an empirical force field scoring function and the in-build Lamarckian genetic algorithm is used by AutoDock for the prediction of free binding energy. AutoDock therefore, accounts user-supplied charges, which especially in the case of active site metals can determine the prediction accuracy of a docking model (Chen *et al*., 2007b).

Aiming to develop a docking model with improved prediction power, nine docking runs were performed with incremental alterations to the active site metal charges (Mg^{0+} - Mg^{2+}). A gridbox analogous to that used in AutoDock Vina with 58 x 64 x 77 Å size, centred at -73.904, -78.164, 118.268 (x, y and z, respectively), and 0.375 Å spacing was used. Each dataset was correlated to the experimental % inhibitions using Bravais-Pearson correlation analysis.

Overall, the higher the assigned active site metal charges, the stronger the negative correlation between the experimental and *in silico* datasets with the only exemptions at +0.25 and zero charge (Figure 3.9). Docking with +2 metal charges, which is the highest tested charge, showed the stronger linear relationship with the experimental dataset (R^2=0.22, r=-0.5) and predicted binding affinities as low as -10.911 kcal/mol (Figure 3.9; Table 3.6).

Docking with +2 metal charges, which is the highest tested charge, showed the stronger linear relationship with the experimental dataset (R^2=0.22, r=-0.5) and predicted binding affinities as low as -10.911 kcal/mol (Figure 3.9; Table 3.6). Although with an R^2 at 0.22 the correlation between the predicted binding energy and experimental inhibition for a particular small-molecule remains poor, docking with active site metal charges at +2 appears efficient for prioritization of strong HsFEN1 binders that are likely to have a more significant inhibitory effect over HsFEN1 (Figure 3.9). For example, among two compounds A and B, both with ≥60% inhibition in vitro and predicted binding affinities lower or equal to -7 kcal/mol, the compound with the lowest binding energy will not necessarily have the highest experimental inhibition over HsFEN1. Therefore, conclusions on potency based on the *in silico* docking results, even under +2 active site metal charges, may not be appropriate. However, both compounds A and B will appear as "hits" for downstream screening applications.

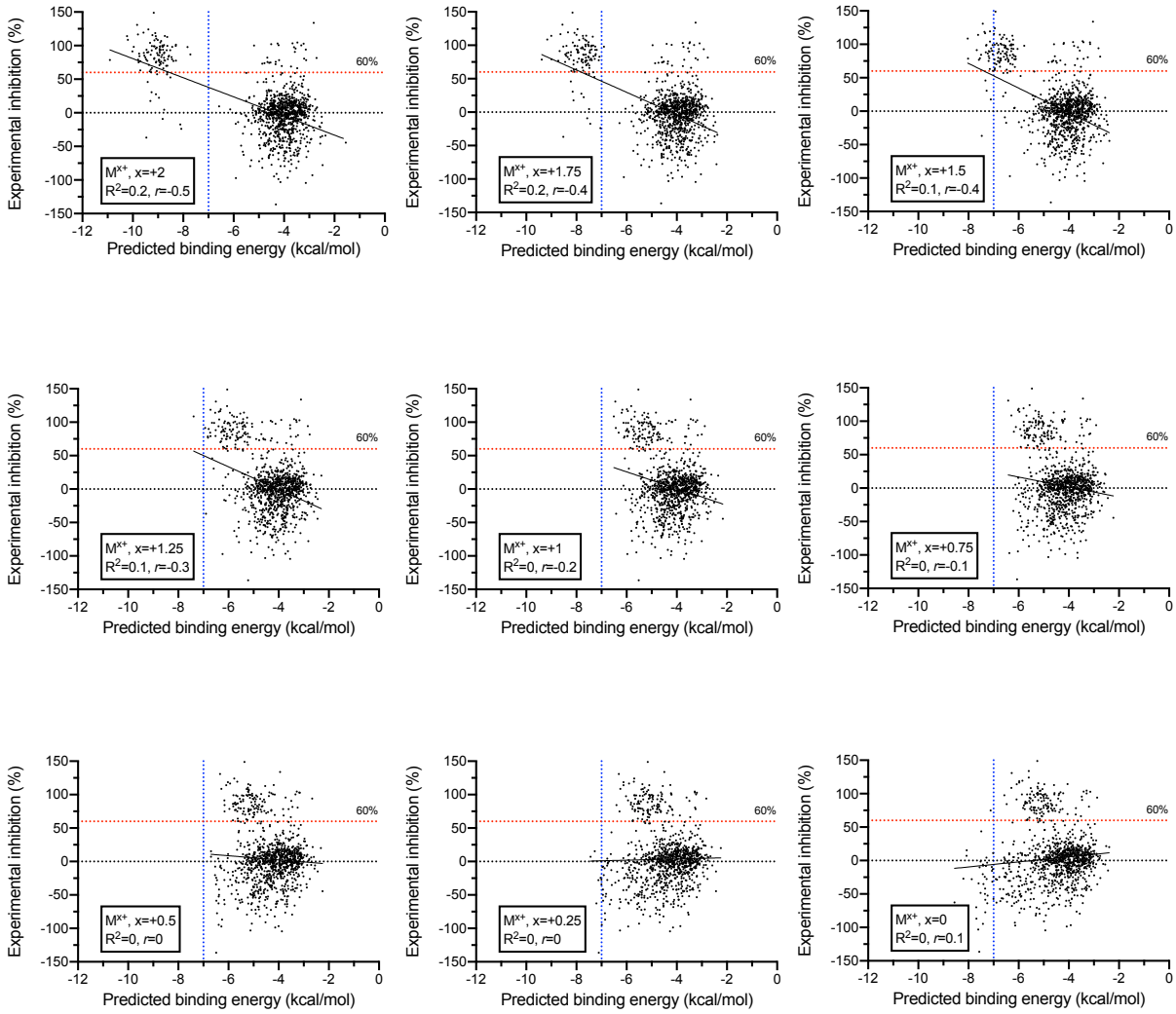


Figure 3.9 Effect of incremental increases of active site metal ion charges on the correlation between the experimental % inhibition and predicted binding energy of MaybridgeRO3 library. The plotted binding energy values represent the mean binding energy of the top 10 poses with minimized energy conformations for each individual ligand. Red and blue dotted lines are indicative of the 60% and -7 kcal/mol thresholds used for the identification of top scoring HsFEN1 inhibitors, respectively. R^2 ($1 \geq R^2 \geq 0$), representative of the strength, and r ($1 \geq r \geq -1$) representative of the direction of the linear correlation between the two datasets were calculated using Bravais-Pearson analysis in GraphPad Prism 8.2. Data points corresponding to six compounds with extremely negative % inhibitions *in vitro,* were excluded from analysis to allow plotting the correlation data within a sensible axis range.

Comparative analysis between the number of predicted hits and the number of experimentally (FRET assay) inhibitors, also supported the above observation. Using an initial cutoff of -7 kcal/mol, previously shown to be the "significance" cutoff for the efficient differentiation of strong and non-specific/weak binders, the number of hits per docking run was determined (Chang *et al.*, 2007). Docking run performed with +2 metal charges, revealed 114 compounds with predicted binding affinities lower or equal to -7 kcal/mol. Out of these, 99 molecules

(86.8%) were found to inhibit ≥ 60% of HsFEN1 activity *in vitro*, which represent the 78% of the total number of hits identified *in vitro* (Table 3.6). Although the dataset collected at +1.75 metal charges did not show significant difference compared to that of +2 metal charges, the prediction power observed at +1.5 or lower was significantly decreased. In fact, datasets collected at +1, +0.75, +0.5 and +0.25 metal charges did not show any predictive power (Table 3.6).

Since in these runs the lower predicted binding energies were higher than runs performed with +2 and +1.75 metal charges, the possibility of obtaining equal or improved prediction power with a loosened redefined threshold was explored. Predicted hits were therefore quantified using a -6 kcal/mol cutoff. As expected, the use of a higher binding affinity threshold led to higher numbers of predicted hits (Table 3.6). Similarly, the numbers of predicted hits with confirmed *in vitro* inhibition activity over HsFEN1 were also increased. However, with the exception of docking runs with +2, +1.75 and +1.5 metal charges, the overall prediction power remained low (Table 3.6). Applying the -6 kcal/mol cutoff for the dataset obtained with +2 led to no increase in the number of predicted hits confirmed experimentally, but to a slight increase for the datasets obtained at +1.75 metal charge (78% compared to 77.2%). However, that extra hit responsible for the observed 0.8% increase, has to be identified among three additional predicted false-positive hits (117 compared to 113).

Generally, the efficiency of a docking strategy should not only be determined by the number of predicted hits that have been confirmed experimentally to have the desired effect over the target protein. Maintaining a high ratio between the predicted and experimentally determined hits is equally important. For example, between two docking runs with 100 and 150 predicted hits respectively, but only 50 hits confirmed *in vitro*, the most efficient is the one with 50% success rate on the number of predicted hits. The corresponding success rates of docking runs performed with +2 and +1.75 metal charges (86.8%) at -7 kcal/mol cutoff were found to be higher than those at -6 kcal/mol cutoff (85.3% and 84.6%, respectively). Based on these, a cutoff of -7 kcal/mol and active site metal charges of +2 were selected as the optimum conditions for the *in silico* identification of HsFEN1 inhibitors. Despite the fact that there is no significant difference in the success rates at +2 and +1.75 metal charges, the stronger separation between HsFEN1 inhibitors and non-inhibitors observed at +2 metal charge was the main criteria driving that decision (Figure 3.9).

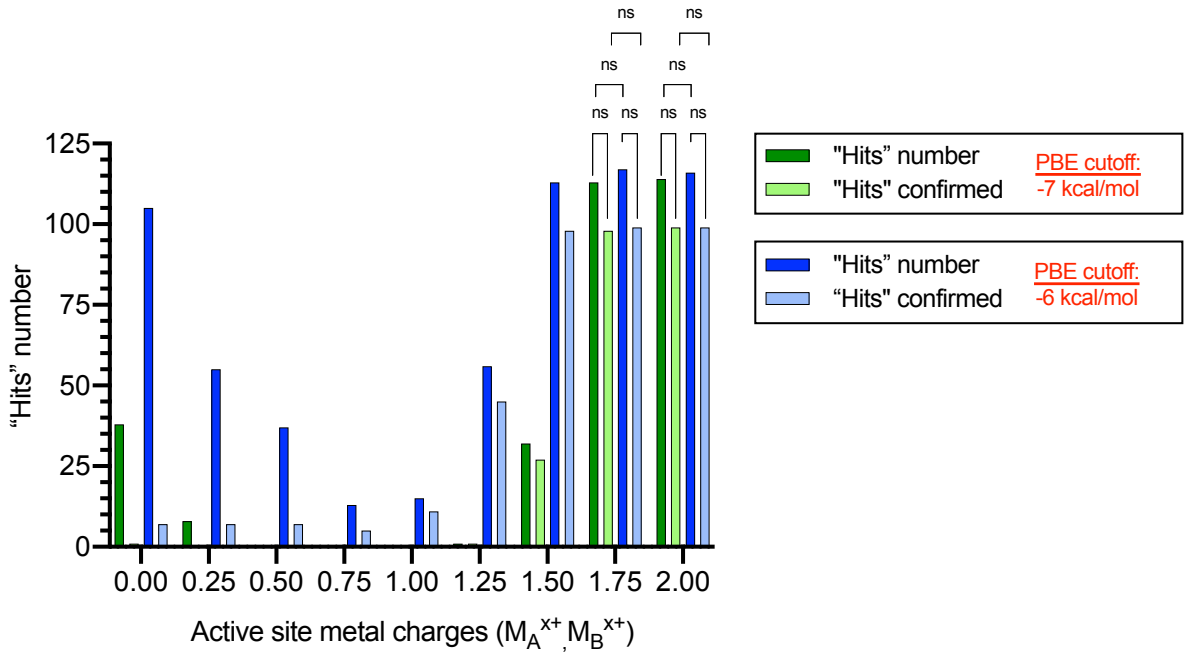


Figure 3.10. *In silico* predicted and *in vitro* confirmed "hits" from MaybridgeRO3 compound docking, under tweaked active site metal charges. "Hits" numbers (green and blue) are representative of the predicted number of "hits" per docking run, performed under differing active site metal charges (0 to +2). Numbers of "hits" confirmed (light green and light blue), are indicative of the number of predicted "hits", whose inhibitory activity over HsFEN1 has been confirmed *in vitro* by FRET-based HTS. In each case, the plotted "hits" numbers and numbers of "hits" confirmed correspond to the emerging numbers, using a PBE threshold of either -7 kcal/mol (shades of green) or -6 kcal/mol (shades of blue). Significance levels were analysed using *the two-way ANOVA* model in GraphPad Prism 8.4. and are only shown for the docking runs with the stronger prediction powers.

Table 3.6 Comparison between the number of "hits" predicted during docking with tweaked active site metal charges

Binding energy cutoff (kcal/mol)		Charge on active site Mg^{x+} ions								
		+2	**+1.75**	**+1.5**	**+1.25**	**+1**	**+0.75**	**+0.5**	**+0.25**	**0**
-7	[1]No. of "hits"	114	113	32	1	0	0	0	8	38
	[2]No. of "hits" confirmed *in vitro* (%)	99 (78%)	98 (77.2%)	27 (21.2%)	1 (0.8%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	1 (0.8%)
-6	[1]No. of "hits"	116	117	113	56	15	13	37	55	105
	[2]No. of "hits" confirmed *in vitro* (%)	99 (78%)	99 (78%)	98 (77.2%)	45 (35.4%)	11 (8.7%)	5 (4%)	7 (5.5%)	7 (5.5%)	7 (5.5%)
	Lowest predicted binding energy (kcal/mol)	-10.9	-9.4	-8.	-7.4	-6.5	-6.4	-6.7	-7.4	-8.6

[1]Number of "hits" per docking run is determined using a predicted binding energy threshold of either -7 or -6 kcal/mol, as indicated above.

[2]Number of "hits" confirmed by both AutoDock and FRET-based HTS. Also expressed as the percentage of the number of "hits" predicted per docking run out of the 127 "hits" identified *in vitro*.

Of course, the above observations and conclusions presuppose the acceptance of the experimental data as the "gold" standard. However, both FRET-based HTS screening and molecular docking have their own limitations. Even when using the docking model with the strongest prediction power among those tested (Mg^{2+}), 15 (22%) experimentally confirmed hits were not predicted as hits by the selected *in silico* model. These are likely to be false positives of the HTS screening or "true" inhibitors, driving inhibition through a mechanism that is not supported by the proposed *in silico* model. The increased number of predicted hits at zero magnesium charge also support that statement. Using the -6 kcal/mol cutoff, 7 predicted hits at zero magnesium charge were found to have significant inhibitory activity against HsFEN1 *in vitro*. Interestingly, two of these appeared to be unique predictions of the zero-magnesium charge model, indicating that their inhibition mechanism does not involve chelation of the active site metals. These class of strong HsFEN1 binders that are non-metal chelators, can potentially be an interesting class of inhibitors to study as an alternative for HsFEN1 inhibition. However, given the high number of hits predicted at zero magnesium charge and the extremely low percentage of the "true" hits confirmed in vitro, non-metal chelators seem to be the most difficult and least cost-/time-efficient to identify at an early *in silico* stage.

3.3.4 Molecular docking of MaybridgeRO3 fragment library against the entire HsFEN1 molecule using AutoDock

Aiming to examine whether docking against the entire HsFEN1 macromolecule, instead of the predefined active-site focused search space used so far, increases the prediction power of the proposed docking model, two additional docking runs were performed. Typically, the predicted binding energies and poses are limited to the ligand search space, used for docking. Although docking against the entire protein macromolecule, also referred to as blind docking, is computationally prohibitive, it is often required for the potential identification of alternative ligand-preferred binding pockets, other than the known active sites (Forli *et al*., 2016).

Additional docking runs were performed in the presence (Mg^{2+}, Mg^{2+}) and absence (Mg^{0}, Mg^{0}) of active site metals, against the entire HsFEN1 molecule, as the defined ligand search space. Comparison between the "hits" numbers predicted here, and those predicted in the corresponding runs against the active site-focused search space in Section 3.3.3, revealed statistically insignificant differences (Figure 3.11). In the absence of active site metals or metals at 0 charge, none of the predicted hits was found to inhibit ≥ 60% of HsFEN1 activity, when docked against the entire HsFEN1 molecule, and only one, when docked against the active site-focused search space (Table 3.7). However, given the number of experimentally confirmed hits, docking against the active site-focused search space, with active site metals charge at +2, was found to still possess the strongest prediction power among those tested (Table 3.7). In addition, the calculated binding energy for the top scoring ligand was predicted to be -10.9 kcal/mol, lower than the predicted binding energy during docking with +2 metal charge and search space the entire macromolecule (PBE: -9.3 kcal/mol).

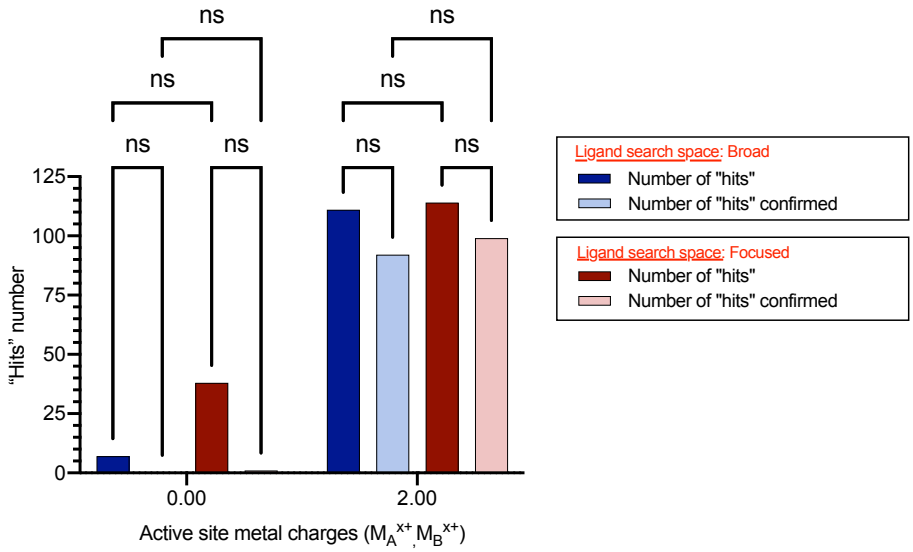


Figure 3.11 *In silico* predicted and *in vitro* confirmed "hits" from MaybridgeRO3 compound docking, against the entire HsFEN1 molecule or an active site-focused search space. "Hits" numbers (blue and red) are representative of the predicted number of "hits" per docking run, performed in the presence (Mg^{2+}, Mg^{2+}) or absence (Mg^{0}, Mg^{0}) of active site metal charges and under two distinct search spaces. Numbers of "hits" confirmed (light blue and light red), are indicative of the number of predicted "hits", whose inhibitory activity over HsFEN1 has been confirmed *in vitro* by FRET-based HTS. In each case, the plotted "hits" numbers and numbers of "hits" confirmed correspond to the emerging numbers, using a PBE threshold of -7 kcal/mol. Significance levels were analysed using *the two-way ANOVA* model in GraphPad Prism 8.4.

Table 3.7 Comparison between the number of "hits" predicted during docking in the presence of Mg^{2+}/Mg^{0}, against the entire HsFEN1 molecule or an active site-focused search space

[1]Ligand search space	Broad	Focused	Broad	Focused
Metal ion charge	+2	+2	0	0
[2]No. of "hits"	111	114	7	38
[3]No. of "hits" confirmed *in vitro* (%)	92 (72.4%)	99 (78%)	0 (0%)	1 (0.8%)
[4]No. of poses ≤-7 kcal/mol	1012	1139	139	349
Lowest predicted binding energy (kcal/mol)	-9.3	-10.9	-8.0	-8.6

[1]Focused search space, refers to the predefined search space used in Section 3.3.3., with 58 x 64 x 77 Å size, centred at -73.904, -78.164, 118.268 (x, y and z, respectively), and 0.375 Å spacing. Broad search space, refers to the entire HsFEN1 protein molecule.
[2]Number of "hits" per docking run is determined using a predicted binding energy threshold of -≤-7 kcal/mol.
[3]Number of "hits" confirmed by both AutoDock and FRET-based HTS. Also expressed as the percentage of the number of "hits" predicted per docking run out of the 127 "hits" identified *in vitro*.
[4]Number of ligand poses with PBE ≤-7 kcal/mol, given that a total of 10,000 poses with minimized energy conformations are output for the entire 1,000 MaybridgeRO3 ligands screened.

For both docking runs in the presence and absence of active site metals, correlation analysis revealed linear relationships between the binding energies predicted, during docking against the entire HsFEN1 molecule and the active site-focused search space (Figure 3.12). Whilst this is in line with the statistically insignificant differences in the numbers of predicted and experimentally confirmed hits, top scoring binding poses distribution suggests that, at least for the 1,000 MaybridgeRO3 ligands docked, "true" inhibitors-HsFEN1 interactions are concentrated at protein's active site and its surrounding residues (Figure 3.12). Here, ligand binding regions, other than the predefined active site-focused search space, do not appear to drive HsFEN1 inhibition *in vitro*. Given these, it is concluded that docking against the entire HsFEN1 molecule does not enhance the sensitivity and prediction power of the proposed docking model, using the active site-focused search space and active site metal charges of +2.

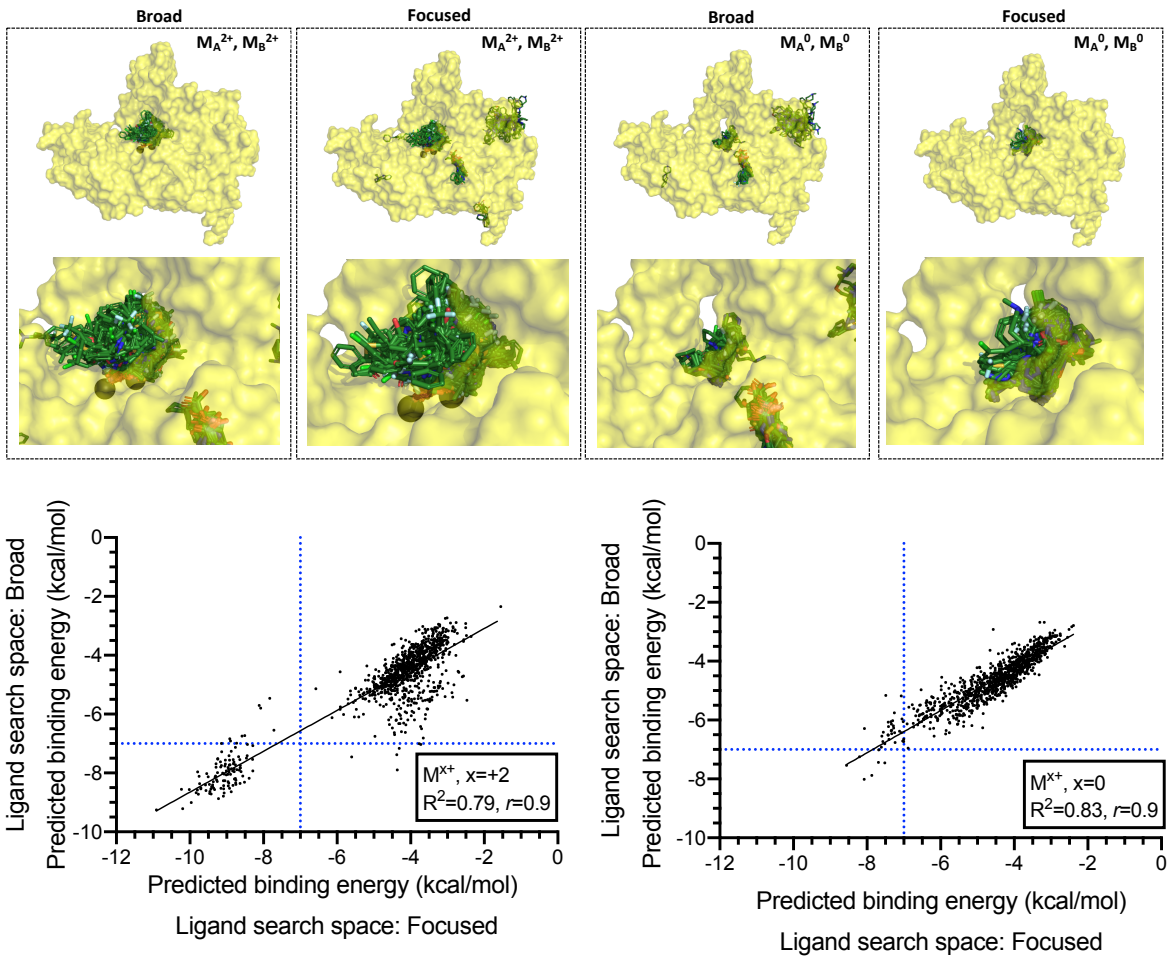


Figure 3.12 Top scoring binding poses distribution of MaybridgeRO3 compounds, when docked against the entire HsFEN1 molecule or an active site-focused search space. (a) Compound binding poses (green) with PBE ≤-7 kcal/mol across HsFEN1 (yellow) are shown in zoomed in and zoomed out, for each of the docking runs performed by AutoDock. Docking runs were carried out in the presence (Mg^{2+}, Mg^{2+}) and absence (Mg^{0}, Mg^{0}) of active site metals (grey spheres), against the entire HsFEN1 molecule or a predefined focused search space analogous to that of Section 3.3.3. (b) Correlations between the predicted binding energies of each compound docked under two differing search spaces. The plotted binding energy values represent the mean binding energy of the top 10 poses with minimized energy conformations for each individual small-molecule. Blue dotted lines are indicative of the -7 kcal/mol threshold used for the identification of top scoring HsFEN1 inhibitors, respectively. R^2 ($1 \geq R^2 \geq 0$), representative of the strength, and r ($1 \geq r \geq -1$) representative of the direction of the linear correlation between the two datasets were calculated using Bravais-Pearson analysis in GraphPad Prism 8.2. Figures have been generated using the modified 5K97 PDB file, used as the target macromolecule for molecular docking, and rendered by PyMOL™ version 2.3.5 Schrödinger, LLC.

3.3.5 Molecular docking of MaybridgeRO3 fragment library in the presence of individual metals and complete absence of metals

Aiming to explore whether identification of HsFEN1 inhibitors with alternative inhibition mechanisms is possible at an early *in silico* stage, additional docking runs were performed in

the presence of each individual active site metal and in complete absence of both. Docking runs were performed using the optimized +2 magnesium charge for the metals presented in the runs and the resultant datasets were correlated to the experimental % inhibitions.

As expected due to the absence of one of the active site metals or even both, the lowest predicted binding energy observed in each dataset was higher than that in the presence of both active site metals (Table 3.8). Bravais-Pearson analysis revealed correlation patterns similar to those observed before (Figure 3.13). The correlation of the binding energies predicted during docking in the presence of either M_A^{2+} or M_B^{2+}, were found to be similar to those performed with both active site metals at +0.75 and +1.75 metals charges, respectively. This, however, was not supported by the number of predicted hits (Table 3.8). Similar correlation patterns and number of predicted hits were only obtained in the datasets collected in the absence of both active site metals and active site metals with zero charge (Figure 3.13).

Table 3.8 Comparison between the number of "hits" predicted in the presence of M_A^{2+}, M_B^{2+} or complete absence of both active site metals

Binding energy cutoff (kcal/mol)		**M_A^{2+}**	**M_B^{2+}**	**No metals**	**[1]Mg_A^0 & Mg_B^0**	**[1]M_A^{2+} & M_B^{2+}**
-7	[2]No. of "hits"	6	74	47	38	114
	[3]No. of "hits" confirmed *in vitro* (%)	1 (0.8%)	64 (50.4%)	1 (0.8%)	1 (0.8%)	99 (78%)
-6	[2]No. of "hits"	87	116	132	105	116
	[3]No. of "hits" confirmed *in vitro* (%)	25 (19.7%)	99 (78%)	8 (6.3%)	7 (5.5%)	99 (78%)
	Lowest predicted binding energy (kcal/mol)	-7.68	-8.47	-9.12	-8.557	-10.911

[1]Mg_A^0 & Mg_B^0 and M_A^{2+} & M_B^{2+} datasets have been copied from Table 3.6 (Section 3.3.3) to allow direct comparisons.

[2]Number of "hits" per docking run is determined using a predicted binding energy threshold of either -7 or -6 kcal/mol, as indicated above.

[3]Number of "hits" confirmed by both AutoDock and FRET-based HTS. Also expressed as the percentage of the number of "hits" predicted per docking run out of the 127 "hits" identified *in vitro*.

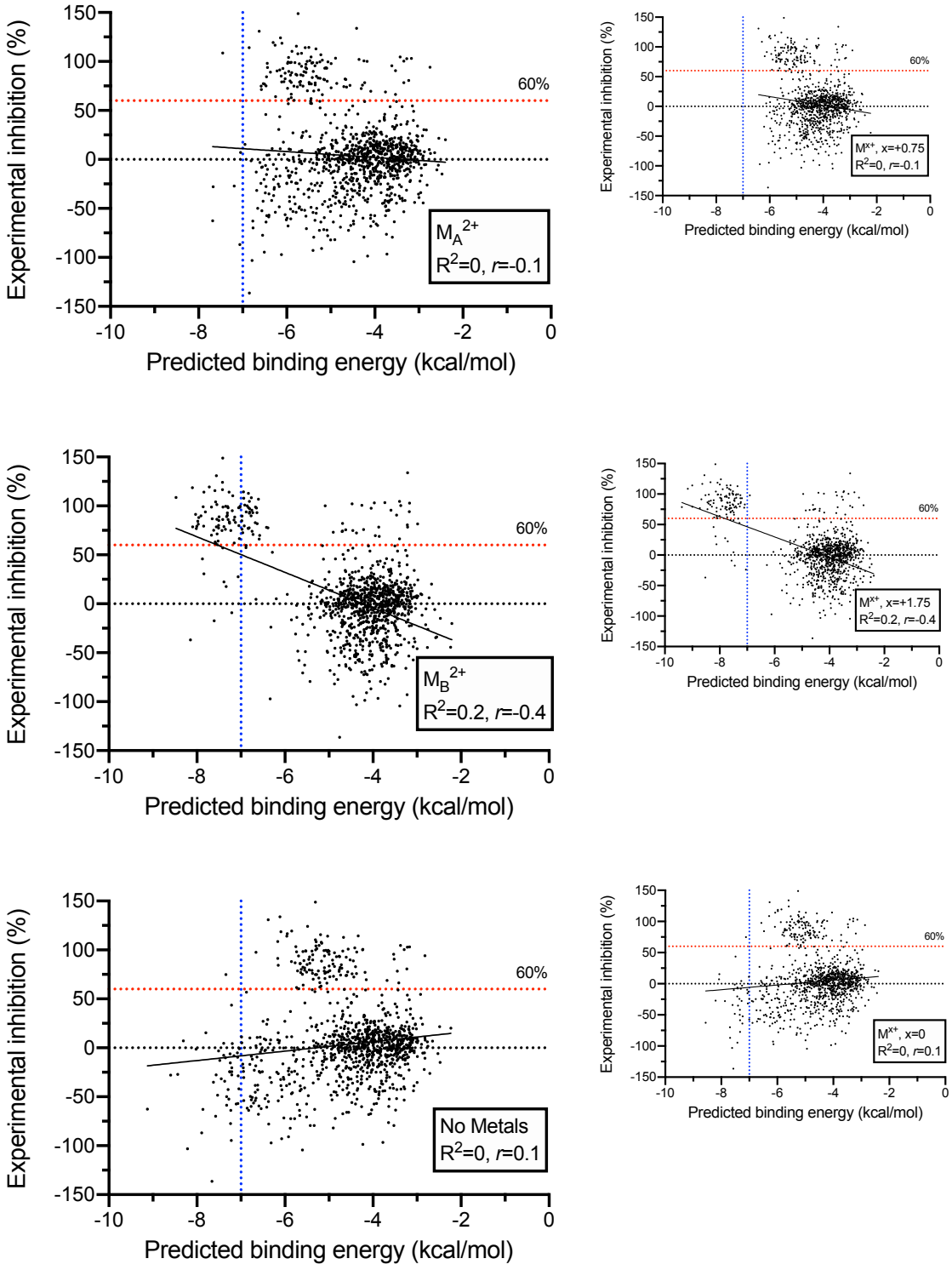


Figure 3.13 Correlation between experimental inhibitions and *in silico* docking datasets, collected in the presence of M_A^{2+}, M_B^{2+} or both. The plotted binding energy values represent the mean binding energy of the top 10 poses with minimized energy conformations for each individual ligand. Red and blue dotted lines are indicative of the 60% and -7 kcal/mol thresholds used for the identification of top scoring HsFEN1 inhibitors, respectively. Graphs on the right are representative of docking runs performed in section 3.3.3 that were found to resemble significantly similar correlation with the docking runs performed in the presence of either M_A^{2+}, M_B^{2+} or both active site metals. R^2 ($1 \geq R^2 \geq 0$), representative of the strength, and r ($1 \geq r \geq -1$) representative of the direction of the linear correlation between the two datasets were calculated using Bravais-Pearson analysis in GraphPad Prism 8.2. Data points corresponding to six compounds with extremely negative % inhibitions *in vitro*, were excluded from analysis to allow plotting the correlation data within a sensible axis range.

The number of hits predicted in the sole presence of M_A^{2+} was found to be lower compared to those predicted in the sole presence of M_B^{2+} and in complete absence of active site metals (Table 3.8). Whether a binding energy cutoff of -7 or -6 kcal/mol was used, none of the identified hits appeared to be a unique prediction of docking in the sole presence of M_A^{2+}. On the other hand, the highest number of predicted hits was obtained during docking with M_B^{2+}, which was found to correlate strongly with the docking run performed in the presence of both M_A^{2+} and M_B^{2+} (Table 3.8; Figure 3.14). This strong correlation, in addition to the fact that the great majority of the identified hits were unique predictions of that run, suggest that metal coordination ligands seem to form tighter interactions with M_B^{2+} and/or the neighboring residues around M_B^{2+} than M_A^{2+}.

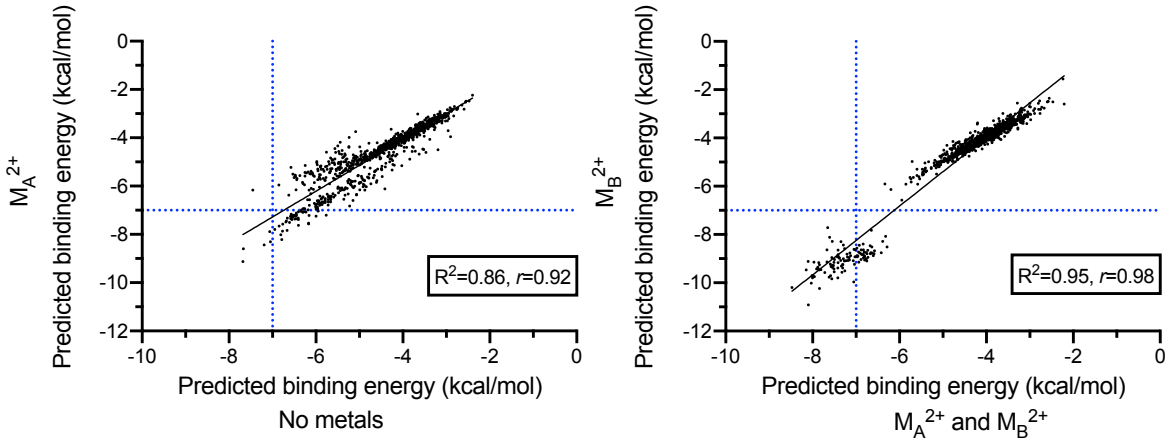


Figure 3.14 Linear correlation between *in silico* datasets collected in the sole presence of with datasets collected in the presence of both and in complete absence of active site metals, respectively. The plotted binding energy values represent the mean binding energy of the top 10 poses with minimized energy conformations for each individual ligand. Blue dotted lines are indicative of the -7 kcal/mol threshold used for the identification of top scoring HsFEN1 inhibitors. R^2 ($1 \geq R^2 \geq 0$), representative of the strength, and r ($1 \geq r \geq -1$) representative of the direction of the linear correlation between the two datasets were calculated using Bravais-Pearson analysis in GraphPad Prism 8.2.

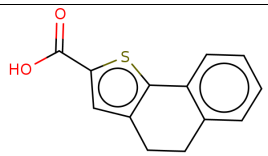
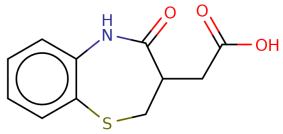
Coordination of active site metals, whether that involves M_A^{2+}, M_B^{2+} or both, seems to be the predominant inhibition mechanism across the top scoring compounds of MaybridgeRO3 library. This, according to the docking poses analysed. Although the majority of these metal chelators was predicted as hits in the presence of both active site metals, dockings carried out in the sole presence of M_B^{2+} have also effectively predicted such ligands. However, due to the absence of M_A^{2+} the observed predicted binding energies were higher, suggesting that with the additional presence of M_A^{2+} chelation of both active site metals is likely to occur. Generally, the numbers of predicted hits suggest a preference of metal chelators for M_B^{2+} and very likely its neighboring residues. Subsequent modelling of the binding poses for some of the top scoring can potentially reveal some key interaction driving that preference.

3.3.6. Preliminary analysis of the top-scoring HsFEN1 inhibitors from MaybridgeRO3 library

As a first step towards the identification of novel HsFEN1 inhibitors, screening of MaybridgeRO3 fragment library, using both a FRET-based HTS assay and *in silico* docking, resulted in a range of ligands with significant inhibitory potential against HsFEN1. In section 3.2.5, based on the *in vitro* data, these compounds were clustered and the identified structural cores were discussed speculating their respective intermolecular interactions. Here, the binding poses predicted during *in silico* docking will be analyzed further on an effort to model the intermolecular interactions for some of the top scoring compounds.

Focusing on those ligands, whose inhibitory activity was confirmed both *in silico* and *in vitro*, ligands 333 (*In vitro* inhibition: 108.7 %; Minimum PBE: -10.2 kcal/mol) and 625 (*In vitro* inhibition: 114.1 %; Minimum PBE: -10.2 kcal/mol), were selected for further interaction analysis. Based on the data collected so far, both ligands represent some of the most potent HsFEN1 inhibitors from MaybridgeRO3 library with predicted binding energies <-10 kcal/mol and *in vitro* inhibition >90% (Table 3.9). Despite their observed potency, these compounds were not classified in any of the identified clusters in the previous similarity analysis (Section 3.2.5). Therefore, analysis of their differing structural cores and the possible interactions with HsFEN1 can potentially unravel critical interaction pairs for HsFEN1 inhibition. Interestingly, even in the absence of active site metals or metals of zero charge, both 333 and 625 ligands were predicted to have binding energies ≤-6 kcal/mol (Table 3.9). Although the lowest predicted binding energies, indicative of a potentially increased inhibition potency, were found in docking runs with assigned active site metal charges ≤+1.5, these data suggested that ligands 333 and 625 are likely to form critically strong interactions with residues other than the active site metals. Of course, due to the presence of -COOH groups, coordination of active site metals upon their presence is expected. Indeed, using the predicted inhibitor-HsFEN1 models, both molecules were found to drive inhibition through chelation of the active site metals. For all top 10 poses, RMSD values, indicative of the distance between the predicted binding poses, were found to be lower than 2 Å. Generally, the chances of a binding pose having been predicted accurately increases when the top 10 binding poses of a particular ligand lie within 2 Å (Houston and Walkinshaw, 2013).

Table 3.9. Comparison between predicted binding energies and experimental inhibition of HsFEN1 by ligands 333 and 625

	Docking run	 [1]Ligand 333	 [1]Ligand 625
Binding energy predicted by AutoDock Vina (kcal/mol)	$M_A^{?}$, $M_B^{?}$	-6.1	-5.5
Binding energy predicted by AutoDock (kcal/mol)	M_A^{2+}, M_B^{2+}	-10.2	-10.2
	$M_A^{1.75+}$, $M_B^{1.75+}$	-9.2	-9
	$M_A^{1.5+}$, $M_B^{1.5+}$	-8.	-7.6
	$M_A^{1.25+}$, $M_B^{1.25+}$	-7.4	-6.5
	M_A^{1+}, M_B^{1+}	-6.5	-6.2
	$M_A^{0.75+}$, $M_B^{0.75+}$	-6.3	-6
	$M_A^{0.5+}$, $M_B^{0.5+}$	-6.3	-6.1
	$M_A^{0.25+}$, $M_B^{0.25+}$	-6.2	-6.1
	Mg_A^{0}, Mg_B^{0}	-6.3	-6
	M_A^{2+}	-7.4	-6.8
	M_B^{2+}	-8.5	-7.8
	No metals	-6.2	-6
***In vitro* inhibition (%)**		108.7	114.1

[1]2D structures of the corresponding ligands were generated using the online Smi2Depict tool from ChemDB Chemoinformatics Portal (Chen *et al*., 2007b).

Despite the intramolecular stabilizing interaction between thiazole sulphur and carbonyl oxygen in ligand 333, the increased number of functional groups presented in ligand 625 are likely to be responsible for the observed increase in potency. Given that in both cases the carboxyl group faces the active site metals, the ability of ligand 625 to bend and bring a second carbonyl group in close proximity to M_B^{2+} seems to be critical for HsFEN1 inhibition (Figure 3.15; Figure 3.16). In addition, across the predicted binding poses, that second carbonyl group in ligand 625 was found to form parallel interactions with some of the active site carboxylates (D86) and other neighboring residues at the bottom of the helical arch (R100, K97 and M37). The active site carboxylates were also found to form hydrogen bonds with the secondary amino group of ligand 625. Generally, although the predicted binding energies of the two ligands did

not differ substantially, preliminary interaction analysis using protein-ligand models supported the increased *in vitro* inhibition of HsFEN1 by ligand 625.

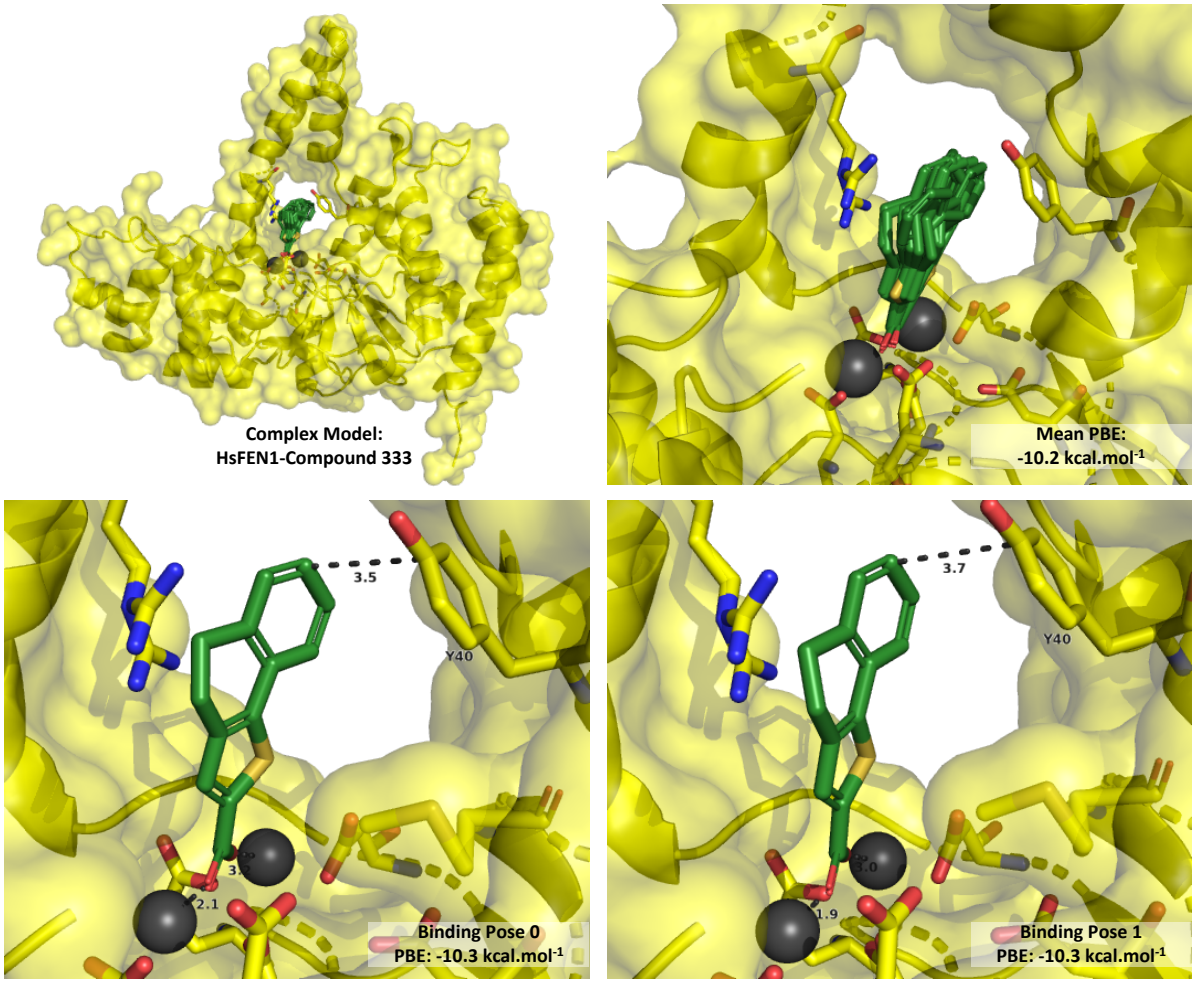


Figure 3.15 Models of HsFEN1-Compound 333 complex. (top left, top right) Open- and close-up views of top 10 binding poses distribution of MaybridgeRO3 inhibitor 333 along with their respective mean predicted binding energy (PBE). (bottom left, bottom right) Models of active site and compound 333 linkage at the top two binding poses with the lowest PBE, as these have been predicted by AutoDock4 in the presence of both active site metals at +2 charge. Structure of HsFEN1 is shown in yellow cartoon, with selected key residues shown in sticks. Active site metals are shown as dark grey spheres, while compound 333 is shown in green. Figures have been generated using the modified 5K97 PDB file, used as the target macromolecule for molecular docking, and rendered by PyMOL™ .

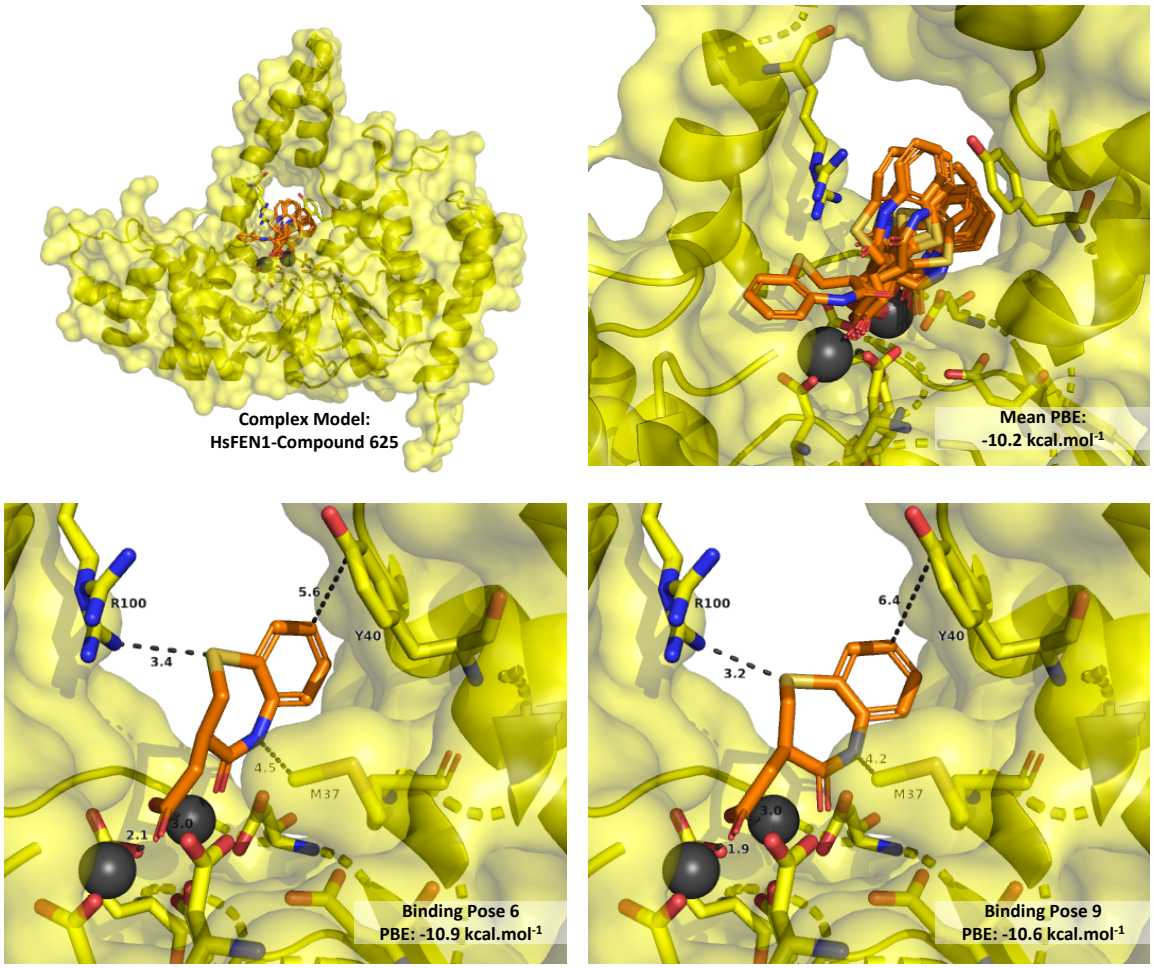
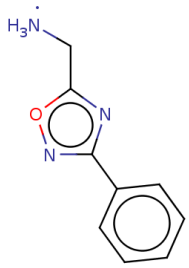
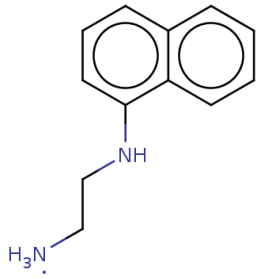


Figure 3.16 Models of HsFEN1-Compound 625 complex. (top left, top right) Open- and close-up views of top 10 binding poses distribution of MaybridgeRO3 inhibitor 625 along with their respective mean predicted binding energy (PBE). (bottom left, bottom right) Models of active site and compound 625 linkage at the top two binding poses with the lowest PBE, as these have been predicted by AutoDock4 in the presence of both active site metals at +2 charge. Structure of HsFEN1 is shown in yellow cartoon, with selected key residues shown in sticks. Active site metals are shown as dark grey spheres, while compound 625 is shown in orange. Figures have been generated using the modified 5K97 PDB file, used as the target macromolecule for molecular docking, and rendered by PyMOL™ .

Among the identified HsFEN1 inhibitors, ligands 43 and 776 were also found to inhibit about 74.9% and 97.9% of HsFEN1 activity, respectively (Table 3.10). Although their *in silico* identification appeared to be more challenging, in fact the experimental FRET-based data suggest that such a class of inhibitors can potentially be an interesting alternative for HsFEN1 inhibition. With predicted binding energies lying between -6 and -7 kcal/mol, ligands 43 and 776 were unique predictions of docking runs performed in the absence of active site metals or active site metals of zero charge. These, in addition to the presence of a positively charged amino group ($-NH_3^+$), on each of these ligands, suggested that ligands 43 and 776 are likely to interact directly with the carboxylate-rich active site of HsFEN1. Indeed, modelling of the predicted binding poses indicated 10 highly similar binding poses for each of these ligands, with the $-NH_3^+$ group always facing the negatively charged active site (Figure 3.17; Figure 3.18). Despite that, the interactions with the active site residues were found to vary depending

on the orientation of this rotatable -NH_3^+ group, attached at the end of a highly flexible side chain. In both cases, the close proximity of D34 with -NH_3^+ group seemed to be maintained, while other interactions with neighbouring residues of the helical arch were formed. Generally, ligands 43 and 776 seemed to probe deeper into the helical arch compared to the previously discussed metal chelating ligands.

Table 3.10 Comparison between predicted binding energies and experimental inhibition of HsFEN1 by ligands 43 and 776

	Docking run	 [1]Ligand 43	 [1]Ligand 776
Binding energy predicted by AutoDock Vina (kcal/mol)	$M_A^?$, $M_B^?$	-5.3	-5.4
Binding energy predicted by AutoDock (kcal/mol)	M_A^{2+}, M_B^{2+}	-4.5	-3.6
	$M_A^{1.75+}$, $M_B^{1.75+}$	-4.6	-3.7
	$M_A^{1.5+}$, $M_B^{1.5+}$	-4.5	-3.8
	$M_A^{1.25+}$, $M_B^{1.25+}$	-4.9	-4.1
	M_A^{1+}, M_B^{1+}	-5.3	-4.4
	$M_A^{0.75+}$, $M_B^{0.75+}$	-5.8	-5.2
	$M_A^{0.5+}$, $M_B^{0.5+}$	-6.2	-5.6
	$M_A^{0.25+}$, $M_B^{0.25+}$	-6.3	-6.
	Mg_A^0, Mg_B^0	-7.1	-6.7
	M_A^{2+}	-6.6	-5.7
	M_B^{2+}	-4.4	-4
	No metals	-7.3	-6.7
***In vitro* inhibition (%)**		74.9	97.9

[1]2D structures of the corresponding ligands were generated using the online Smi2Depict tool from ChemDB Chemoinformatics Portal (Chen *et al*., 2007b).

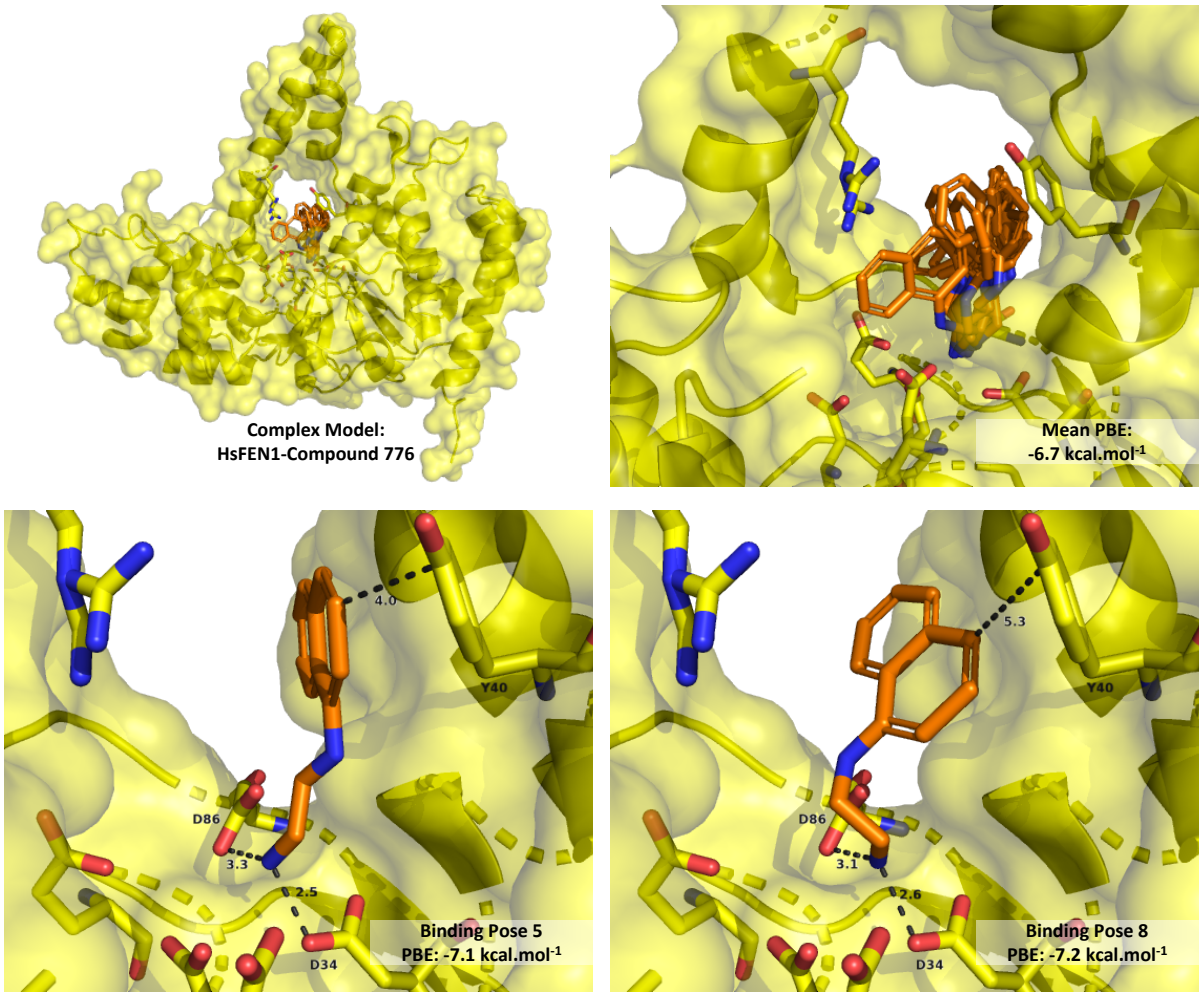


Figure 3.17 Models of HsFEN1-Compound 776 complex. (top left, top right) Open- and close-up views of top 10 binding poses distribution of MaybridgeRO3 inhibitor 776 along with their respective mean predicted binding energy (PBE). (bottom left, bottom right) Models of active site and compound 776 linkage at the top two binding poses with the lowest PBE, as these have been predicted by AutoDock4 in the absence of both active site metals. Structure of HsFEN1 is shown in yellow cartoon, with selected key residues shown in sticks. Active site metals are shown as dark grey spheres, while compound 776 is shown in orange. Figures have been generated using the modified 5K97 PDB file, used as the target macromolecule for molecular docking, and rendered by PyMOL™ .

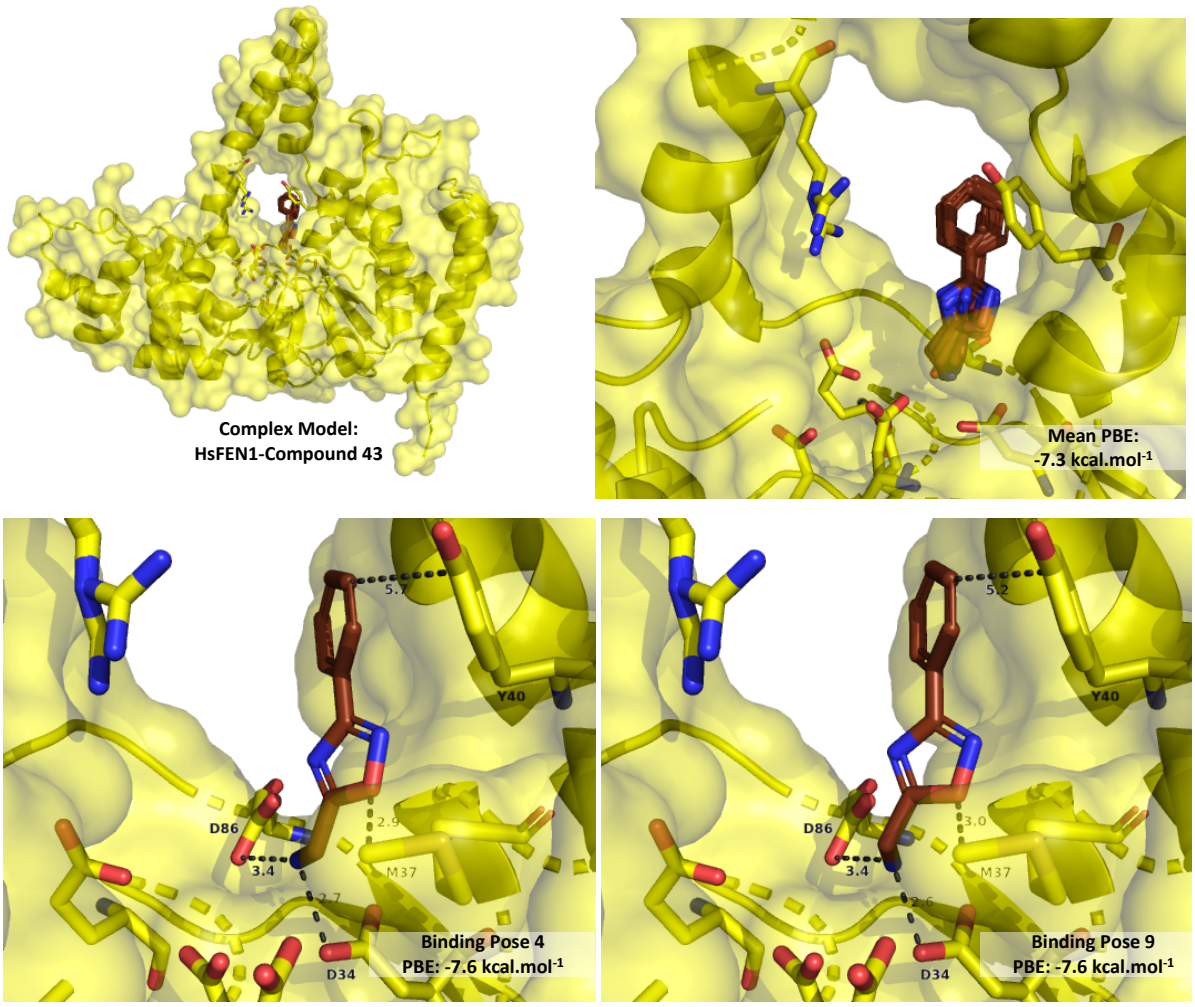


Figure 3.18 Models of HsFEN1-Compound 43 complex. (top left, top right) Open- and close-up views of top 10 binding poses distribution of MaybridgeRO3 inhibitor 43 along with their respective mean predicted binding energy (PBE). (bottom left, bottom right) Models of active site and compound 43 linkage at the top two binding poses with the lowest PBE, as these have been predicted by AutoDock4 in the absence of both active site metals. Structure of HsFEN1 is shown in yellow cartoon, with selected key residues shown in sticks. Active site metals are shown as dark grey spheres, while compound 43 is shown in brown. Figures have been generated using the modified 5K97 PDB file, used as the target macromolecule for molecular docking, and rendered by PyMOL™ .

Careful examination of the intermolecular interactions, likely to occur between the individual inhibitors and HsFEN1, is crucial to understand their inhibition mechanism and identify critical interaction pairs maintained across diverse ligands. Although here preliminary interaction analysis was only performed for four non-metal and metal chelators, analysis of all 127 inhibitors identified from MaybridgeRO3 would have been a long time-consuming process. Subsequent dose-response studies will potentially shorten that list further and allow analysis of the most effective HsFEN1 inhibitors based on their inhibitory concentrations.

3.4 Screening of hit expansion (HE) compounds to identify HsFEN1 inhibitors

3.4.1 Selection and properties of hit expansion compounds set

Following a collaboration between the Sayers' laboratory, Dr Gian Marco Ghiandoni and Prof Valerie J. Gillet (Information School, University of Sheffield), a set of 23 compounds were identified *in silico* as potential inhibitors against a range of FEN1 proteins. Based on the *in vitro* inhibition of top scoring MaybridgeRO3 compounds, Dr Gian Marco Ghiandoni's computational approach on the development of an *in silico* correlation model for the identification of potential HsFEN1 inhibitors, resulted in a set of 23 compounds.

Since this set of compounds was selected for *in vitro* testing against HsFEN1 based on quantitative data from the MaybridgeRO3 screen, similar structural characteristics are expected to be seen. Indeed, preliminary analysis of the 2D structures from that pilot hit expansion round showed not only common functional groups, but also conserved structural cores with some of the highest scoring MaybridgeRO3 top scoring compounds (Section 3.2.5; Table 3.4). Similar to MaybridgeRO3, the majority of these candidate inhibitors appear to have at least one carboxylate (-COOH) group in their structures (Figure 3.19). However, ligands lacking a carboxylate (-COOH) group appear to have at least one carbonyl- (>C=O), hydroxyl- (-OH) and/or amino- (-NH) group, which could support interactions with the active site, particularly the metals in the case of carboxylate compounds.

Based on the structures of these 23 hit expansion compounds, analysis of their physicochemical properties was carried out using DataWarrior 5.2 (Sander et al., 2015). Since these potential inhibitors were selected from an online database of structurally diverse and commercially available chemical compounds, their physicochemical properties would not necessarily comply with the "Rules of three/five" (Lipinski and Hopkins, 2004; Lipinski *et al*., 2001). Unlike MaybridgeRO3, which has been designed in complete compliance with the "Rule of Three", calculation of some major "druglike" properties of the hit expansion compounds revealed varying values beyond Lipinski's rules range (Table 3.11). Both the MW and hydrophilicity of each hit expansion compound were found to lie within the "Rule of Three" range, according to which the MW must be ≤300 Da and hydrophilicity ≤3. The number of H-bond acceptors however, exceeded the maximum number of three, but seemed to be in compliance with the original "Rule of Five" (Lipinski *et al*., 2001). In both cases, the guidelines set by the "Rule of Three" and "Rule of Five" aim to single out non-druglike from druglike molecules, which are likely to exhibit poor absorption and permeability at a later stage. However, this is not necessarily a requirement for early hit expansion screening, since downstream design processes can improve the performance and ADME/Tox properties of a candidate molecule. Hit expansion molecules were generally found to have druglike characteristics, with the exception of the susceptible to the oxidative action of either aldehyde oxidase or potentially air, aldehydic compound HE_17. Because of that, development examples of the prone to oxidation aldehydic drug candidates in literature are associated with increased toxicity and attrition rates (e.g. BIBX1382; Epidermal growth factor receptor (EGFR) inhibitor) (Dittrich *et al*., 2002; Hutzler

et al., 2013; Sanoh *et al*., 2015). In addition, aldehydes are reactive with the amino groups of lysine residues eliminating water and forming the unique to aldehydes products known a "Schiff bases" (Lys-N=C-R; Schiff base). Although "Schiff bases" are generally considered as excellent and easy to synthesize metal chelators, such products have also been associated with adverse toxic effects and immunological responses (Bootorabi *et al*., 2008).

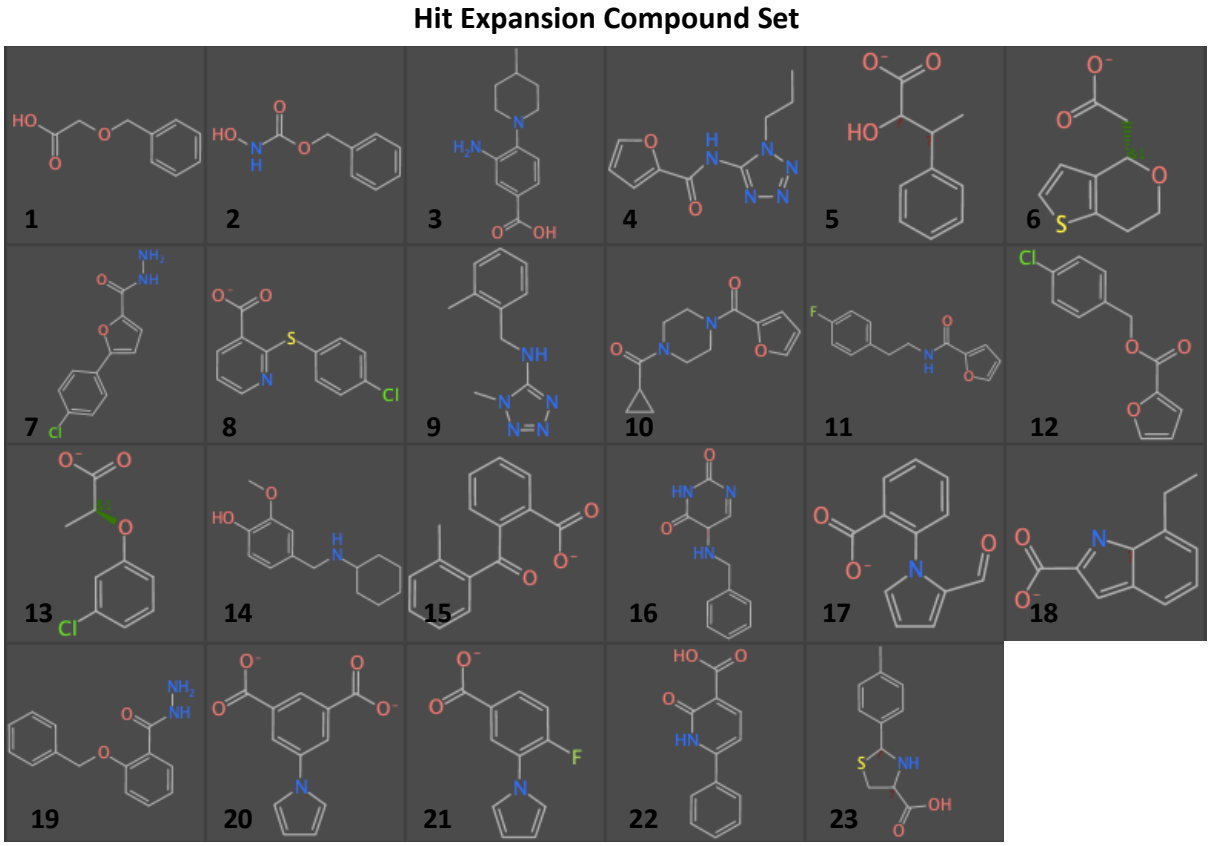


Figure 3.19 Structures of the hit expansion compound set. Numbers 1-23 are representative of MolPort_1 to MolPort_23 ligands, respectively. 2D structures were output using DataWarrior 5.2. open-source software (Sander *et al*., 2015).

Table 3.11 Physicochemical properties of hit expansion compound set

Hit expansion (HE) compound no.	Molecular weight (MW, Da)	Hydrophilicity (cLogP)	Solubility (cLogS)	H-Acceptors	H-Donors	Relative polar surface area (PSA)
1	167.16	0.8078	-2.372	3	1	-0.35903
2	266.73	2.6683	-3.119	4	2	0.3004
3	231.21	0.8147	-3.955	4	2	0.34708
4	215.21	1.3169	-4.29	7	1	0.27764
5	205.19	1.4304	-4.256	3	1	0.21622
6	166.18	0.6202	-1.403	3	0	0.26794
7	198.24	1.1479	-1.803	4	2	0.39351
8	215.21	0.763	-1.788	3	0	0.31165
9	236.65	2.7854	-3.498	5	1	0.2068
10	233.24	2.2361	-3.052	5	0	0.20782
11	248.28	1.0068	-1.308	3	1	0.25274
12	217.23	-0.1692	-2.498	3	0	0.35965
13	243.26	0.7447	-2.265	3	2	0.24323
14	236.66	1.53	-4.178	3	0	0.30638
15	242.28	1.3309	-3.323	3	2	0.25587
16	224.30	-1.1001	-2.178	5	1	0.35135
17	204.26	1.2303	-2.945	4	0	0.31683
18	240.26	2.6186	-4.073	4	0	0.20937
19	180.20	1.0029	-1.833	3	2	0.27836
20	236.33	2.156	-2.788	4	0	0.19311
21	189.21	1.9976	-2.681	5	0	0.27198
22	222.23	0.8678	-2.687	3	2	0.45178
23	234.30	1.5697	-2.857	4	2	0.24638

3.4.2 *In vitro* evaluation of hit expansion compound set using a FRET-based assay

Aiming to evaluate the potential inhibitory effect of the 23 HE compounds, FRET-based HTS assay was carried out in the presence of 0.5 μg/ml HsFEN1 per reaction. Following the established FRET-based assay, each compound was tested at 500 μM concentration alongside a negative/no enzyme (n = 8) and positive/no inhibitor control (n = 8). With the assay's standard deviation at 7.1%, the theoretical hit limit for the identification of HsFEN1 inhibitors was set at 21.3% according to the rule of 3SDs (Zhang *et al*., 1999). Although the majority of the tested compounds led to some level of decrease in the detected HsFEN1 activity, the inhibitory effect was classified as significant (≥21.3%) only for the 13 out of 23 (56.5%) compounds tested. Among the 13 compounds with percentage inhibition above the defined threshold, four with percentage inhibition >40% stood out. Out of these, HE_20 and HE_21 compounds were shown to completely abolish the activity of HsFEN1 with 98.8% and 106.8%

percentage inhibition, respectively (Figure 3.20). The top four compounds were selected for follow up interaction and dose-response analysis to determine their efficacy and potential as leads for HsFEN1 inhibition.

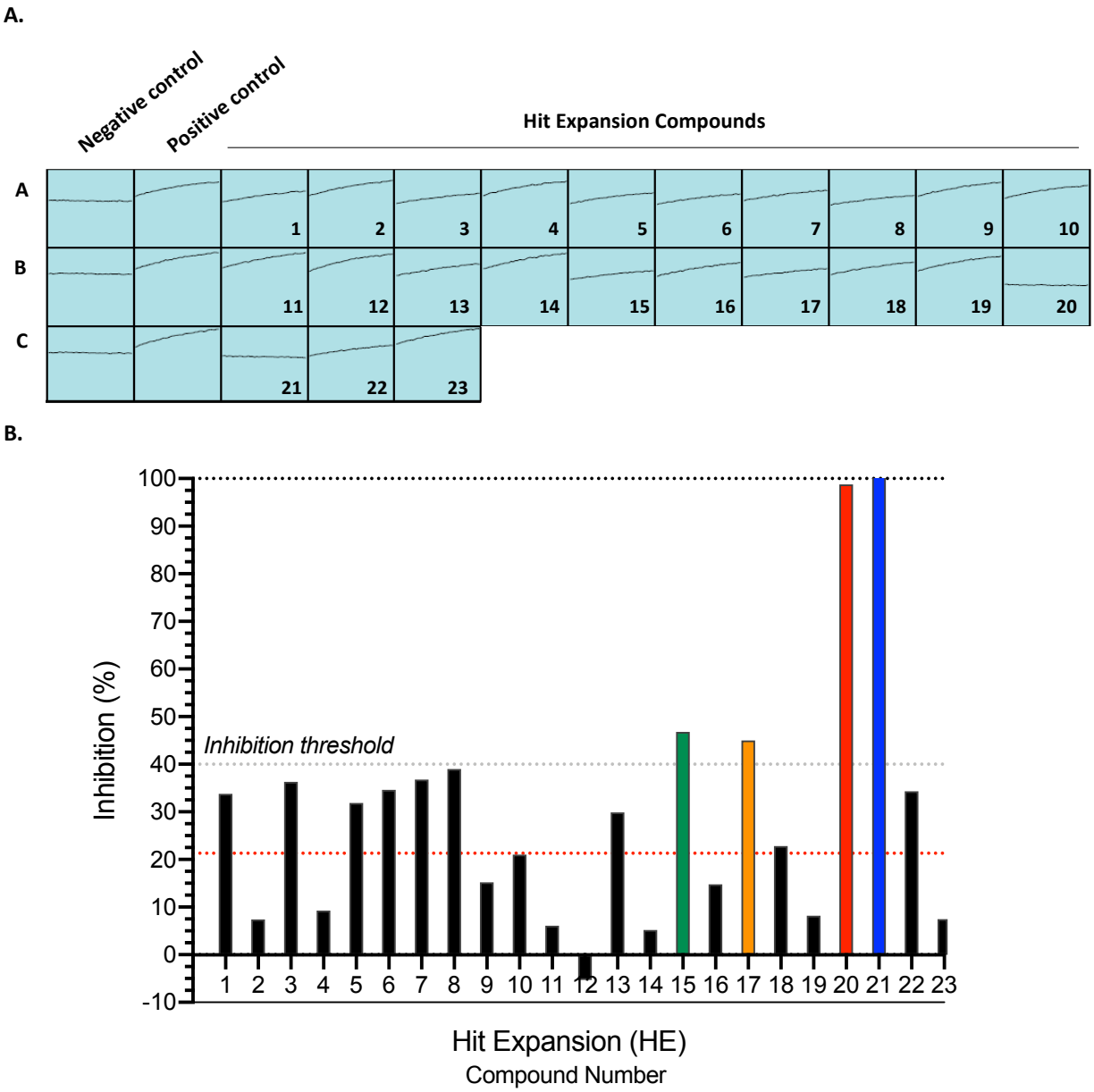


Figure 3.20 *In vitro* evaluation of hit expansion compound set for potential inhibition of HsFEN1. (A) Enzyme-catalysed reaction progress curves monitored by fluorescence change at in the presence of candidate inhibitors at 500 μM concentration. Negative and positive controls correspond to vehicle and enzyme only/no inhibitor controls (N=1, n=8). (B) Percentage inhibition of HsFEN1, calculated based on the formula Inhibition (%) = 100-[(RFU_{sample}-$\bar{x}_{NEG}$)/($\bar{x}_{POS}$-$\bar{x}_{NEG}$)]*100, where $\bar{x}_{POS}$ and $\bar{x}_{NEG}$ are the mean RFU of the positive and negative controls, respectively. Top four compounds with inhibition over the 40% of HsFEN1 activity are coloured. Red dotted line is representative of the assay 3SDs with an overall Z' value, indicative of assay quality, at 0.72.

3.4.3 Dose-response analysis of HE hit compounds

Compounds HE_20 and HE_21 identified as hits, based on the rule 3SDs, were further examined for their inhibition efficacies. In addition to these, compounds HE_15 and the aldehydic HE_17 were also examined for comparison, given that they appeared to have the highest percentage inhibition after HE_20 and HE_21 (Section 3.4.2). Dose-response FRET assays were carried out to estimate the IC_{50} values for each compound. Eight different concentrations from 0.2-500 μM for HE_20 and HE_21 and 1-2500 μM for HE_15 and HE_17 were examined over two independent experiments (*N*=2, *n*=2-3). Compounds HE_15 and HE_17 with IC_{50}>1000 μM were found to be less potent from HE_20 and HE_21, consistent with what was previously suggested by the percentage inhibition data (Figure 3.21). On the other hand, HE_20 with an IC_{50} of ~2±0.09 μM, which is 32-fold lower than that of HE_21 (62.6±3 μM), was identified as the most potent from the HE set of compounds (Figure 3.21). Given that only one of the concentration points used was found to lie in the slope between the two plateaus of HE_20 sigmoidal curve, dose-response assays were repeated at a lower concentration range (0.08-10 μM) for a more accurate determination of compound's HE_20 IC_{50} value. These led to an IC_{50} of 1.7±0.06 μM (Figure 3.21), which to our knowledge is the lowest among HsFEN1 inhibitors derived from computational/virtual screening approaches, with the only exception of NSC281680 inhibitor (1,2,5,6-Tetrazocane; IC_{50} of ~1.2 μM) (Deshmukh *et al*., 2017) (Panda *et al*., 2009). The identification of HE_20 inhibitor expands the current range of reported HsFEN1 inhibitors and may facilitate future inhibitor development studies.

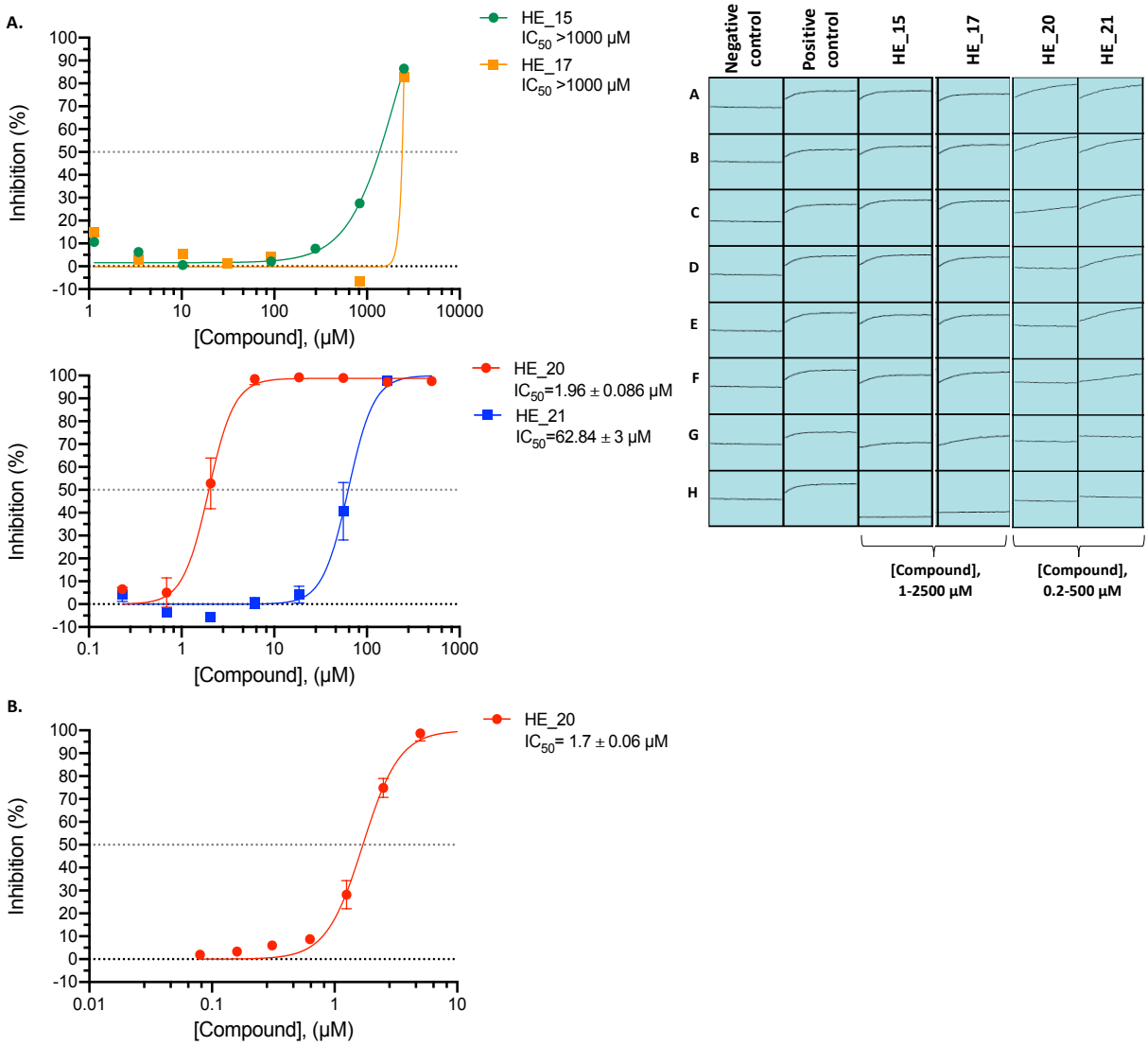


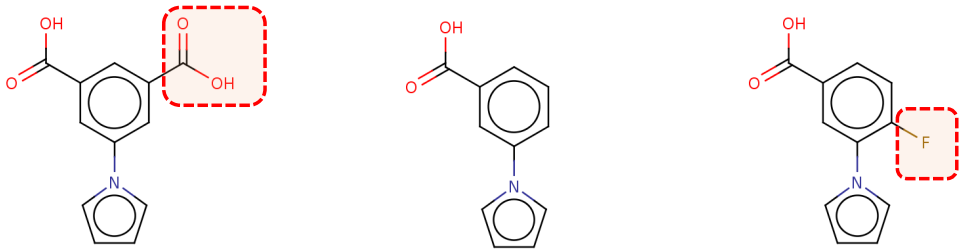
Figure 3.21 Dose-response curves of HE_15, HE_17, HE_20 and HE_21 compounds. (A) Dose-response curves and IC_{50} values of HE_15, HE_17, HE_20 and HE_21 over a 20 min measurement period (*N*=2, *n*=2-3), along with their respective catalysed reaction progress curves. (C) Dose-response curves and IC_{50} value of HE_20 ligand calculated over a 10 min measurement period at a lower concentrations range (*N*=2, *n*=3). Data were analysed using *nonlinear regression [Inhibitor] vs. response -Variable slope (four parameters)* model in GraphPad Prism 8.4. *Z′* values, indicative of the assay quality, were maintained >0.5 for each experiment carried out.

3.4.4 Modelling of HsFEN1-HE_20 and HsFEN1-HE_21 complexes

Aiming to explore further the intermolecular interactions of HsFEN1-ligand complexes that might be responsible for the 32-fold decrease in IC_{50} observed in HE_20 ligand, virtual screening of the two ligands was performed increasing the maximum number of energy evaluations to 25×10^6, instead of the standard 25×10^3. Using the optimised HsFEN1 docking model, prediction of the energetically favourable binding poses of HE_20 and HE_21 ligands

was carried out in AutoDock. Docked HE_20 and HE_21 ligands were initially processed in Open Babel (O'Boyle *et al*., 2011). For the resulting top 10 poses, the average binding energies predicted were -11.8 kcal/mol for HE_20 and -10.4 kcal/mol for HE_21 with <2 Å RMSD values. As expected, the lower average binding energy predicted for HE_20 suggested a more potent HsFEN1 inhibitor, which was in accordance to the *in vitro* dose-response data.

HE_20 and HE_21 ligands share a 3-(1H-pyrrol-1-yl)benzoic acid as their common structural core (Figure 3.22) Their only difference lies in the presence and the position of an additional functional group on the benzene ring. The HE_20 contains a second -COOH group attached to the benzoic acid ring at a *meta*- position, whereas HE_21 contains a fluorine group attached to the benzoic acid ring at a *para*- position (Figure 3.22). However, due to the presence of at least one -COOH group in each ligand, both compounds were predicted to coordinate the active site metals in a similar manner (Figure 3.23). Therefore, the difference between the inhibition potency of the two ligands is likely to rely on interactions formed between the extra -COOH/-F group and HsFEN1.



(a) HE_20 (b) Common core (c) HE_21

Figure 3.22. Structural similarity between HE_20 and HE_21 ligands. Chemical structures of HE_20 (a), HE_21 (c) and their structural core (b) were generated using the online Smi2Depict tool from ChemDB Chemoinformatics Portal (Chen *et al*., 2007b). Faint red boxes highlight main structural difference between the two ligands, relying on the type and position of an additional functional group on benzene ring.

Docking suggested that the second -COOH group of HE_20 could form additional charge-charge interactions with the long side chain of Arg100 (Figure 3.23). Donor-acceptor distances were found to be ~2 Å. Due to the conformational flexibility of Arg100 this distance is likely to be even shorter in reality, resulting in even stronger H-bonds. Fluorine in HE_21 on the other hand, was not found to form short contacts with acidic hydrogens, such as -OH and -NH (Figure 3.23). Such contacts with fluorine (<3 Å), are generally considered extremely rare despite the increased electronegativity of fluorine (Howard *et al*., 1996). Therefore, the presence of a second -COOH group in HE_20, and its absence in HE_21, appears to be critical for the generation of strong H-bonds at the bottom of the helical arch, responsible for the observed increase in inhibition potency.

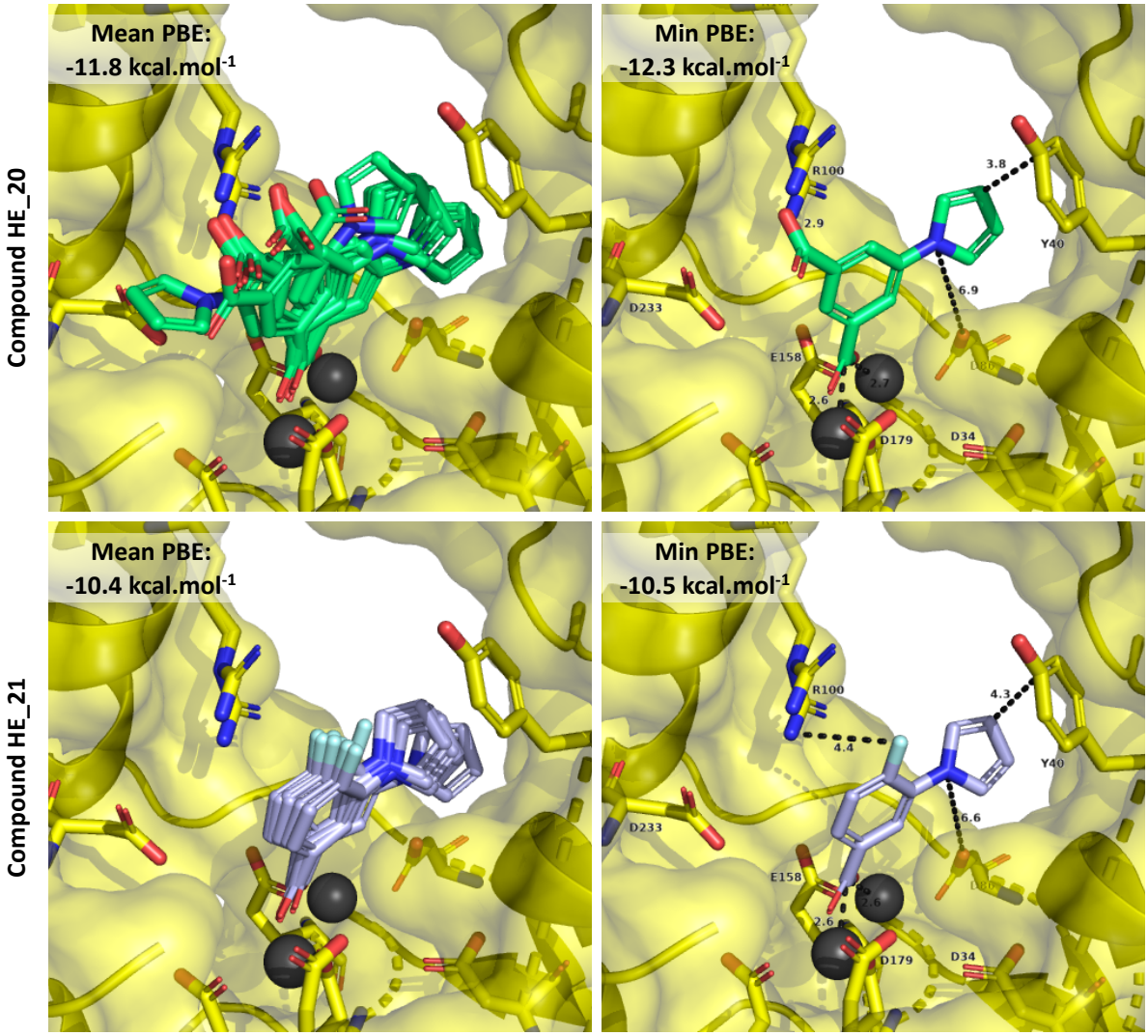
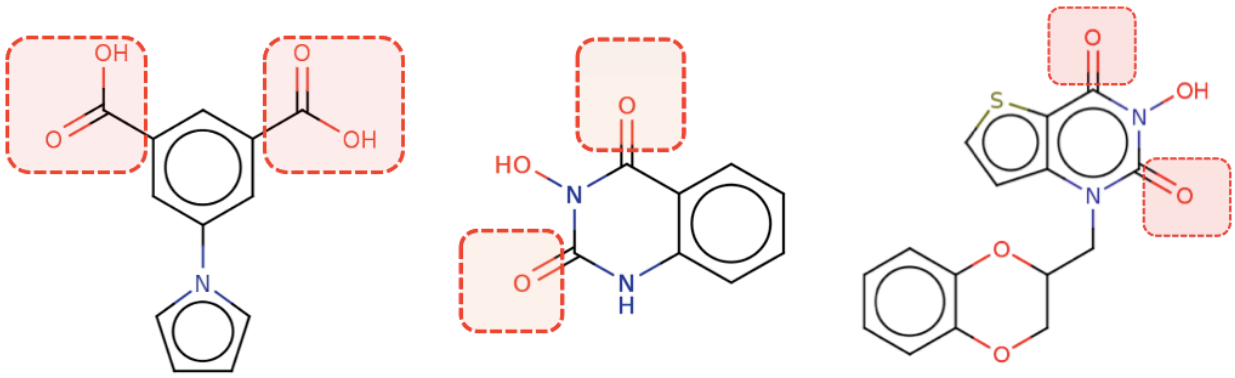


Figure 3.23. Models of HE_20- and HE_21-HsFEN1 complexes. Top 10 binding poses of HE_20 and HE_21 inhibitors within HsFEN1 active site. (Top and bottom left) Close-up views of top 10 binding poses distribution of HE_20 and HE_21 along with their respective mean predicted binding energy (PBE). (Top and bottom left right) Models of active site and inhibitor linkage at the top binding pose with the lowest PBE, as these have been predicted by AutoDock4 in the presence of both active site metals. Structure of HsFEN1 is shown in yellow cartoon, with selected key residues shown in sticks. Active site metals are shown as dark grey spheres, while HE_20 and HE_21 inhibitors are shown in green and light blue, respectively. Figures have been generated using the modified 5K97 PDB file, used as the target macromolecule for molecular docking, and rendered by PyMOL™ .

3.4.5 Comparative analysis between HE_20 and *N*-hydroxyurea inhibitors

In an effort to understand the possible inhibition mechanism of HE_20, a comparative analysis using available structural and biochemical data on the established *N*-hydroxyurea inhibitors was performed. Comparison between their chemical structures, revealed the possibility of a common metal coordination-mediated inhibition mechanism (Figure 3.24; Exell *et al*., 2016).

N-hydroxyurea inhibitors are known metal chelators, in which the presence of two carbonyl groups in proximity allows coordination of both active site metals in HsFEN1, with 4.4 Å distance between them (Exell *et al*., 2016). Overall, the inhibition potencies of the developed *N*-hydroxyurea inhibitors are higher than that of raw HE_20, derived from a pilot hit expansion study (Exell *et al*., 2016; Tumey *et al*., 2005). Two of the known *N*-hydroxyurea inhibitors were also tested for their IC_{50} values, using our own FRET-based assay, confirming the increased potency over HE_20. *N*-hydroxyurea inhibitors SN1125958263/ AZ13153160 and SN1049696325/ AZ13623940, provided from AstraZeneca, were found to have IC_{50} of 162 nM and 34.6 nM, respectively (Figure 3.25).



(a) HE_20
5-(1H-pyrrol-1yl)benzene-1,3-dicarboxylic acid)

(b) *N*-hydroxyurea
3-hydroxy-1H-quinazoline-2,4-dione)

(c) Crystallised *N*-hydroxyurea
1-[(2,3-dihydro-1,4-benzodioxin-2-yl)methyl]-3-3-hydroxy-1H,2H,3H,4H-thieno[3,2-d]pyrimidine-2,4-dione

Figure 3.24 Chemical structures of HE_20 and *N*-hydroxyurea inhibitors. 2D structures were generated using the online Smi2Depict tool from ChemDB Chemoinformatics Portal (Chen *et al*., 2007b). Faint red boxes highlight similarities between the two ligands. Ligand (b) represents only one example of the known *N*-hydroxyurea class of HsFEN1 inhibitors.

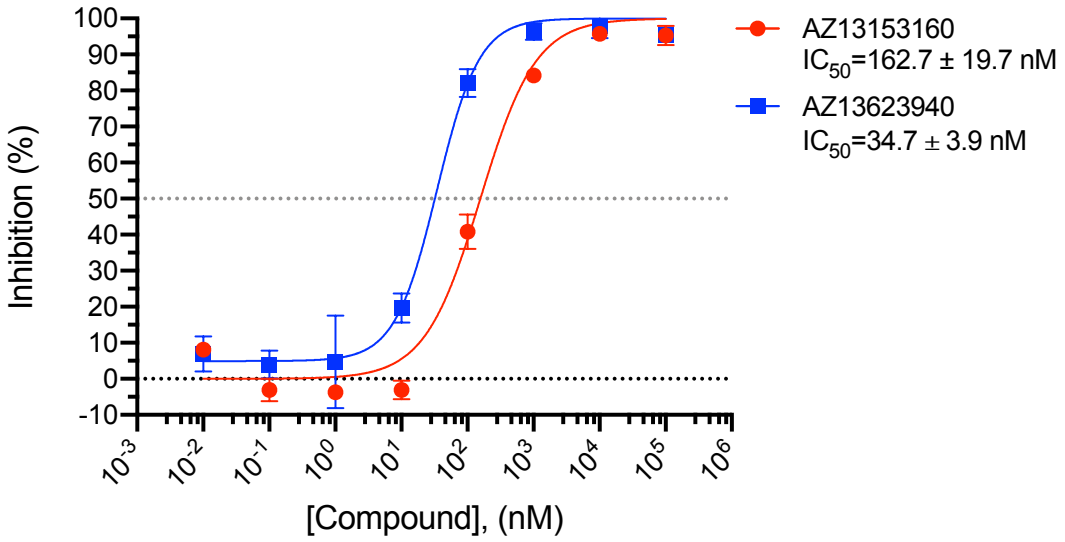


Figure 3.25. Dose-response curves of SN1125958263/ AZ13153160 and SN1049696325/ AZ13623940 compounds. (A) Data were analysed using *nonlinear regression [Inhibitor] vs. response -Variable slope (four parameters)* model in GraphPad Prism 8.4. *Z'* value, indicative of the assay quality, was maintained >0.5 (*N*=1, *n*=3).

Generally, the *N*-hydroxyurea inhibitors, are molecules with larger PSA and more H-acceptors/donors in their structures than HE_20 (Table 3.12). These might explain the observed difference in potency and can be particularly useful for future development studies of HE_20 inhibitor or even selection of compounds for additional screening. Analysis of the intermolecular interactions between HsFEN1 and an *N*-hydroxyurea inhibitor, using docking simulations and the only structure currently available, indicated that the -OH group presented between the two carbonyls in the ionized state was able to form double bridging interactions with the active site metals (Figure 3.26; Exell *et al*., 2016). Such interactions, formed at 2.3-2.6 Å (-O-Mg) distance are crucial for HsFEN1 inhibition and expected to increase potency.

Docking of the crystallized *N*-hydroxyurea inhibitor against HsFEN1 with an IC_{50} value of 0.05 μM, about 34-fold lower than that of HE_20, led to predicted binding energies analogous to those predicted for HE_20 top poses (Table 3.12). Based on the HsFEN1:HE_20 model and HsFEN1:*N*-hydroxyurea structure, common interactions between the two molecules were identified, in addition to their conserved metal-coordination ability (PDB ID:5FV7; Exell *et al*., 2016). Whilst the distance between the two active site metals differed for the used HsFEN1:inhibitor model (4.1 Å) and structure (4.4 Å), both HE_20 and *N*-hydroxyurea inhibitor were shown to coordinate the two active site metals. In addition, the thiophene ring, in the case of *N*-hydroxyurea, and pyrrole ring, in the case of HE_20 inhibitor, were oriented towards the same hydrophobic pocket formed by Met37, Tyr40 and Val133 (Exell *et al*., 2016). The favourable contacts between Met37 sulphur and the 2,4-dione ring of *N*-hydroxyurea inhibitor, are also favourable with the pyrrole ring of HE_20 inhibitor, both distanced at ~4 Å. Hydrophobic side chain contacts are also formed with residues Met37 and Tyr40, which due to the flexibility of 2,3-dihydro-1,4-benzodioxine ring in the case of *N*-hydroxyurea inhibitor are less directional (Exell *et al*., 2016).

Due to the presence of a second carboxyl group, HE_20 inhibitor was also found to form polar interactions with Gly2 and the side chain of Arg100. Whilst Gly2 interactions also occur in the *N*-hydroxyurea inhibitor through benzodioxine ring, no contacts between Arg100 and the *N*-hydroxyurea inhibitor could be identified due to the absence of helical gateway from the crystallized HsFEN1:inhibitor structure. Superimposition with a HsFEN1 structure including the helical gateway indicated potential hydrophobic interactions between Arg100 and the 2,3-dihydro-1,4-benzodioxine ring. However, assuming that adopted inhibitor binding pose and fold in the crystallized HsFEN1:inhibitor structure is the one favourable for inhibition, the contacts between Arg100 and *N*-hydroxyurea inhibitor are not as strong as those formed with HE_20 inhibitor.

Although the predominant contribution to binding and HsFEN1 inhibition is through active-site metal coordination for the *N*-hydroxyurea inhibitors, and will more likely be for HE_20, key side chain contacts are however important for optimum inhibitor binding (Exell *et al*., 2016). Given that Arg100 is important for catalysis and post-translational regulation of HsFEN1 activity, contacts with Arg100, which have not been highlighted in previous inhibitor

studies, may be of importance to maintain and optimise in future inhibitor development studies (Algasaier *et al*., 2016; Exell *et al*., 2016; Guo *et al*., 2010; Wadhwa *et al*., 2015).

Table 3.12 PBE, IC_{50} values and physicochemical properties of HE_20 and *N*-hydroxyurea-based inhibitors

	Inhibitor[1,2]		
	HE_20	***N*-hydroxyurea**	**Crystallised *N*-hydroxyurea**
Molecular weight (MW, Da)	236.33	178.15	332.34
Hydrophilicity (cLogP)	2.16	-0.148	1.47
Solubility (cLogS)	-2.79	-2.28	-3.76
H-Acceptors	4	5	7
H-Donors	0	2	1
Relative polar surface area (PSA)	0.193	0.439	0.386
IC_{50} (μM)	1.7	-	0.05
Mean PBE (kcal/mol)	-11.8	-	-11.8
Min PBE (kcal/mol)	-12.3	-	-12.2

[1]Mean and min PBE values are derived from docking simulations using the predefined search space of Section 3.3.3 and active site metal charges of +2.

[2]IC_{50} for *N*-hydroxyurea inhibitors were extracted from Exell *et al*., (2016) and physicochemical properties calculations were carried out using DataWarrior 5.2 (Sander *et al*., 2015).

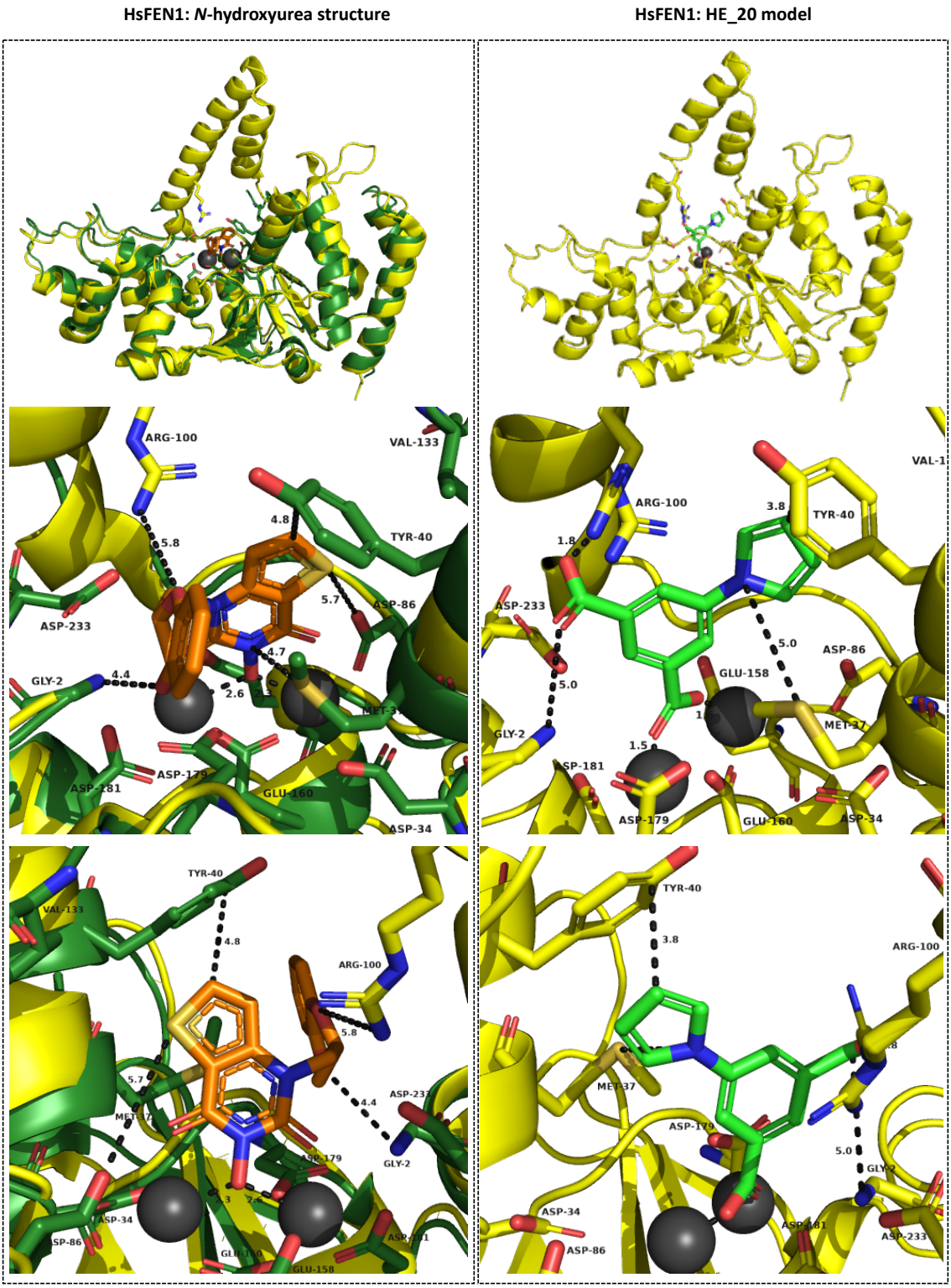


Figure 3.26 HE_20 and *N*-hydroxyurea binding to HsFEN1. (top) Open-up views of *N*-hydroxyurea (orange sticks; left) and HE_20 (green sticks; right) inhibitors bound to HsFEN1. Green cartoon corresponds to the structure of HsFEN1:*N*-hydroxyurea inhibitor complex (PDB ID:5FV7; Exell *et al*., 2016). Yellow cartoon corresponds to the superimposed HsFEN1 to facilitate interaction analysis with the helical gateway, absent from HsFEN1:*N*-hydroxyurea inhibitor complex (PDB ID: 3Q8L; Tsutakawa *et al*., 2011). (middle, bottom) Close-up vies of the occurring intermolecular interactions in HsFEN1:*N*-hydroxyurea complex structure and HsFEN1:HE_20 model (Top pose; PBE: 12.3 kcal/mol).

3.5 Screening of BIONET library for the identification of HsFEN1 inhibitors

3.5.1 Assay rationale and key features of BIONET library

With the aim to identify small molecules with increased inhibition potency over HsFEN1, the 2nd generation BIONET fragment library (Key Organics, UK) was selected for *in silico* screening and validation of the proposed +2 active site charge metals model for HsFEN1 inhibitor identification. Similar to MaybridgeRO3, BIONET is another chemically diverse fragment library in strict compliance with the aforementioned Lipinski's "Rule of Three" or the revised "Rule of Five" criteria (Lipinski and Hopkins, 2004; Lipinski *et al*., 2001). Composed of 1166 ligands with experimentally determined solubility in both DMSO and water, BIONET has been developed to overcome common issues of fragment library screening, related to false positives and high attrition rates at the later stages of drug discovery. In addition, the fact that BIONET's compound selection is composed mainly of cyclic structural motifs, found in FDA-approved drugs, increases the chances of identifying a molecule suitable for subsequent medicinal chemistry development.

3.5.2 Molecular docking of BIONET fragment library against HsFEN1

The 1166 ligands, composing the BIONET fragment library, were initially screened against HsFEN1 *in silico* for the identification of those compounds with potential inhibitory activity over HsFEN1 in subsequent *in vitro* experiments. Two docking runs were performed using AutoDock under default docking parameters. Targeting mainly metal chelators, the first docking run was carried out in the presence of both active-site metals at +2 charge (M_A^{2+}, M_B^{2+}), which was shown to possess the greatest prediction power among the *in silico* docking models tested before (Section 3.3.3). On the other hand, targeting a potentially interesting class of non-metal chelators, the second docking run was performed in complete absence of both active site metals. Analysis of the average binding energies calculated from the predicted top 10 poses of each BIONET ligand, revealed 130 ligands as potential HsFEN1 inhibitors in the presence of both M_A^{2+} and M_B^{2+} (Figure 3.27). Number of "hits" was determined using the predefined -7 kcal/mol threshold, under which docking in the absence of both active site metals resulted in only 7 hits. These represent about 0.6% of the entire BIONET fragment library, which, although remains lower than that predicted for MaybridgeRO3 (1.9%), is still a lower proportion compared to the predicted metal chelators. Of course, this depends on the intrinsic characteristics and composition of each fragment library. However, the 11.1% of BIONET ligands, predicted as potential HsFEN1 inhibitors, was found to resemble that predicted for MaybridgeRO3 (11.3%). These data suggest a potentially consistent performance of the *in silico* docking strategy, developed in Chapter 3, for screening of unbiased structurally diverse fragment libraries. What remains now is the *in vitro* validation of the top-scoring compounds (<-7 kcal/mol). This will allow identification of those fragments with increased inhibition potential against HsFEN1, and comparative analysis of the *in silico* and *in vitro* data to determine how the efficacy of the developed docking model translates for BIONET library.

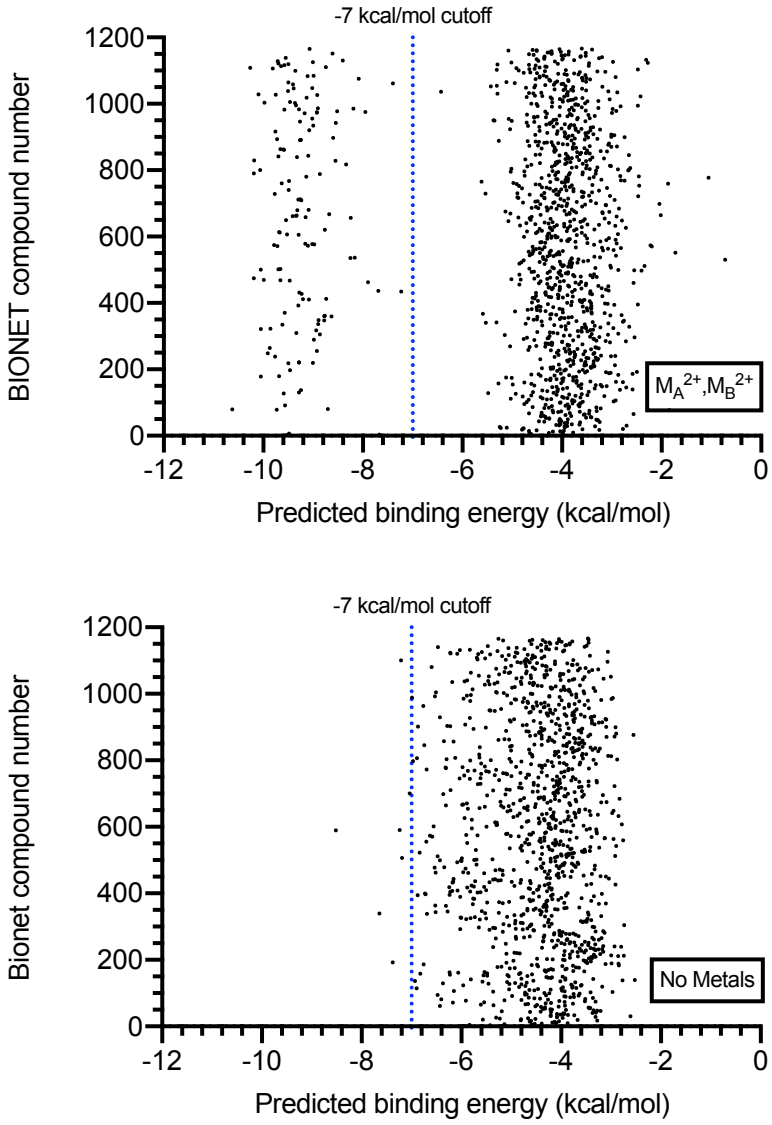


Figure 3.27 Distribution of the average binding energies predicted from the top 10 poses of each BIONET ligand. The average binding energies predicted for BIONET ligands 1-1166 were plotted using GraphPad Prism 8.2. Blue dotted lines indicate the predefined -7 kcal/mol threshold, used to identify ligands with potential inhibitory effect over HsFEN1 worth of further *in vitro* validation.

3.5.3 *In vitro* evaluation of BIONET top scored fragments using a FRET-based assay

BIONET fragments with average binding energies ≤-7 kcal/mol, predicted either in the presence or absence of both active site metals, were short-listed for validation of their *in vitro* inhibitory activity against HsFEN1. Using the optimised FRET-based HTS assay protocol for HsFEN1, each of the short-listed BIONET compounds was tested at 1 mM concentration in a 96-well format, alongside a negative/no enzyme (n = 8) and positive/no inhibitor control (n = 8). Given the limited amounts of stock compounds and following the general principle of HTS, each BIONET compound was tested as singlet (n = 1) over one single experiment (N = 1). A total of four assays was performed with their individual SD and *Z′* values, indicative of assay

quality, shown in Table 4.2. Overall, *Z′* values were maintained above the desired 0.5, while the mean SD was calculated at 8.2% (Table 3.13). Based on this and the rule of 3SDs, the hit identification limit was defined at 24.6%. About 87.6% of the *in silico* predicted hits were found to have significant inhibitory effect over HsFEN1, above the predefined 24.6% threshold. Interestingly, this is similar to the previously confirmed MaybridgeRO3 hits *in vitro* (86.8% - 99 confirmed out of the 114 hits predicted; Section 3.3.3), following the active site +2 metal charge model. With the great majority of the *in silico* predicted MaybridgeRO3 and BIONET hits having a significant inhibitory activity *in vitro*, docking against HsFEN1 active site with metal charges of +2 suggests a potentially efficient strategy for the early-stage identification of HsFEN1 inhibitors. However, without a complete *in vitro* evaluation of all 1166 molecules constituting BIONET library, it is unclear how the 130 *in silico* predicted hits relate to the overall number of true inhibitors existed in BIONET library. Although docking against HsFEN1 active site with +2 metal charges effectively predicted 78% of the total number of experimentally confirmed hits in MaybridgeRO3 library, no direct comparisons with BIONET library and the model's predictive power can be made.

Table 3.13 Quality of individual FRET-based assays during *in vitro* validation of BIONET top scoring compounds

Assay number	*Z′*	SD
1	0.52	13.3
2	0.67	7.1
3	0.79	6.2
4	0.67	7.9
Mean	0.64	8.2

[1]The overall assay quality was determined by the means of the calculated *Z′*from each of the three assay repeats. $Z' < 0.5$ and $1 > Z' \geq 0.5$, correspond to "marginal" and "excellent" assay quality, respectively.

None of the 7 predicted hits in the absence of both active site metals were found to have inhibitory activity over HsFEN1 *in vitro*. On the other hand, combination of the collected datasets for all 130 evaluated hits revealed 14 top compounds with experimental inhibition ≥ 60% for follow-up assessment (Figure 3.28). These top 14 compounds predicted as hits *in silico* in the presence of both M_A^{2+} and M_B^{2+}, suggested a potential inhibition of HsFEN1 activity through active site metal coordination. These compounds constitute 10.8% of the total number of BIONET compounds predicted as hits in the presence of both active site metals and only 1.2% of the entire BIONET library. Targeting the HsFEN1 inhibitors with the highest possible inhibitory effect over HsFEN1, the top 14 BIONET compounds were prioritized for follow-up dose-response analysis. However, future studies may also consider potency evaluation of more experimentally confirmed hits.

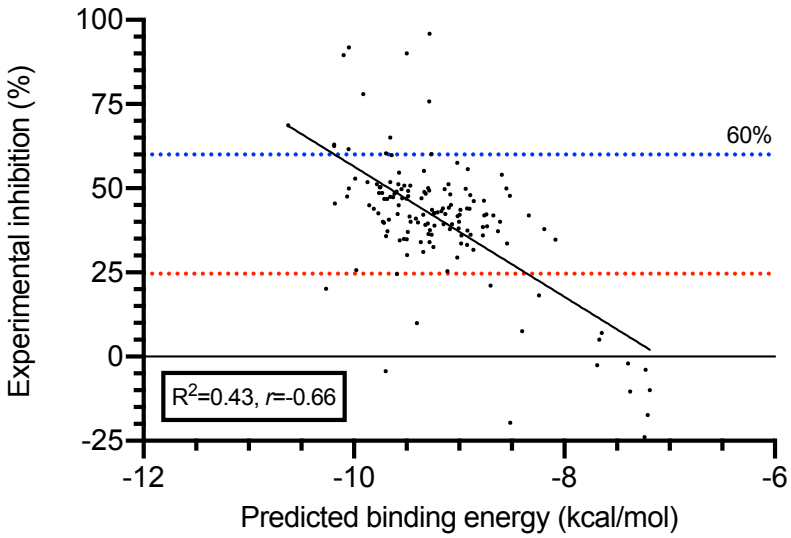


Figure 3.28 Experimental inhibition of BIONET top scoring ligands and correlation between the *in silico* and *in vitro* datasets. The predicted binding energies of BIONET top scoring ligands (≤-7 kcal/mol) were plotted against their calculated experimental inhibitions (%). Red dotted line indicates the 60% inhibition cutoff, used to short-list compounds for follow-up assays. Blue dots are representative of the top scoring ligands identified *in silico* in the absence of both active site metals, whereas black dots are representative of the top scoring ligands identified *in silico* in the presence of M_A^{2+} and M_B^{2+}. R^2 ($1 \geq R^2 \geq 0$), representative of the strength, and r ($1 \geq r \geq -1$) representative of the direction of the linear correlation between the two datasets were calculated using Bravais-Pearson analysis in GraphPad Prism 8.2.

3.5.4 Dose-response analysis of BIONET most potent inhibitors

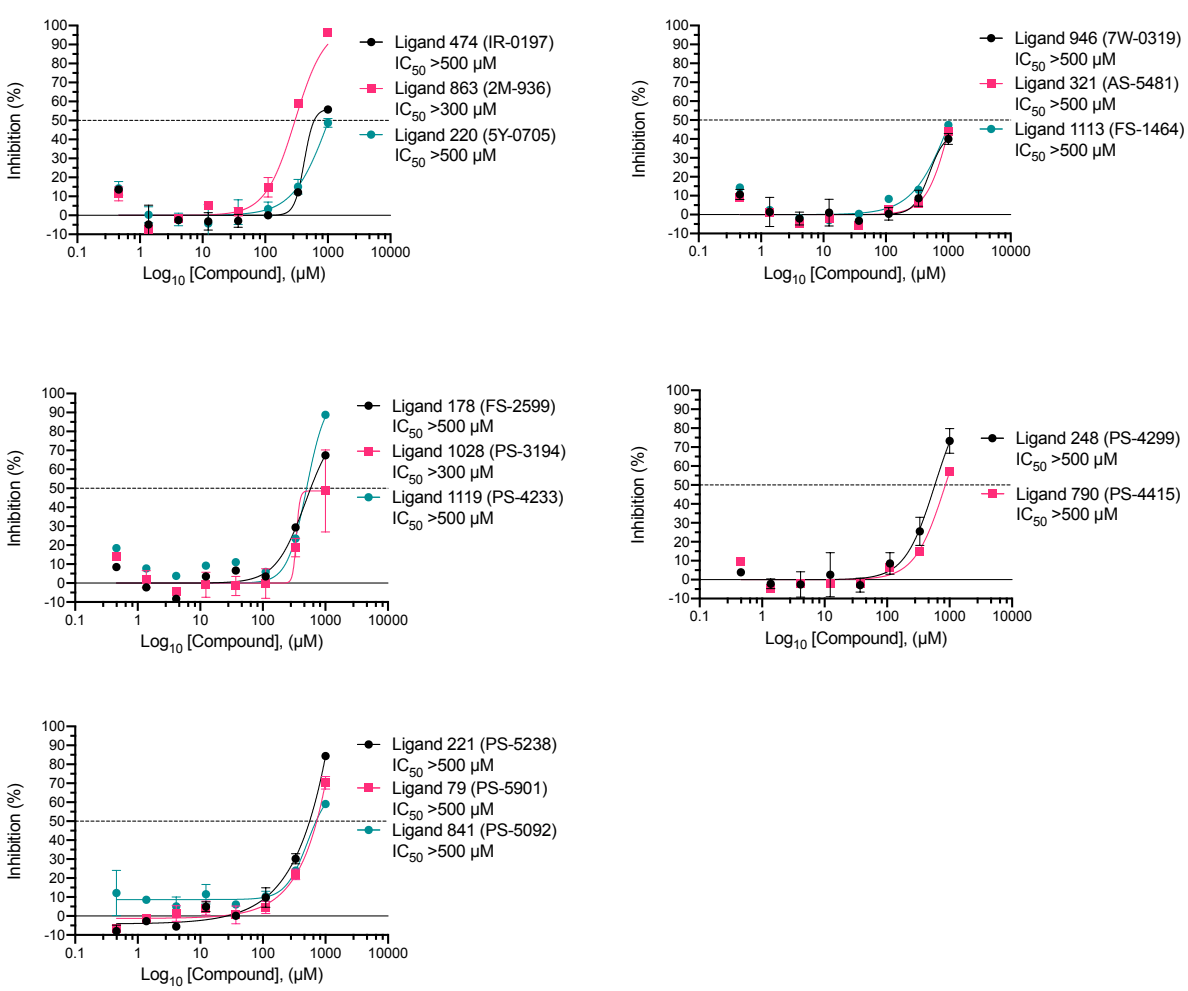
Top 14 BIONET compounds, identified with percentage inhibition ≥ 60%, were further evaluated for their half-maximal responses using the same well-established FRET assay protocol. Aiming to determine their inhibition efficacy, eight different drug concentrations ranging between 0.5 and 1000 μM were initially tested in triplicate over a single experiment (N=1, n=3). Based on the generated dose response curves, all 14 compounds were estimated to have IC_{50} values above 500 μM with ligands 863 and 1028 representing the only exemptions (Figure 3.29).

Generally, an IC_{50} cutoff of less than 1 μM is used to distinguish strong from weak inhibitors, with substantially decreased potential as drug candidates at the early stages of drug discovery (Kishimoto *et al*., 2014). Of course, this is not absolute and is often loosened in such early phases. Important information with regards to key intermolecular interactions between ligands and the protein of interest, can also be acquired from less potent ligands. Therefore, identifying and studying multiple small molecules, even with compromised potencies, can reveal critical functional motifs for the subsequent stages of drug design and development. Although the dose-response curves do not reach 100% inhibition and the IC_{50} values only represent rough estimates, further experiments to obtain more accurate IC_{50} values were not carried out. This, was not considered necessary due to the limited compound stocks, the low potential of such

weak HsFEN1 inhibitors, as well as the non-granted compound solubility at much higher concentrations.

Despite the high IC_{50} values, comparison of all top 14 BIONET compound structures were found to have one carboxylate group in their structures, indicating the importance of metal coordination for HsFEN1 inhibition. Compounds 220, 221 and 248 were found to share a common structural motif with the hit expansion compound HE_20, which however had an at least 300-fold lower IC_{50} value. The presence of a second carboxylate group in the structure of HE_20, which is not presented in those of BIONET compounds 220, 221 and 248, seems to be responsible for that massive increase in potency. On the other hand, the addition of a second nitrogen in the aromatic rings of compounds 220 and 221, which is not presented in that of HE_20, seemed to have negligible/no effect in their respective inhibition potencies.

A.



B.

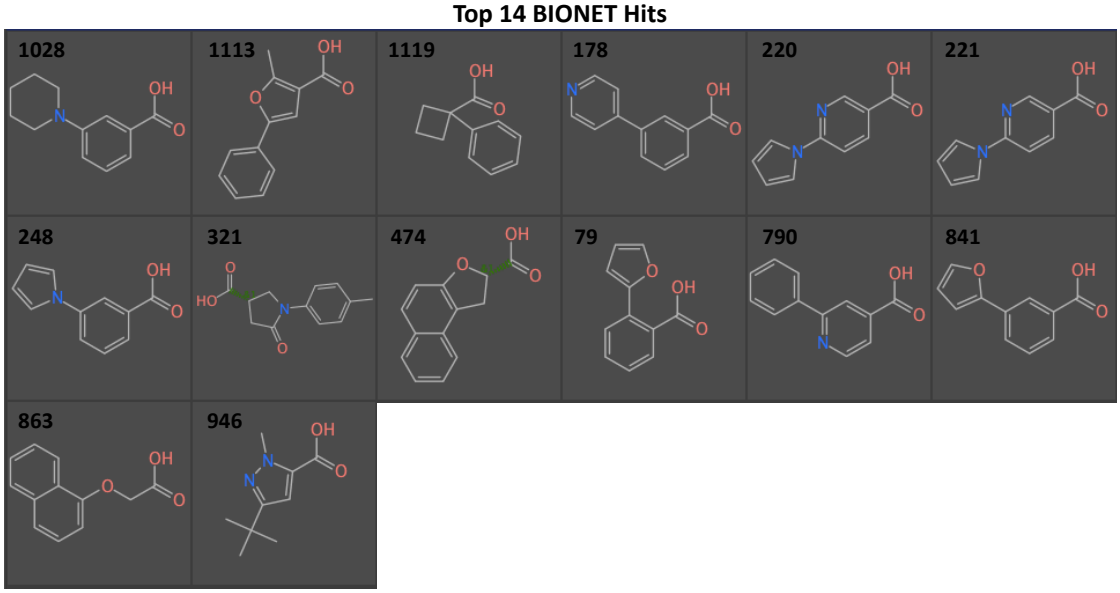


Figure 3.29 Dose-response curves and structures of top 14 BIONET compounds. (a) Data collected over a 20 min measurement in a single experiment period (N=1, n=3) were analysed using *nonlinear regression [Inhibitor] vs. response -Variable slope (four parameters)* model in GraphPad Prism 8.4. Z' values, indicative of the assay quality, were maintained ≥0.5 and IC_{50} values were estimated to lie >500 μM for the majority of BIONET top 14 compounds with the only exemptions of ligands 863 and 1028. (b) Structures of top 14 BIONET inhibitors output using DataWarrior 5.2. open-source software (Sander et al., 2015).

3.6 Conclusion

Following a dual approach of virtual and physical screening, manipulation of the active site metal charges has led to an effective *in silico* docking strategy with strong predictive power for the potential identification of HsFEN1 inhibitors. Docking of the commercially available MaybridgeRO3 ligands in the presence of +2 charge active site metals was shown to effectively predict the great majority of the *in vitro* confirmed HsFEN1 inhibitors, with a conserved ≤13% false positive rate during further model validation using BIONET library. While these data revealed a total of 127 potential HsFEN1 inhibitors, a collaborative HE study has led to a HsFEN1 inhibitor with an IC_{50} of ~1.7 μM, classified among the most potent known HsFEN1 inhibitors derived from computational/virtual screening approaches. This expands the reported to date range of HsFEN1 inhibitors facilitating future inhibitor development studies. Finally, the evaluated docking strategy not only suggests the rapid and cost-effective identification of HsFEN1 inhibitors at an early *in silico* stage, but also holds the potential to effectively be applied for the identification of microbial FEN1 inhibitors and the development of novel antibiotics.

Chapter 4 – Structural and mechanistic insights into HsFEN1 catalysis

4.1 Introduction

4.1.1 Structural studies of HsFEN1: Current knowledge

Since the establishment of HsFEN1 as a chemotherapeutic target and recognition of its inhibitors as potential drug candidates, the number of structural studies seeking to address its catalytic mechanism have dramatically increased. With only three crystal structures of HsFEN1:DNA complexes available by 2016, with the initiation of this PhD project, substrate specificity and cleavage of the ss 5′-flap with or without DNA threading through the conserved helical gateway was still enigmatic (Table 4.1).

Although these structures were a real breakthrough at the time revealing essential aspects of HsFEN1 substrate recognition and binding, such as the disordered-to-ordered transition of the highly conserved helical gateway, the absence of an energy source to support threading of the free DNA end led to further debates. Following these and the findings of AlMalki and co-workers, showing for the first time threading of the ss 5′-flap through the helical gateway of a bacteriophage HsFEN1 homolog, T5FEN, later structures of HsFEN1:DNA complexes published in 2017 had come to end the long-standing debate (AlMalki *et al*., 2016; Tsutakawa *et al*., 2017). Co-crystallisation of truncated HsFEN1 mutants with DNA substrates, containing a common double-stranded core with either a 2-, 3- or 4-nt long ss 5′-flap, have shown threading of the free DNA end through the helical gateway in a unique and extraordinary manner. Electrostatic interactions were found to drive steering of the free phosphodiester backbone and translocation of the dsDNA, which dynamically promote the upturned threading (DNA bases facing the active site metals) of the ss 5′-flap for catalysis (Tsutakawa *et al*., 2017; Chapter 1 – Introduction, Section 1.4.3). Despite the conserved DNA double-stranded core, the three complexes highlighted significant functional differences in the scissile phosphate bond positioning, Tyr40 orientation and base-pairing/unpairing, accommodating 5′-flap threading. Complete base-pairing, partial and full unpairing mechanisms were all found to be catalytically competent in D86N, R100A and D233N HsFEN1:DNA complexes, respectively. Whether the observed differences are attributed to the induced mutations and/or the substrate DNA used, remains unclear.

A year later, in 2018, three additional pre-threading structures of truncated HsFEN1 mutants with a 3-nt long flap DNA substrate were published by Xu *et al*., (2018). Focusing on HsFEN1 regulatory mechanisms, crystallisation of a catalytically active and inactive mutants analogous to methylated HsFEN1 (R192F), the primary HsFEN1 methylation site, revealed a unique loop-to-helix transition in the β-pin region (Chapter 1 – Introduction, Figure 1.10). Rigidifying of this β-pin region maintains and enhances hydrophobic interactions with the upstream duplex DNA core, as well as HsFEN1 interacting protein partners. Despite the in-line 5′-flap recognition observations, over secondary ss flaps and/or unstructured DNA, the complete absence of the structurally conserved helical gateway, cap and active-site metals in all three

structures, held back significant structural rearrangements. Strikingly, the 3-nt long flap was not oriented towards the active site and remained unprocessed, with the bases surrounding the scissile phosphate fully paired to the template strand.

Despite the existing array of HsFEN1:DNA structures and the, now, established inverted threading mechanism, the exact process of how DNA base-pairing/unpairing proceeds to accommodate cleavage of 5′-flaps at the scissile phosphodiester bond in the native WT HsFEN1 remains unclear. Given the chemotherapeutic potential of HsFEN1 inhibitors, efficient targeting of HsFEN1 could be aided through understanding of its catalytic mechanism and the intermolecular interactions at the HsFEN1 active site. Additional structures of HsFEN1:DNA complexes, perhaps with the WT HsFEN1, will enable direct comparisons to the structures of Tsutakawa *et al*., (2017) and potentially determine the occurring base-pairing/unpairing within WT HsFEN1 active site.

Table 4.1 Overview of published HsFEN1:DNA structures to date

PDB ID	WT/Mutant HsFEN1	5′-Flap length	Co-factors	Resolution (Å)	Reference
5ZOE	D181A HsFEN1	3-nt	K	1.95	(Xu *et al*., 2018)
5ZOG	R192F HsFEN1	3-nt	-	2.3	
5ZOF	D181/R192F HsFEN1	3-nt	K	2.2	
5UM9	D86N HsFEN1	3-nt	K, Sm	2.8	(Tsutakawa *et al*., 2017)
5K97	D233N HsFEN1	2-nt	K, Sm	2.1	
5KSE	R100A HsFEN1	4-nt	K, Sm	2.1	
3Q8K	WT HsFEN1	4-nt	K, Sm	2.2	(Tsutakawa *et al*., 2011)
3Q8L	WT HsFEN1	1-nt	K, Sm	2.3	
3Q8M	D181A HsFEN1	1-nt	K	2.6	

All published structures of HsFEN1 shown here have been determined using X-ray crystallography, based on the available information deposited in PDB database. HsFEN1 structures are shown in a chronological order, starting with the most recent (Date accessed: July, 2021).

4.1.2 Rationale and aims

For the identification and development of potent mechanism-based inhibitors, it is important that the HsFEN1 catalytic mechanism is fully understood. At the commencement of this project the only available crystal structures of HsFEN1-DNA complexes, involved structures of HsFEN1 with product DNA. The interactions mediating processing of the single-stranded 5′ flap were highly ill-defined and the controversy between the proposed DNA threading/tracking

models continued to exist (Tsutakawa *et al*., 2011). Aiming to investigate the catalytic mechanism of HsFEN1 and facilitate future drug design experiments, crystallisation trials of HsFEN1 complexed with multiple DNA substrates were carried out to identify those conditions allowing determination of high-resolution crystal structures. These can potentially unravel not only key interactions for DNA threading and cleavage, but also any potential conformational changes mediating catalysis, which could be targeted for selective and efficient HsFEN1 inhibition.

Following two different approaches, the full-length WT HsFEN1 and a catalytically inactive truncated D179K HsFEN1 mutant were used for crystallisation trials to test the hypothesis that DNA threads through the helical gateway for cleavage of the single-stranded 5′ flap. Similar to the unique "metal-mimic" mutagenesis approach followed by AlMalki and co-workers for crystallisation of D153K T5FEN with intact DNA substrate, the analogous truncated D179K HsFEN1, previously produced in Sayers' laboratory, was used for crystallisation (AlMalki *et al*., 2016). Substituting the positively charged ε-amino group of lysine's long side chain for a catalytically essential amino acid D153K in T5FEN had led to an inactive enzyme, able to bind DNA but not hydrolyse it. We suspected that a similar effect could be observed in the analogous D179K HsFEN1 mutant, leading to a catalytically inactive enzyme that retained DNA binding. We supposed that this inactive mutant would allow threading of the single-stranded 5′ flap and facilitate crystallisation efforts. As an alternative approach, crystal trials were also set using the native full-length WT HsFEN1. Although this is a catalytically active enzyme capable of processing the single-stranded 5′ flap, substitution of the preferred Mg^{2+} ions by Ca^{2+}, which is thought to inhibit the activity of HsFEN1, can also lead to the desired crystal structures of HsFEN1 with intact DNA substrates (Harrington and Lieber, 1994).

4.2 RESULTS & DISCUSSION: Production and purification of HsFEN1 proteins

4.2.1 Overexpression of HsFEN1 proteins by auto-induction

To facilitate production of truncated D179K and full-length WT HsFEN1, large-scale expression was carried out using Studier's well-established auto-induction method. The respective codon-optimised cDNA sequences encoding both HsFEN1 proteins, previously cloned in pET21a(+) vector, a lactose-inducible system, were expressed in *E. coli* strain BL21(DE3).

Expression strains were initially grown to saturation in non-inducing overnight shaking cultures at 37°C. MDG was typically used as the non-inducing medium, in which the presence of 0.5% (w/v) glucose prohibited any inducing activity from basal T7 RNA polymerase expression. Large-scale induction of protein expression was carried out in either 3 or 5 L shaking cultures at 22°C, using fermenter at 1:100 non-inducing to auto-inducing medium ratio.

Each auto-inducing culture was sampled at regular, timely intervals to monitor bacterial cell growth and the respective cell protein expression. Culture densities were estimated spectrophotometrically, at 600 nm, and total protein expression was determined by Coomassie-stained SDS-PAGE. Cell densities were typically as high as A_{600}≈10-12, with typical expression of the respective HsFEN1 proteins as shown in Figure 4.1. In parallel with the increased cell densities, expression of total cell proteins including the target HsFEN1 protein, was also shown to increase over time.

Induction at 22°C and the use of the commercially available Super Broth auto-induction medium led to higher cell densities than those obtained previously using "homemade" ZYM-5052 at 27°C, by Dr Sarah L. Oates ($A_{600\ nm}$≈4-6). The lower induction temperature, at which the solubility of oxygen is higher, in addition to the differing selection of auto-induction medium are likely to be the main reasons for the increased culture cell densities observed (Studier, 2005).

Auto-induction is an advantageous method enabling high density cultures to be achieved prior to the spontaneous induction of protein expression upon glucose depletion. Typically, the emerging cell densities are much higher compared to the conventional isopropyl **β**-D-1-thiogalactopyranoside (IPTG) method (Studier, 2005). Basal T7 RNA polymerase and, therefore, unintended expression of the target protein, which can often have a toxic effect on the cell host and compromise the resulting protein yield, are limited to the increased concentration of glucose used in the non-inducing media. Once glucose, the favoured energy source of *E. coli* hosts, is depleted, lactose is converted to the inducible allolactose by **β**-galactosidase. Allolactose induces expression of T7 RNA polymerase from the chromosomal lacUV5 promoter, which in turn transcribes the targeted gene under the control of T7 promoter.

In all cases, for every 1 L of auto-induction culture, about 8-10 g of cells were harvested. The resulting cells expressing strongly either the truncated D179K or full-length WT HsFEN1, were frozen at -80°C for storage and used routinely for downstream production and purification of the respective proteins.

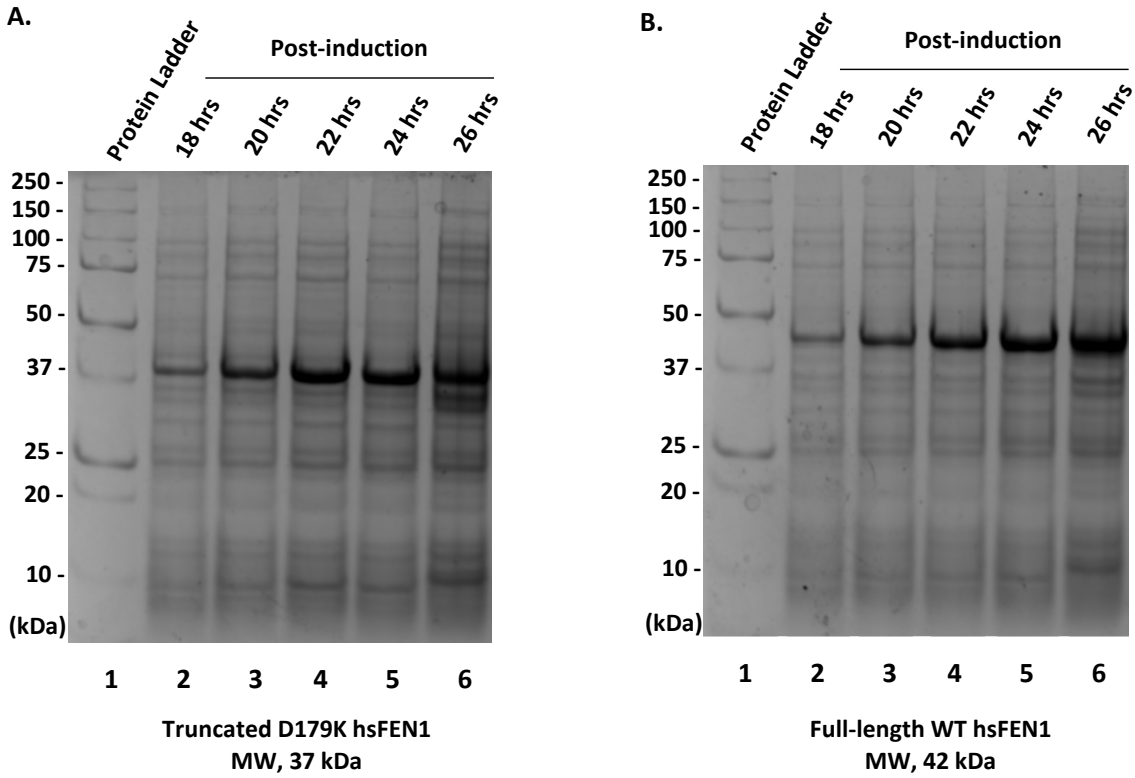


Figure 4.1 Large-scale overexpression of (a) truncated D179K and (b) full-length WT HsFEN1 proteins. Coomassie-stained 10% SDS-PAGE gels representative of the progressive overexpression of target HsFEN1 proteins in *E. coli* by auto-induction. Bands corresponding to truncated D179K and full-length WT HsFEN1 proteins are shown at ~37.7 kDa and ~42.6 kDa, respectively. Lane 1: protein ladder and Lanes 2-6: Expression of target HsFEN1 protein at different time point intervals from 18 to 26 hrs post-induction.

4.2.2 Cell lysis and extraction of HsFEN1

Depending on the required scale of protein production, which is determined by the amount of highly purified HsFEN1 proteins needed for downstream structural/activity analyses, cell pastes of 5-30 g weight were used for lysis. Cell lysis was carried out as described in Materials and Methods (Section 2.5.4-2.5.6). In all cases, the emerging soluble and insoluble fractions were analysed by SDS-PAGE electrophoresis to track the extraction and isolation process of the truncated D179K and full-length WT HsFEN1 proteins.

Both of the expressed HsFEN1 proteins were found to be soluble following lysis by lysozyme and sonication (Figure 4.2). However, minor bands, corresponding to only a small fraction of the target proteins were also found in the pellet, indicating perhaps an incomplete cell lysis or the existence of a partially insoluble form of these proteins.

Soluble fractions containing the great majority of target HsFEN1 proteins were typically purified as described in Section 2.5, which included removal of nucleic acids with PEI precipitation and ammonium sulphate cut. Although a minor fraction of both HsFEN1 proteins

is expected to be lost by the addition of these two extra steps prior to chromatographic purification, elimination of interfering nucleic acids preserves target protein integrity and can lead to increased final purities. Nucleic acids precipitation was carried out using minimal PEI concentrations, whereas close to saturation $(NH_4)_2SO_4$ concentrations were used to concentrate bulk and target HsFEN1 proteins. To recover the solubility and concentrate proteins and decrease the $(NH_4)_2SO_4$ concentration the resulting pellets were redissolved in minimal volumes of a low ionic strength buffer and dialysed prior to further purification by chromatography.

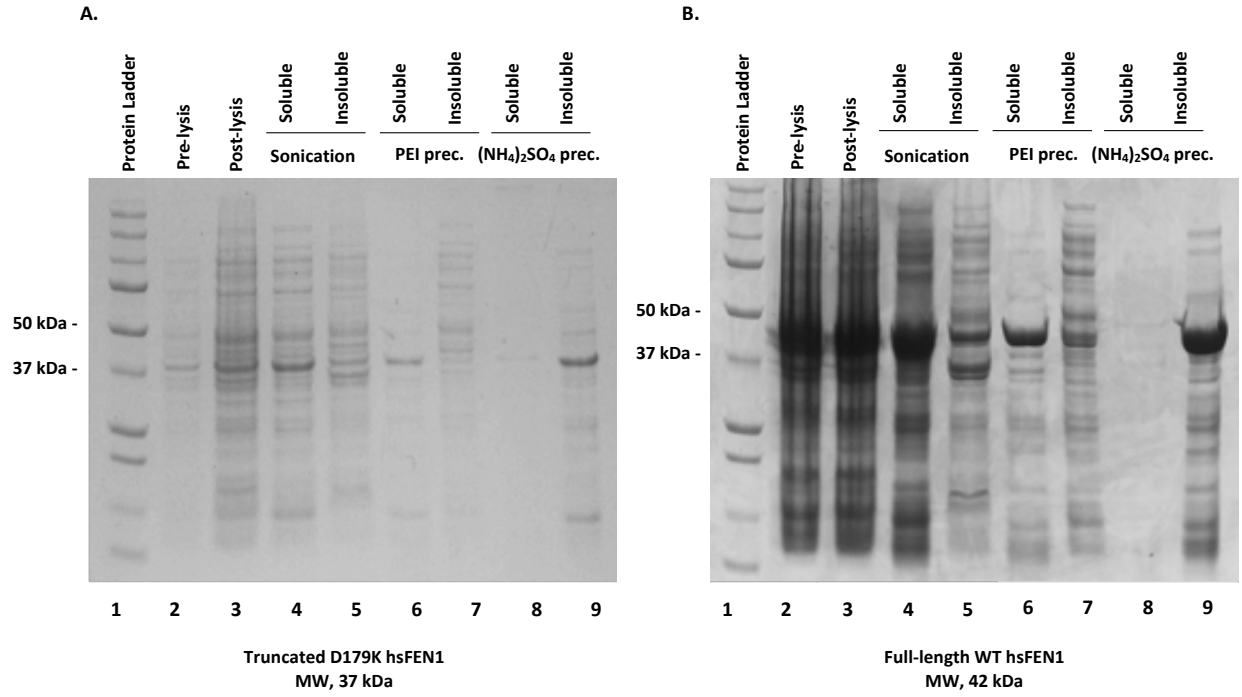


Figure 4.2 Cell lysis and extraction of (a) truncated D179K and (b) full-length WT HsFEN1 proteins. Coomassie-stained SDS-PAGE of soluble and insoluble protein fractions collected during the multi-step cell lysis and extraction process. Bands corresponding to truncated D179K and full-length WT HsFEN1 proteins are shown at ~37.7 kDa and ~42.6 kDa, respectively. Lane 1: protein ladder, Lane 2: Starting cell lysis pellet, Lane 3: Lysozyme-induced lysis, Lanes 4&5: Post-sonication soluble and insoluble fractions, Lanes 6&7: Post-PEI precipitation soluble and insoluble fraction and Lanes 8&9: Post-$(NH_4)_2SO_4$ precipitation soluble and insoluble fractions.

4.2.3 Purification of HsFEN1 proteins by affinity, ion-exchange and size-exclusion chromatography

Aiming at the production of high purity HsFEN1 proteins suitable for downstream biochemical and structural analyses, the latter of which usually requires exceptionally homogeneous protein products for successful crystallisation, extensive purification was carried out. Following dialysis, the resulting protein pools, rich in truncated D179K and full-length WT HsFEN1, respectively, were subjected to a series of affinity and ion-exchange chromatography steps.

In both cases, a heparin column was used for the first purification step, which was then followed by a combination of cation and anion exchange. A sulfopropyl (SP) column coupled to quaternary ammonium (Q) column for the purification of truncated D179K HsFEN1 and a Q column coupled to SP, for the purification of full-length WT HsFEN1 were used. Column, buffer and pH selection was adjusted based on the calculated isoelectric points (pI) of the target HsFEN1 proteins, and facilitated by previously optimised FEN1 purification protocols, described in Sayers' laboratory.

Heparin is a known cation exchanger, whose DNA-mimicking polyanionic structure often enables separation of DNA-binding proteins based on affinity, independently of their charge. SDS-PAGE electrophoresis of the collected at increasing ionic strength fractions showed recovery of both target HsFEN1 proteins and their separation from the great majority of intracellular *E. coli* proteins (Figures 4.3-4.4). Separation of the remaining contaminants was carried out using a combination of SP and Q columns for the truncated D179K HsFEN1 and Q and SP columns for the full-length WT HsFEN1. Given that the truncated D179K HsFEN1, a protein with a 6.5 theoretical pI, is expected to have a negative net charge at pH 8, the combination of SP coupled to Q column used, enables the mutant to pass through SP and be retained on Q. On the other hand, the full-length WT HsFEN1 is expected to have a positive net charge, given its 8.8 theoretical pI, requiring a combination of Q coupled to SP column, enabling it to pass through the Q column and be retained on the SP. For both target HsFEN1 proteins, increasing ionic strength conditions led to the recovery of significant target protein concentrations.

Finally, both HsFEN1 proteins were subjected to a final purification step using size-exclusion chromatography to improve sample homogeneity, which is crucial for successful crystallisation. Operational details with regards to the used purification protocols can be found in Chapter 2 – Materials and Methods (Section 2.7). Collectively, the 3-step purification processes followed for truncated D179K and full-length HsFEN1 lead to proteins of high-purity homogeneous state, shown as strong individual bands on Coomassie-stained SDS-PAGE gels, even at overloaded wells (Figures 4.3-4.4). Based on purity both HsFEN1 proteins produced through the described process appear to be suitable for downstream biochemical and structural experiments.

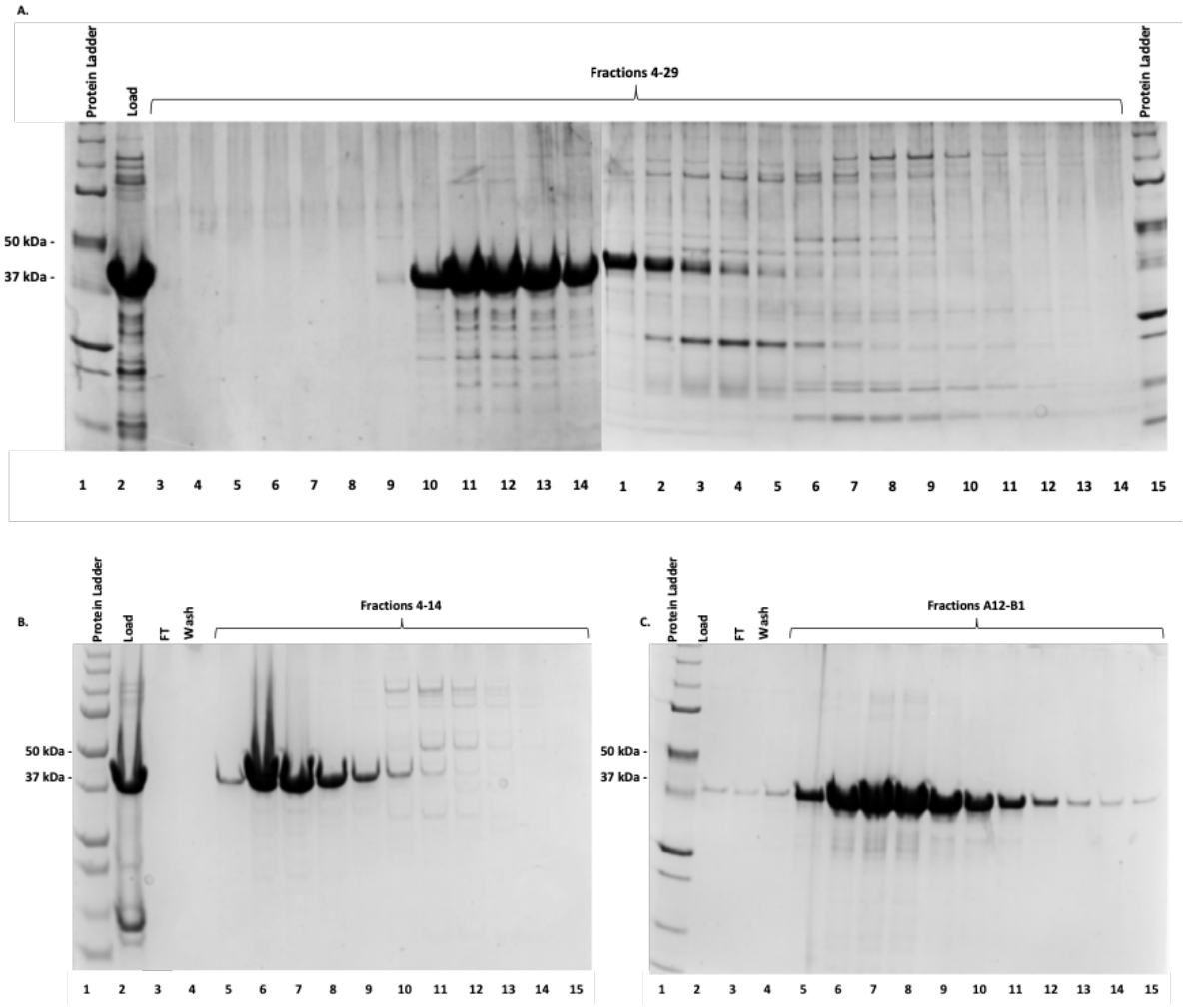


Figure 4.3 Analysis of truncated D179K HsFEN1 purification. Coomassie-stained 10% SDS-PAGE gels representative of the progressive 3-step chromatographic purification of the ~37.7 kDa truncated D179K HsFEN1. (a) 1st step: Purification of truncated D179K HsFEN1 through Heparin affinity column. (b) 2ndt step: Purification of truncated D179K HsFEN1 through the strong anion exchanger, Q column. (c) 3rd step: Final purity state of the resulting truncated D179K HsFEN1 pool from size-exclusion chromatography used for storage and downstream analysis.

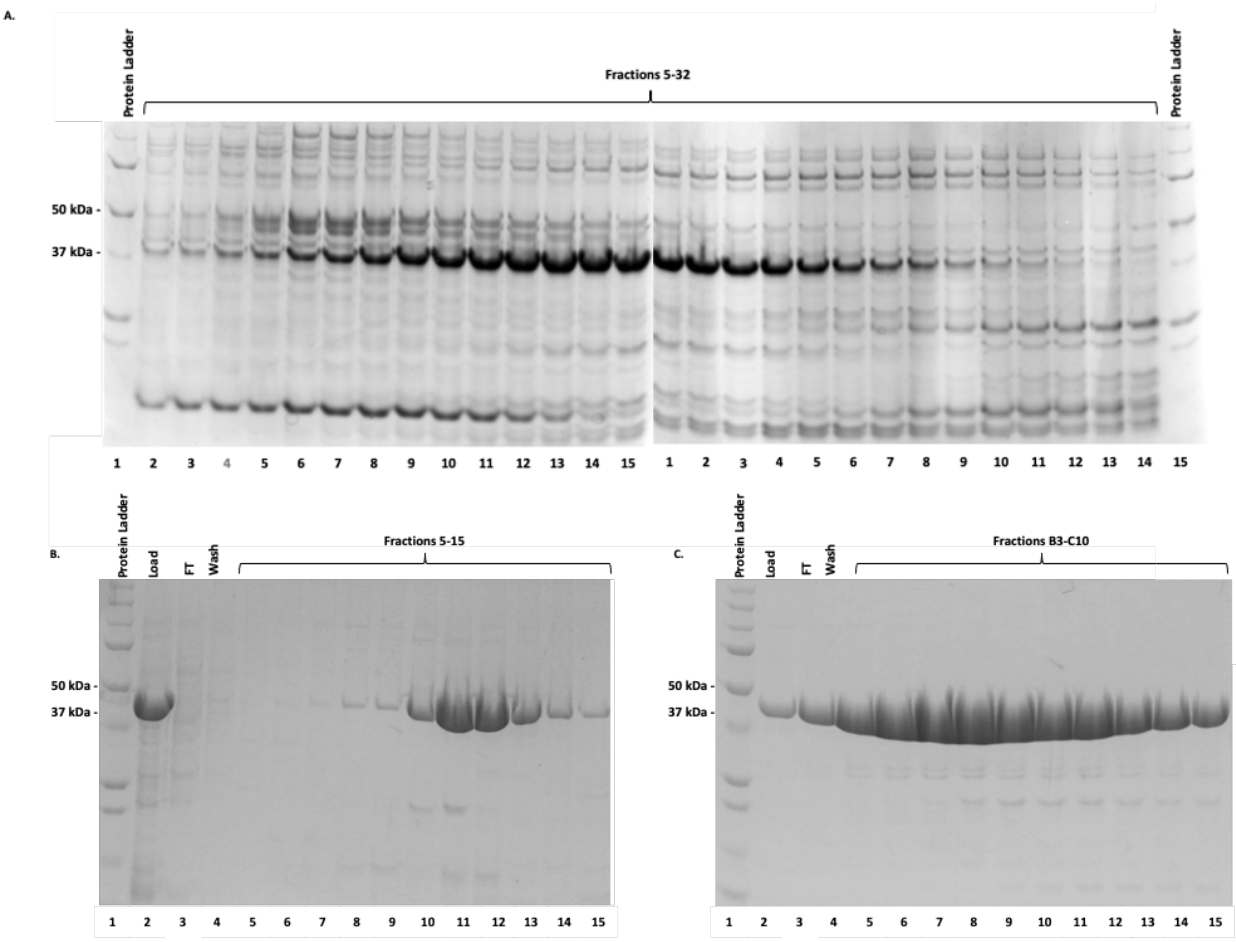


Figure 4.4 Analysis of full-length WT HsFEN1 purification. Coomassie-stained 10% SDS-PAGE gels representative of the progressive 3-step chromatographic purification of the ~42.6 kDa full-length WT HsFEN1. (a) 1st step: Purification of full-length WT HsFEN1 through Heparin affinity column. (b) 2ndt step: Purification of full-length WT HsFEN1 through the strong cation exchanger, SP column. (c) 3rd step: Final purity state of the resulting full-length WT HsFEN1 pool from size-exclusion chromatography used for storage and downstream analysis.

4.2.4 Zymogram-PAGE electrophoresis of purified HsFEN1 proteins

Aiming to establish the suitability of the produced HsFEN1 proteins for downstream structural and biochemical analyses, which will be determined by the presence or absence of any contaminating endogenous nucleases, zymogram-PAGE electrophoresis was carried out. Increasing amounts of both the purified truncated D179K and full-length WT HsFEN1 were electrophoresed on a 10% polyacrylamide gel containing type XIV DNA, with which HsFEN1 proteins allowed to react overnight. The use of type XIV DNA, as substrate, targets the exonucleolytic activity of WT HsFEN1, which was however, shown to be specific for recessed 5′-ends (Harrington and Lieber, 1994). Given this, and the fact that truncated D179K HsFEN1 is expected to be a catalytically inert active site mutant, the increasing protein amounts loaded per gel lane, aimed to increase sensitivity to the lowest possible exonuclease activity for

detection. In this way, even traces of nucleases with contaminating activity, whose detection would not have been possible at lower loading doses, are likely to be detected.

Ethidium bromide staining revealed single bands of increasing intensity at ~42.6 kDa, corresponding to the increasing amounts of full-length WT HsFEN1 loaded per gel lane (Figure 4.5). On the other hand, no bands, at ~37.7 kDa, corresponding to the truncated D179K HsFEN1 were detected at the lower loading doses tested (2.5 and 5 μg). At highly overloaded lanes (15 and 25 μg) however, single faint bands of increasing intensity were identified, corresponding to the binding capacity of the expected catalytically inert truncated D179K HsFEN1 (Figure 4.5). None of the produced batches of truncated D179K and full-length HsFEN1 was found to have any contaminating nuclease activity, indicating their suitability for downstream structural and biochemical analyses.

Detection of contaminating nucleases in any of the produced highly purified HsFEN1 batches would be likely to lead to entire batch rejection. Since the produced HsFEN1 proteins are meant for downstream analysis, the presence of any contaminating nucleases with detectable activity would be likely to interfere with the exo- and endonuclease activities of HsFEN1, skewing results and obstructing crystallisation. Whilst band detection on the zymogram gel is indicative of nuclease activity, its absence does not necessarily exclude the possibility for the presence of proteins with nucleolytic cleavage capabilities. Whether these proteins are contaminating nucleases or an expected catalytically inactive mutant, substrate binding and cleavage are dependent on their respective substrate specificities, reaction environment and, in the case of zymogram-PAGE electrophoresis, their renaturation efficiency in the polyacrylamide gel. Although the reaction environment used contained all required co-factors for HsFEN1 catalysis, type XIV DNA, a high MW linear DNA, is not the optimal substrate for HsFEN1.

Zymogram-PAGE electrophoresis here, is used as qualitative measure to determine the suitability of the produced protein batches for downstream analyses and not to characterize the nucleolytic activity of HsFEN1 proteins. The increasingly strong bands, corresponding to the activity of full-length WT HsFEN1, serve as positive controls for the absent/low intensity bands, corresponding to truncated D179K HsFEN1 and any contaminating nucleases even at the heavily overloaded lanes. Therefore, both protein batches are suitable for downstream structural and biochemical analyses.

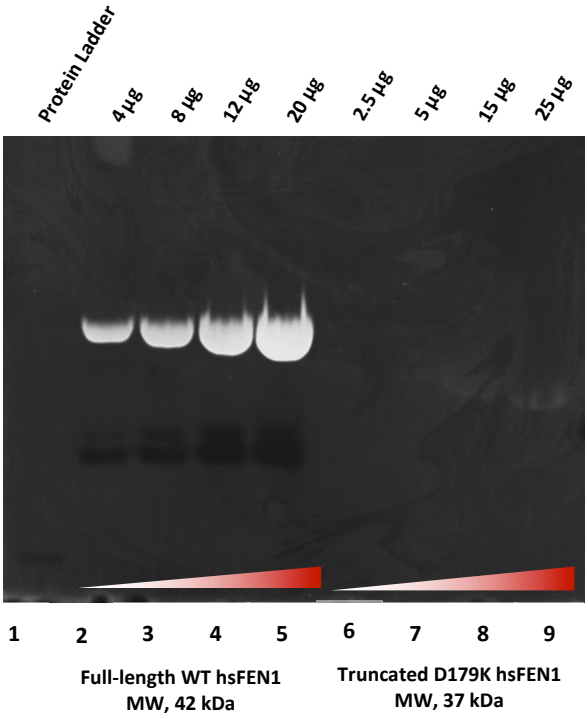


Figure 4.5 Zymogram-PAGE electrophoresis of full-length WT and truncated D179K HsFEN1 proteins. Increasing concentrations of the produced HsFEN1 proteins were subjected to 10% substrate-containing SDS-PAGE gel, the composition of which is described in Materials and Methods (Section 2.8.1.2). Electrophoresis of HsFEN1 proteins was carried out alongside a blue-stained protein ladder (non-fluorescent) for calibration of protein sizes (Lane 1). Lanes 2-5 and lanes 6-8, correspond to the increasing full-length WT (4, 8, 12 and 20 µg) and truncated D179K HsFEN1 (2.5, 5, 15, and 25 µg) concentrations, respectively. Detectable bands correspond to exonuclease activity.

4.2.5 Endonucleolytic cleavage of WT HsFEN1 and truncated D179K HsFEN1 proteins

Aiming to confirm the production of a catalytically competent full-length WT HsFEN1, capable of endonucleolytic cleavage, and a catalytically inactive truncated D179K HsFEN1 mutant, free of any detectable contaminating endonuclease activity, a FRET-based assay was carried out. Using a 200 nM constant concentration of a fluorescently dual labelled double-flap DNA substrate with a 4-nt long 5′-flap and a 1-nt long 3′-flap, the optimal DNA substrate of HsFEN1, truncated D179K HsFEN1 (1.33 nM, 13.3 nM and 133 nM) was evaluated for its endonucleolytic cleavage capabilities against 13.3 nM full-length WT HsFEN1. Their respective FRET signals were recorded over a 20 min period and compared to a negative/no enzyme control.

Detection of enhanced fluorescence signals, corresponding to endonucleolytic activity due to the prohibited energy transfer between the two fluorophores, was only possible in the presence of full-length WT HsFEN1 (Figure 4.6). The truncated D179K HsFEN1 was found to be catalytically inactive at all three concentrations tested, with no significant difference between them and the negative/no enzyme control (Figure 4.6).

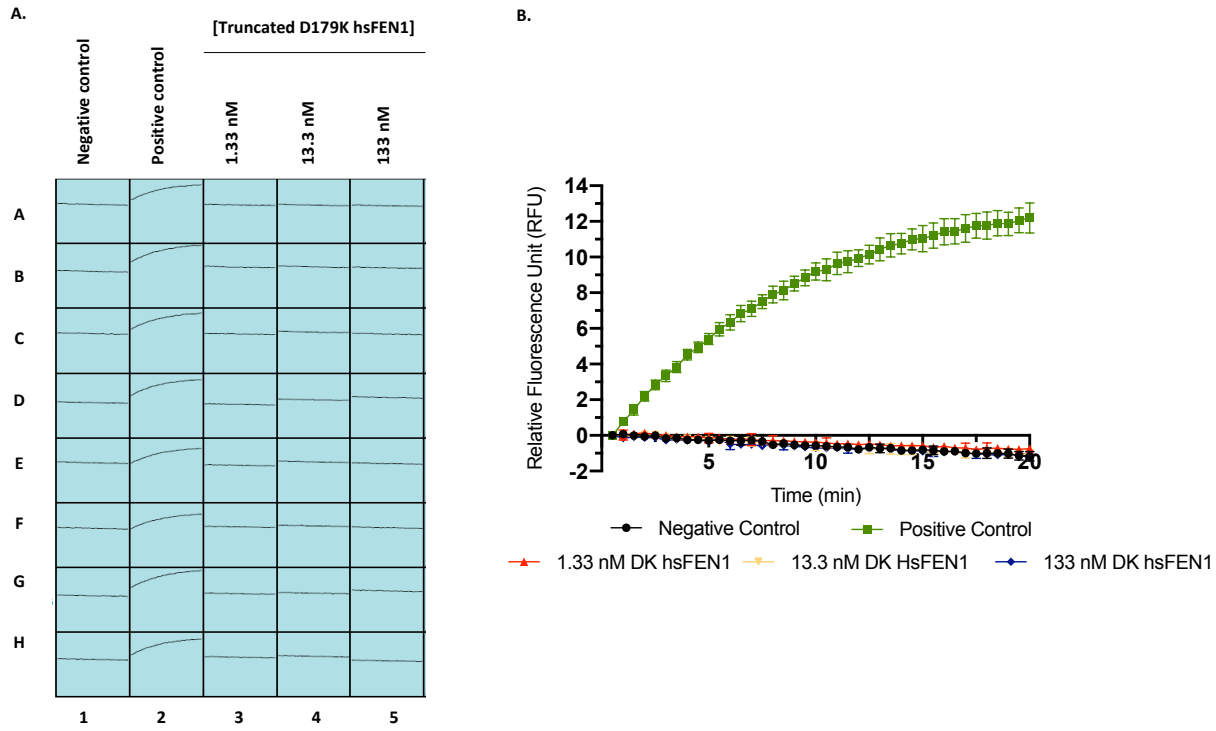


Figure 4.6 Structure-specific endonuclease activity of truncated D179K and full-length WT HsFEN1. (a) Fluorometric report representative of the full emission spectra recorded over a period of 20 min in the presence of A3Invader (double-flap) DNA substrate. Columns 1-5 represent the negative/no enzyme control, positive/WT HsFEN1 control (13.3 nM) and truncated D179K hsFEN concentrations from 1.33 nM, 13.3 nM and 133 nM, respectively. Rows A-H are representative of the 8 technical replicates. (b) Corresponding fluorescence emission data of truncated D179K and full-length WT HsFEN1. Cleavage potential by truncated D179K HsFEN1 was evaluated at 1.33 nM, 13.3 nM and 133 nM, against 13.3 nM of full-length WT HsFEN1 and a negative/no enzyme control. Relative fluorescence was plotted against time and analysed by non-linear regression following Michaelis-Menton equation (GraphPad Prism 8.2). Error bars represent the SD of the normalised mean, calculated from 8 technical replicates performed in a single experiment (i.e. $N = 1, n = 8, Z' = 0.75$)

4.2.6 Mass spectrometry analysis of purified HsFEN1 proteins

Samples of the purified truncated D179K and full-length WT HsFEN1 proteins, were sent for mass spectrometry analysis using Bruker Reflex III Matrix-assisted laser desorption/ionisation (MALDI) mass spectrometer to verify their identity and integrity. The resulting deconvoluted protein masses were compared to their respective theoretical masses and the difference between the two was utilised to determine their suitability for downstream processes. Mass spectrometry typically revealed single peaks at 37740.83 and 42461.4 Da corresponding to the truncated D179K and full-length WT HsFEN1 proteins, respectively (Supplementary Information; Figure SI.6). For both, the calculated difference between the theoretical and deconvoluted masses was between 130.75-131.58 Da (Table 4.2). This is a commonly seen difference in protein mass spectrometry corresponding to the N-terminal methionine, often processed during protein maturation (Giglione *et al*., 2004). The remaining 1 Da difference for truncated D179K

HsFEN1 is possibly due to ionization (pH>pI). Given the interactions between Gly2 and HsFEN1 active site, shown in previous HsFEN1 structures, the presence of N-terminal methionine would likely inactivate the enzyme (Tsutakawa *et al*., 2011; Tsutakawa *et al*., 2017).

Table 4.2 Deconvoluted and theoretical masses of truncated D179K and full-length WT HsFEN1 proteins

HsFEN1 protein	Theoretical Mass (Da)	Deconvoluted Mass (Da)	Difference (Da)
Truncated D179K	37871.58	37740.83	130.75
Full-length WT	42592.98	42461.4	131.58

Theoretical mass refers to the estimated mass using the ExPASy protein parameters (ProtParam) tool, based on the provided amino acid sequence of each protein. Deconvoluted mass refers to the estimated protein mass by electrospray ionization spectrometry.

4.3 Crystallisation of HsFEN1-DNA complexes

4.3.1 Screening for crystallisation conditions

Aiming to expand the current knowledge on HsFEN1 catalysis and potentially unravel the basis of the reported functional differences in base-pairing/unpairing, scissile phosphate bond and Tyr40 orientation, crystallisation trials of the produced high purity HsFEN1 proteins with DNA substrates were performed. Using symmetric DNA palindromes with a single 5′-flap at each side of an annealed molecule, including 5ov4 previously crystallised with T5FEN, crystallisation trials were carried out following the two different approaches presented in Section 4.1.2.

Both the truncated D179K and the native full-length WT HsFEN1 proteins were used for crystallisation, increasing the chances of obtaining any high resolution crystal structure(s). In all cases, proteins were supplemented with constant concentrations of the co-factors (K^{+}, M^{2+}) that were repeatedly shown to be essential for HsFEN1 catalysis. KCl was used as the source of K^{+}, mediating interactions with the upstream duplex DNA region, and $MgCl_2$/$CaCl_2$ as the sources of divalent metals with binding affinity for HsFEN1 catalytic site.

Capturing of intermediate catalytic states is required for direct comparisons to the published structures of HsFEN1:DNA complexes, aiming to understand the basis of important conformational differences that were previously reported (differences in base-pairing/unpairing, scissile phosphate bond and Tyr40 orientation). Therefore, crystallisation trials of HsFEN1:DNA complexes aimed primarily in capturing intact 5′-flaps threading through the conserved helical gateway. Mg^{2+}, the physiological co-factor of HsFEN1, was only used for crystallisation trials with the catalytically inactive truncated D179K HsFEN1, whereas Ca^{2+}, known to eliminate the HsFEN1 catalysed reaction rates, was used for trials with the active full-length WT HsFEN1 (Harrington and Lieber, 1994).

Preparation of HsFEN1:DNA complex mixtures for crystallisation trials, required selection of the protein:DNA molar ratios to be used, protein concentration and concentrations of the respective co-factors essential for catalysis. Reviewing the crystallisation conditions of published HsFEN1:DNA structures, though these refer to other HsFEN1 mutants or truncated WT HsFEN1 proteins, complex mixtures at 10 and 20 mg/ml protein concentration and 1:1/2:1 protein:DNA molar ratios were used (Table 4.3). In each case, screening was carried out in the presence of 50 mM KCl/ 5 mM $CaCl_2$/ 5 mM DTT for full-length WT HsFEN1:DNA mixture and 50-100 mM KCl/ 10-100 mM $MgCl_2$/ 1-5 mM DTT for the truncated D179K HsFEN1:DNA mixture.

Table 4.3 Crystallisation conditions of published HsFEN1:DNA structures to date[1]

PDB ID	WT/Mutant HsFEN1	5′-Flap length	Resolution (Å)	[Protein], (mg/ml)[2]	Crystallisation buffer[3]	Temperature (°C)	Reference
5ZOE	D181A HsFEN1	3-nt	1.95	5	35-40% (w/v) PEG 3350/ 20 mM Tris, pH 7.8/ 100 mM KCl/ 10 mM $MgCl_2$	20	(Xu *et al.*, 2018)
5ZOG	R192F HsFEN1	3-nt	2.3	5		20	
5ZOF	D181/R192F HsFEN1	3-nt	2.2	5		20	
5UM9	D86N HsFEN1	3-nt	2.8	4.75	12% mPEG 2,000/ 20% saturated KCl/ 5% ethylene glycol/ 100 mM HEPES, pH 7.5/ 2.2 mM SmSO4	15	(Tsutakawa *et al.*, 2017)
5K97	D233N HsFEN1	2-nt	2.1	4.1	24% mPEG 2,000/ 20% saturated KCl/ 5% ethylene glycol/ 100 mM HEPES, pH 7.5/ 1.6 mM SmSO4	15	
5KSE	R100A HsFEN1	4-nt	2.1	4.75	22% mPEG 2,000/ 20% saturated KCl/ 5% ethylene glycol/ 100 mM HEPES, pH 7.5/ 1.8 mM SmSO4	15	
3Q8K	WT HsFEN1	4-nt	2.2	12.5	16% mPEG 2,000/ 20% saturated KCl/ 0.5 mM Sm2(SO4)3/ 10 mM Mg(NO3)2/ 5% ethylene glycol/ 100 mM HEPES, pH 7	15	(Tsutakawa *et al.*, 2011)
3Q8L	WT HsFEN1	1-nt	2.3	12.5	17% mPEG 2,000/ 20% saturated KCl/ 0.5 mM Sm2(SO4)3/ 5% ethylene glycol/ 50 mM Bis-Tris, pH 7.5	15	
3Q8M	D181A HsFEN1	1-nt	2.6	12.5	20% mPEG 2,000/ 20% saturated KCl/ 10 mM CalCl2/ 5% ethylene glycol/ 50 mM Bis-Tris, pH 7.5	15	

[1]All published structures of HsFEN1 shown here have been determined using X-ray crystallography, based on the available information deposited in PDB database (Date accessed: July, 2021). HsFEN1 structures are shown in a chronological order, starting with the most recent.

[2][Protein], refers to the final protein concentration in the crystallisation drop.

[3]Crystallisation buffer, refers to the buffer used at 1:1 volumetric ratio with the respective protein:DNA complex solutions to set the crystallisation drops. The reported concentrations correspond to those of the stock buffer used, and not in the crystallisation drop.

Crystallisation drops were set using 1:1/1:2 volumetric ratio of the prepared protein:DNA mixture and crystallisation buffer. PACT, JCSG+ and Proplex screens, suitable for crystallisation of protein complexes, are among those tested. Silver Bullets, an additive screen, was also screened for crystallisation conditions of the full-length WT HsFEN1 in complex with 5cCv5 and 5ov4ex DNA substrates (Figure 4.7). Although Silver Bullets is primarily used for optimisation of promising conditions, it is a powerful alternative for crystallisation of macromolecular complexes, whose crystal formation does not seem to be dependent on precipitant concentration and/or pH (Larson *et al*., 2008).

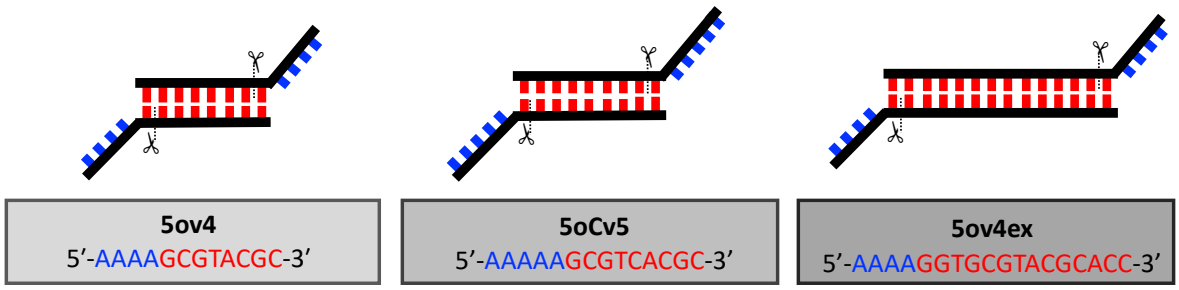


Figure 4.7 Nucleotide sequences and structures of 5ov4, 5oCv5 and 5ov4ex DNA substrates. Schematic representation of the palindromic 5ov4, 5oCv5 and 5ov4ex 5′-flap structures. 5ov4 is representative of the substrate DNA previously co-crystallised with T5FEN, which was used as the basis for designing 5oCv5 (having a central C:C mismatch) and 5ov4ex oligonucleotides co-crystallised with full-length WT HsFEN1 during initial screening with Silver Bullets screen. In all cases, the resulting substrate molecules of the individual self-annealing oligonucleotides is composed of a duplex region with an either 4- or 5-nt long overhang at each 5′-end of 5ov4ex and 5oCv5, respectively. Scissors are indicative of the scissile phosphodiester bond.

Screening for crystallisation conditions with the Silver Bullets screen was performed using a 0.1 M HEPES, pH 7/ 20% PEG3350 buffer, the selection of which was based on the crystallisation conditions of publicly available HsFEN1:DNA structures (Table 4.3). Each full-length WT HsFEN1:5oCv5 and full-length WT HsFEN1:5ov4ex complex mixture used, was prepared at 10 and 20 mg/mL protein concentration, 2:1 protein:DNA molar ratio and 50 mM KCl/ 5 mM $CaCl_2$/ 5 mM DTT. Crystallisation drops were set using sitting-drop vapour diffusion and a 1:0.8:0.2 complex mixture:buffer:additive volumetric ratio. After 3-days at 17°C, clusters of thin needle-shaped crystals of full-length WT HsFEN1 with both 5oCv5 and 5ov4ex were grown (Figure 4.8; Table 4.4). In their majority, the identified crystals seemed to grow at both 10 and 20 mg/mL protein concentration, under multiple conditions with distinctly different additive mixtures (Figure 4.8; Table 4.4). In all cases, the protein nature of the emerging crystals was confirmed using methylene blue, utilizing the ability of protein crystals to absorb and retain the small-molecule dye through their solvent channels.

This needle-shaped morphology of 1D crystals is typically problematic and can often become extremely difficult to work with (Civati *et al*., 2021). They can easily break and/or crack, thereby not enabling harvesting of individual needles for data collection. Even when loop

mounting is possible, their positioning within/over the micro-tubes of standard circular loops is also difficult. Over-hydration of needle-shaped crystals is an additional issue increasing background X-ray scatter and ice ring formation during flash cooling and data collection, compromising diffraction power. Given these, further optimizations were required to produce crystals of sufficient size and quality for data collection and potentially structure determination.

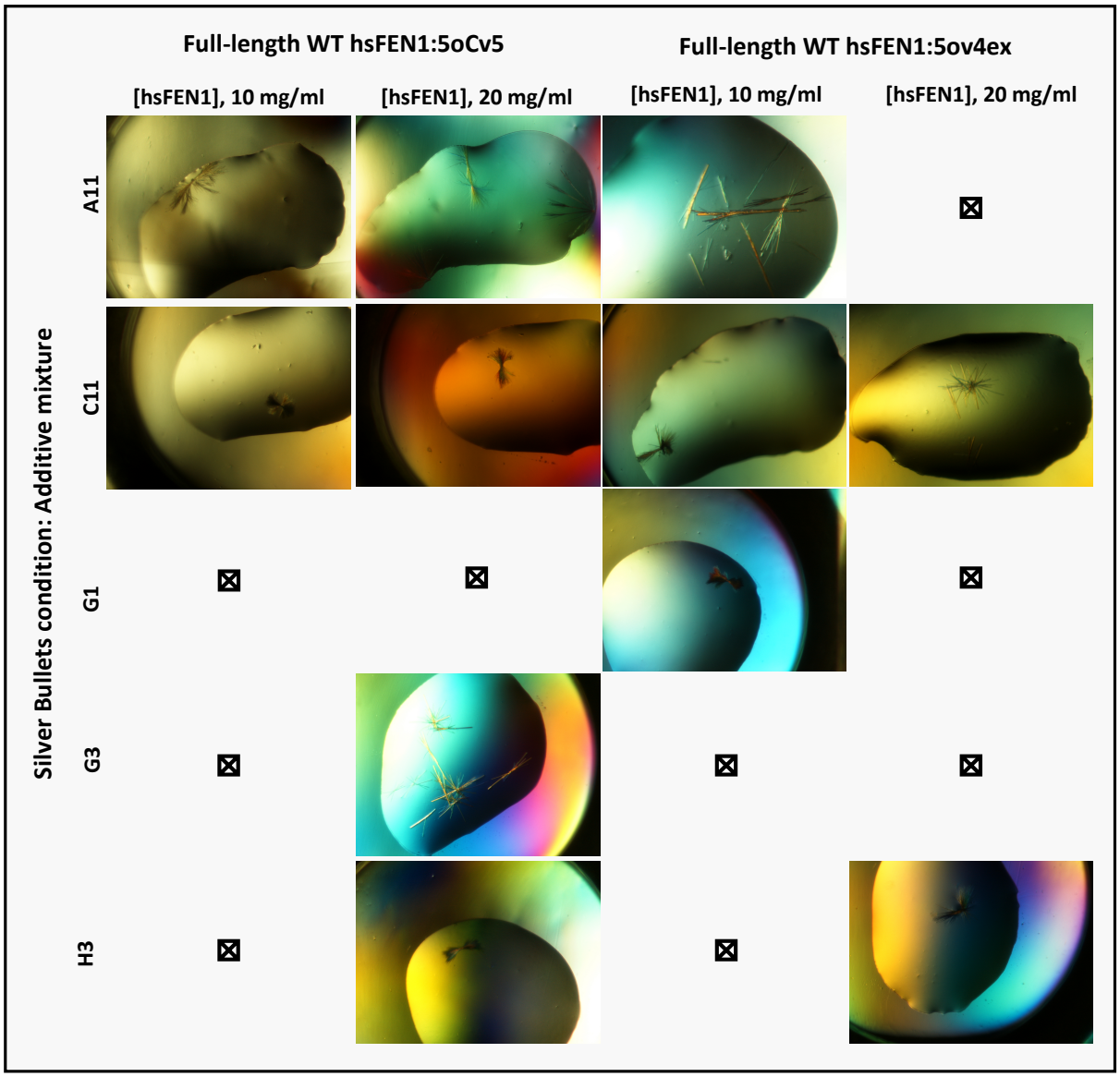


Figure 4.8 Initial screening hits of full-length WT HsFEN1:5oCv5 and full-length WT HsFEN1:5ov4ex complexes. Initial screening using Silver Bullets screen produced crystals of full-length WT HsFEN1:5oCv5/5ov4ex grew in crystallisation drops of 400 nl total volume, using 1:0.8:0.2 complex mixture:buffer:additive volumetric ratio. Compositions of buffer, complex and additive mixtures, corresponding to the crystallisation conditions of the presented crystals are shown in Table 4.4. "Initial" refers to crystals grown 3-days post crystallisation trial set up.

Table 4.4 Summary of initial crystallisation conditions of full-length WT HsFEN1:DNA using Silver Bullets screen[1,2]

Full-length WT HsFEN1:5oCv5[3]		Full-length WT HsFEN1:5ov4ex[3]		Precipitant	Buffer	Silver Bullets condition: Additive mixture
[HsFEN1], 10 mg/mL	[HsFEN1], 20 mg/mL	[HsFEN1], 10 mg/mL	[HsFEN1], 20 mg/mL			
X	X	X	-	20% (w/v) PEG3350	0.1 M HEPES, pH 7	**A11:** 0.33% (w/v) 2,5-Pyridinedicarboxylic acid/ 0.33% (w/v) 4-Nitrobenzoic acid/ 0.33% (w/v) Mellitic acid/ 0.02 M HEPES-Na, pH 6.8
X	X	X	X			**C11:** 0.16% (w/v) Azelaic acid/ 0.16% (w/v) m-Benzenedisulfonic acid-Na_2/ 0.16% (w/v) Mellitic acid/ 0.16% (w/v) Pimelic acid/ 0.16% (w/v) Pyromellitic acid/ 0.16% (w/v) trans-Cinnamic acid/ 0.02 M HEPES-Na, pH 6.8
-	-	X	-			**G1:** 0.16% (w/v) 5-Sulphosalicylic acid dihydrate/ 0.16% (w/v) Dodecanedioic acid/ 0.16% (w/v) Hippuric acid/ 0.16% (w/v) Mellitic acid/ 0.16% Oxalacetic acid/ 0.16% (w/v) Suberic acid/ 0.02 M HEPES-Na, pH 6.8
-	X	-	-			**G3:** 0.25% (w/v) 2.2'-Thioglycolic acid/ 0.25% (w/v) Azelaic acid/ 0.25% (w/v) Mellitic acid/ 0.25% (w/v) trans-Aconitic acid/ 0.02 M HEPES-Na, pH 6.8
-	X	-	X			**H3:** 0.2% (w/v) 3,5-Dinitrosalicylic acid/ 0.2% (w/v) 4-Aminobenzoic acid/ 0.2% (w/v) Benzamidine hydrochloride/ 0.2% (w/v) Hexamminecobalt(III) chloride/ 0.2% (w/v) Mellitic acid/ 0.02 M HEPES-Na, pH 6.8

[1]The crystallisation conditions presented here, only refer to those in which spontaneous nucleation occurred (post 3-days).

[2]Concentrations refer to stock solutions pre-crystallisation drops set up, using the 1:0.8:0.2 complex mixture:buffer:additive volumetric ratio.

[3]Complex mixtures was prepared at 10 and 20 mg/mL protein concentration, 2:1 protein:DNA molar ratio and 50 mM KCl/ 5 mM $CaCl_2$/ 5 mM DTT.

4.3.2 Selection of full-length WT HsFEN1:DNA crystal growth condition(s) for optimisation

Given the unsuitable morphology of the initial hit crystals for the collection of any diffraction data, and all the more so for structure determination, optimisation experiments were carried out to potentially produce individual crystals of sufficient size and quality. Despite the existence of the well-established grid strategy for optimisation of initial crystallisation conditions, the identification of multiple hits with spontaneous nucleation (post 3-days) using Silver Bullets screening, pre-required wise selection of the hit condition(s) for further optimisation.

Screening hit conditions were composed of a common buffer/precipitant system (0.1 M HEPES, pH 7/ 20% PEG3350) with however, distinctly different additive mixtures (Table 4.5). Given that no crystals of full-length WT HsFEN1:DNA complexes were previously obtained under the same buffer/precipitant system in the absence of any additives or in the presence of additives other than those of Silver Bullets, it was rational to hypothesise that nucleation and crystal growth was dependent on the respective Silver Bullets additive mixtures.

Aiming to facilitate selection of hit condition(s) for optimisation, composition analysis of Silver Bullets additive mixtures identified in the initial hit conditions was carried out. Composition analysis focused on the identification of commonalities, variables and/or trends among these apparently different additive mixtures, that can be potentially utilised for optimisation. Since buffer and pH (0.02 M HEPES -Na, pH 6.8) were universal throughout the hit additive mixtures with an expected insignificant effect to stabilisation and lattice formation, composition analysis focused on the individual small molecules composing each additive mixture. The respective small molecules constituted aliphatic and/or aromatic moieties combined with charged and/or uncharged functional groups (Table 4.5). In their majority, ≥2 carboxyl groups and/or combinations of carboxyl-, amino-, hydroxyl-, nitro- and/or sulfonyl-groups, were in the possession of each of these molecules. Such molecules, are likely to interpose themselves between molecules of the target protein and interact intermolecularly by restraining its dynamic nature and thus, stabilizing intermediate conformations (McPherson and Cudney, 2006). These stabilising effects are likely to promote crystal lattice formation and impact crystallisation behaviour (McPherson and Cudney, 2006).

Despite the common multicomponent nature of hit additive mixtures, the individual chemicals composing each these mixtures varied significantly. Mellitic acid or benzenehexacarboxylic acid (IUPAC name) was interestingly found in all hit additive mixtures at concentrations starting from 0.07% to 0.33 % (w/v) of the total 1% (w/v) concentration of additive mixtures in each condition (Table 4.5). Although these observations suggest a perhaps important role of mellitic acid in nucleation and crystal formation of full-length WT HsFEN1:DNA complexes, the crystals formed suggest no impact of the differing concentrations, at the range tested, on the resulting morphology.

Given these and the universal multivalent nature of hit additive mixtures, selection of an initial crystallisation condition for downstream optimisation was based on "trial and error". Additive mixture H3 was the condition initially selected for optimisation, as the only hit condition containing a heavy metal coordination complex, without however ruling out the possibility of optimisations on the basis of the other additive mixtures.

Table 4.5 Composition of Silver Bullets additive mixtures in initial hit conditions, during crystallisation of full-length WT HsFEN1:DNA complexes

Silver Bullet	Silver Bullets screen condition				
	A11	C11	G1	G3	H3
2,5-Pyridinedicarboxylic acid	X				
4-Nitrobenzoic acid	X				
Mellitic acid	X	X	X	X	X
Azelaic acid		X		X	
m-Benzenedisulfonic acid-Na_2		X			
Pimelic acid		X			
Pyromellitic acid		X			
trans-Cinnamic acid		X			
5-Sulfosalicylic acid dihydrate			X		
Dodecanedioic acid			X		
Hippuric acid			X		
Oxalacetic acid			X		
Suberic acid			X		
2,2'-Thiodiglycolic acid				X	
trans-Aconitic acid				X	
3,5-Dinitrosalicylic acid					X
4-Aminobenzoic acid					X
Benzamidine hydrochloride					X
Hexamminecobalt(III) chloride					X
HEPES-Na, pH 6.8	X	X	X	X	X

4.3.3 Optimisation of crystal growth conditions for full-length WT HsFEN1:DNA complexes

Aiming to obtain 3D crystals of sufficient quality for the collection of diffraction data and ultimately structure determination, a series of optimisation experiments was carried out on the basis of condition H3. H3 set up using a 1:0.8:0.2 complex mixture:buffer:additive volumetric ratio, where buffer is 0.1 M HEPES, pH 7/ 20% PEG3350 and protein, a 20 mg/mL full-length WT HsFEN1 mixed at 2:1 molar ratio with 5oCv5/5ov4ex (also supplemented with 50 mM KCl/ 5 mM $CaCl_2$/ 5 mM DTT), enabled optimisation in various directions.

Initial optimisations focused on expanding the tested range of pH and precipitant concentration. Using sitting-drop vapour diffusion and the same drop set up, optimisations were carried out at pH 6.5 to 9 and 18-24 % (w/v) PEG3350 concentration, in the presence and absence of an additional 50 mM $CaCl_2$ in the crystallisation buffer. Although 3-days later the formation of

crystals at multiple conditions with ≥20% (v/v) PEG3350 and pH ≥7 suggested a successful optimisation, crystal morphology remained unsuitable for the collection of diffraction data. About a month later however, promising 3D microcrystals were also found to grow along with the growing clusters of thin needle-shaped crystals (Figure 4.9). Birefringence under polarized light confirmed the protein nature of the slow-growing 3D crystals. Whilst some of these crystals were harvested for the data collection (Supplementary Information; Figure SI.5 and Table SI.5), seeding stocks were generated for further optimisations. Four seeding stock dilutions were prepared for optimisation following the standard procedure described in Materials and Methods (Section 2.11.4).

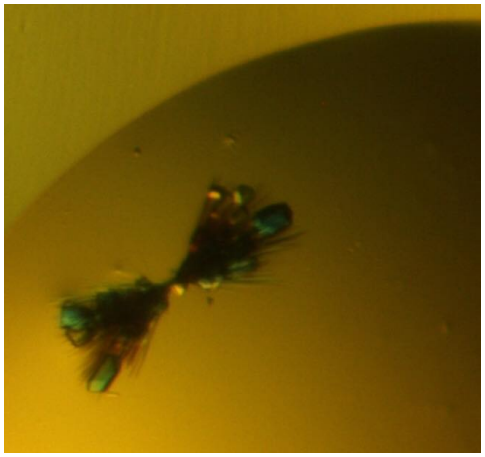


Figure 4.9. Full-length WT HsFEN1:5ov4ex crystals obtained during optimisation of initial crystallisation conditions. The presented crystals of mixed 1D and 3D morphologies were grown in 400 nL crystallisation drop, set up using a 1:0.8:0.2 complex mixture:buffer:additive volumetric ratio. Buffer is 0.1 M HEPES, pH 8/ 20% PEG3350/ 50 mM $CaCl_2$ and protein, a 20 mg/mL full-length WT HsFEN1 mixed at 2:1 molar ratio with 5ov4ex (also supplemented with 50 mM KCl/ 5 mM $CaCl_2$/ 5 mM DTT). Crystal image is representative of the morphologies obtained, but crystal size is not comparable to the previously presented crystals images due to the different scales used for imaging. Collected preliminary diffraction data of the presented 3D microcrystals are shown in Supplementary Information (Figure SI.5; Table SI.5).

Aiming to further improve shape, size and quality of the identified 3D crystals, a new series of optimisations by seeding were carried out using sitting-drop vapour diffusion at a constant drop volume of 400 nL. Several variables were manipulated throughout the downstream process, for each of which a 0.1 M HEPES pH 7.5-8.5/ 20-22% PEG3350 buffer was used, in the presence or absence of an additional 50 mM $CaCl_2$. Optimised variables involved seeding stock dilution, substrate DNA, complex mixture:buffer:additive:seeds volumetric ratio and the concentrations of the respective additive mixture components (H3).

Full-length WT HsFEN1:DNA complexes with alternative DNA substrates, other than 5oCv5/5ov4ex palindromes, were used for optimisation aiming at improved morphologies. These involved a series of optimal DNA oligonucleotides for HsFEN1 activity, all bearing a 1-nt long 5′-flap and various lengths 5′-flaps (Figure 4.10). Preparation of full-length WT HsFEN1:JT complex mixtures was carried out as described earlier for complexes with 5oCv5/5ov4ex oligonucleotides at 20 mg/mL protein concentration and 1:1 molar ratio.

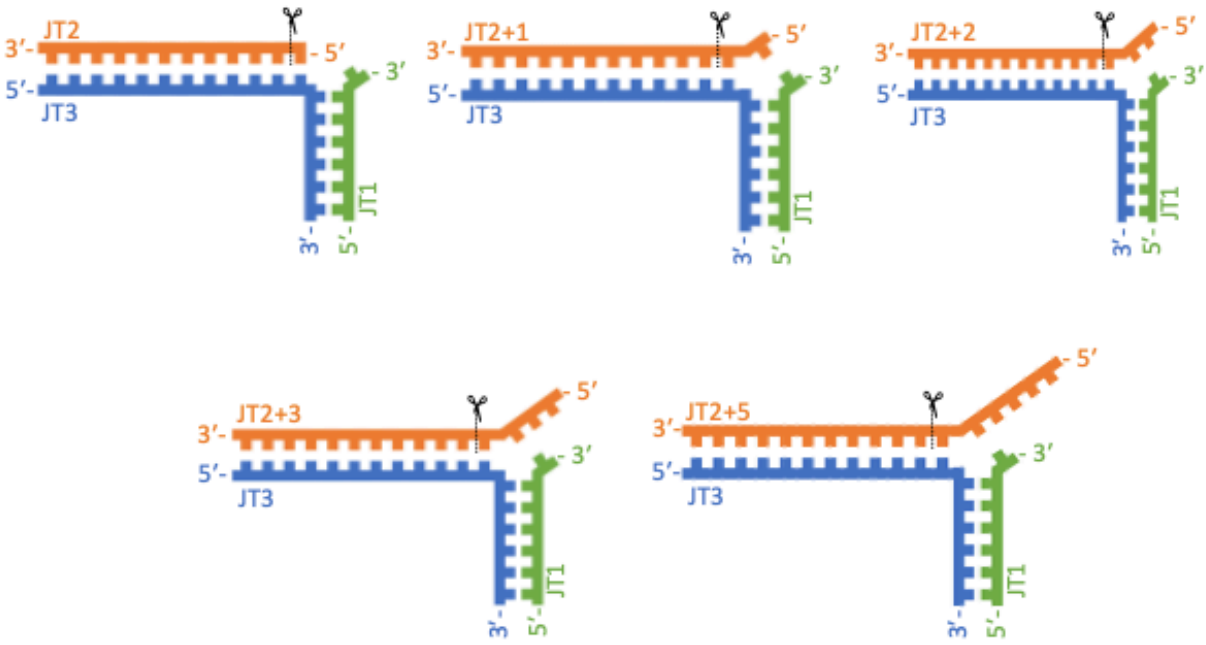


Figure 4.10 Structures of JT oligonucleotides used for crystallisation. Construction of each of the presented JT oligonucleotide, occurs through a combination of three individual oligonucleotides annealed together. These involve in all cases JT1 (green) and JT3 (blue), with a range of JT2+X oligonucleotides, where X determined the length of the protruding 5′-flap. Scissors are indicative of the scissile phosphodiester bond. Nucleotide sequences are shown in Materials and Methods (Section 2.10; Table 2.2).

On the other hand, given that nucleation and/or crystal growth were believed to be highly dependent on additive, at least for this particular optimised condition, further optimisations were carried out by breaking down the H3 SB additive mixture. The individual effect of each H3 additive mixture component on crystal morphology, was evaluated at concentrations ranging from 0.25 to 2% (w/v) in the crystallisation drop.

Optimization of the selected crystallization condition was successful leading to an array of crystals with diverse morphologies (Figure 4.11). Although some retained the initial needle-shaped morphology, multiple others, grew consistently in the presence of 4-aminobenzoic acid and/or 3-aminobenzoic acid, seemed to have 3D morphologies suitable for data collection (Figure 4.11). Interestingly, these 3D-growing crystals, possessing the highest potential, were mostly obtained during optimization with HsFEN1:JT complexes and seeding with the highest seeding stock dilution (seeding solutions 3/4; Section 2.11.4). Crystal morphology seemed to be independent of the 5′-flap length of the used JT oligonucleotides, indicating a common effect of JTs in complex stabilization for crystallisation.

Crystal growth was also accelerated (typically 3-days to 1-week), with however unclear whether this is due to seeding, the optimization method used, and/or the individual effect of 4-aminobenzoic acid and its meta- form. In addition, despite the consistent requirement in

buffer/precipitant and additive type, their respective concentrations at which promising 3D crystals were found to grow, seemed to vary. Both hit buffer/precipitant and additive concentrations span the entire tested range of 0.1 M HEPES pH 7.5-8.5/ 20-22% PEG3350 and 0.25 to 2% (w/v) additive(s) in the crystallisation drop.

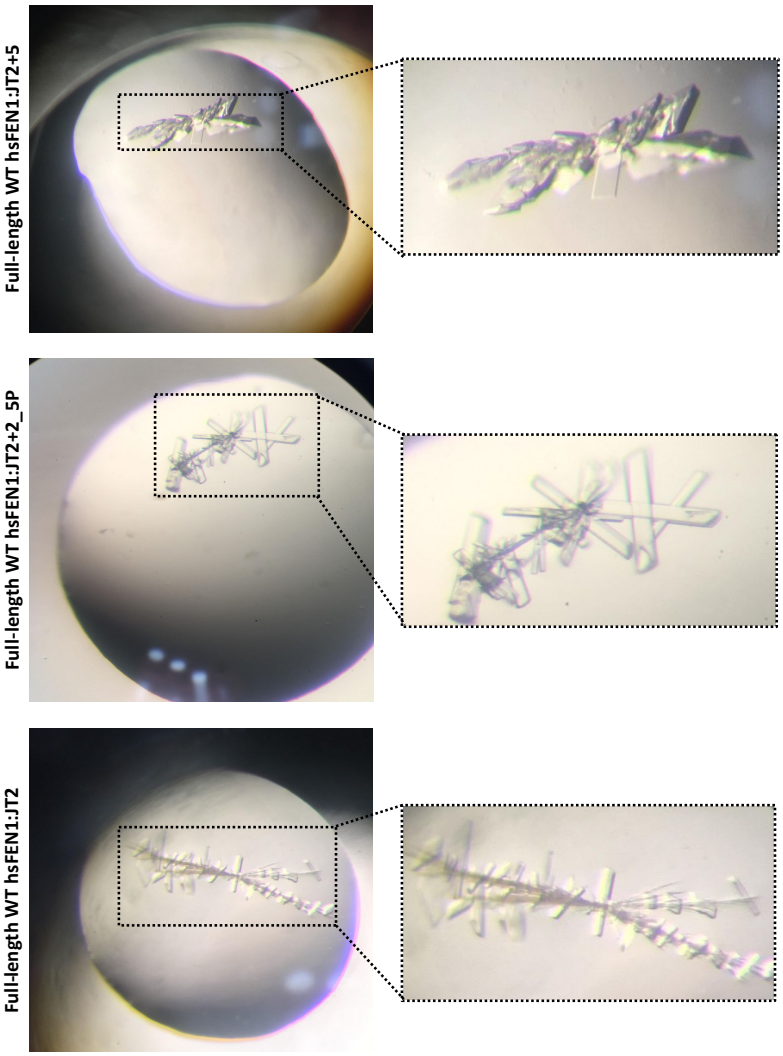


Figure 4.11 Examples of full-length WT HsFEN1:DNA crystals obtained during optimisation. Open- and close-views of improved in shape and size 3D crystals, grown in 400 nL crystallisation drops during optimisation. 3D crystals were grown either as clusters or along with the original needle-shaped clusters of 1D crystals. In their majority, the promising for data collection 3D crystals, involved plate-shaped crystals. The respective crystallisation conditions of the presented crystals are shown in Table 4.6.

Table 4.6 Overview of crystallisation conditions of full-length WT HsFEN1:DNA crystals, diffracted at ≤2.5 Å resolution[1,4]

Crystallised complex	Drop ratio[2]	Buffer	Additive % (w/v)	
			4-Aminobenzoic acid	4-Aminobenzoic acid/3-Aminobenzoic acid
HsFEN1:JT2+5	1:0.6:0.2:0.2	0.1 M HEPES pH 7.5/ 20% PEG3350	-	7.5
HsFEN1:JT2+2_5P[3]	0.8:0.6:0.4:0.2	0.1 M HEPES pH 7.5/ 20% PEG3350/ 50 mM $CaCl_2$	-	2.5
HsFEN1:JT2	0.9:0.7:0.2:0.2	0.1 M HEPES pH 8.5/ 24% PEG3350	17.5	-

[1]Each full-length WT HsFEN1:DNA complex mixture used was prepared at 20 mg/mL protein concentration, 1:1 protein:DNA molar ratio and 50 mM KCl/ 5 mM $CaCl_2$/ 5 mM DTT.
[2]Drop ratio, refers to the respective complex mixture:buffer:additive:seeds volumetric ratio used for crystallisation drop set up. For HsFEN1:JT2+2_5P crystal, additive corresponds to a mixture composed by 1:1 volumetric ratio of the original H3 SB additive mixture and 4-/3-aminobenzoic acid as indicated above.
[3]JT2+2_5P bears a 1-nt long 5′-flap and a phosphorylated 2-nt long 5′-flap.
[4]The collected diffraction data of full-length HsFEN1:DNA complexes, including DISTL plots, maps and data statistics are shown in Sections 5.3.3.

Overall, 3D-growing crystals of HsFEN1:JT complexes seemed to be the combinatorial result of multiple optimized variables within the same buffer/precipitant and additive range. The specific optimization conditions, reflective of the slight tweaks in concentrations and the range of HsFEN1:JT complexes crystallised, for the top quality crystals obtained are summarized in Table 4.6.

4.3.4 Data collection and processing

Full-length WT HsFEN1:DNA complexes were crystallised in minor variations of the same crystallisation buffer. Given that their discrimination was impossible by simply inspecting the respective crystal morphologies, collection of diffraction data was carried out to determine their quality and suitability for subsequent structure determination.

Prior to data collection, the optimised crystals of interest were harvested by loop mounting, cryo-protected appropriately and flash frozen for storage and transfer to Diamond light source, UK (Materials and Methods; Section 2.11.6). Full diffraction datasets of 1800-3600 images were typically collected using Diamond's iØ3 beamline at 1.8-2.2 Å resolution, 0.050-0.080 s exposure, 60x50 μm beamsize; 0.10° Ω Osc, 0.9119 Å Wavelength, 100% Transmission (Type: SAD). Processing of the collected diffraction datasets was carried out automatically using

fast_dp, xia2 DIALS and xia2 3dii pipelines, with xia2 DIALS typically resulting to better resolutions and quality indicators for subsequent structure determination (data not shown).

Despite the optimisations carried out aiming to achieve the highest possible resolution under the same slightly tweaked crystallisation conditions, no datasets were collected at ≤2 Å resolution for the outer/highest resolution shell. However, more than 20 datasets were collected between 2-3 Å resolution at buffer/precipitant and additive concentrations spanning the entire tested range. Among the collected datasets only 7 were found to lie ≤2.5 Å resolution, the top three of which, with diffraction data statistics within the standard cutoff criteria, were short-listed for downstream analysis. The respective crystallisation conditions and diffraction data for the top three datasets, representative of HsFEN1:JT2+5 and HsFEN1:JT2 complexes, are summarised in Table 4.7 and Figure 4.12. Standard cutoff criteria for prioritization of diffraction data to use for further structure determination are typically based on Pearson correlation coefficient ($CC_{1/2}$), multiplicity-corrected R value (R_{meas}), signal-to-noise ratio (Mean I/sig(I)) and completeness. $CC_{1/2}$ and R_{meas} are often refer to as "consistency or precision" metrics. $CC_{1/2}$ has a cutoff of ≤1 and shells by the correlation of the intensity values from two half datasets, whereas R_{meas} has a cutoff of ≤0.8 and evaluates the space group symmetry of multiple datasets (Karplus and Diederichs, 2012). In addition, a minimum value of two is used as a cutoff for the average signal-to-noise ratio of the measured intensities, while completeness indicative of the number of all theoretically possible unique observations measured in the collected dataset must be at nearly 100%.

A clear preference for crystallisation of full-length WT HsFEN1 in complex with JT oligonucleotides in P 1 2 1 (P2, composed by four 2 screw axis each of which equivalent to a rotation of 180°) or P 1 2_1 1 ($P2_1$, composed by four 2_1 screw axis each of which equivalent to a rotation of 180° and followed by a one-half translation of the repeat unit) space groups appeared to exist among the collected datasets ≤3 Å resolution, with however P 1 2 1 predominating (Table 4.7). Resolution of the outer/highest resolution shell was retained between 2.5-2.6 Å for HsFEN1:JT2+2_5P and HsFEN1:JT2 complexes and 2.1 Å for HsFEN1:JT2+5. Unit cells were in all three cases were analogous to monoclinic unit cells ($a \neq b \neq c$, $\alpha=\gamma$, $\beta \neq 90^{\circ}$) with not unusual dimensions for FEN1:DNA complexes (Tsutakawa et al., 2017). Therefore, indicating an increased likelihood of similar packing arrangement in the respective protein crystals, expected to be seen in the following phasing stage.

Table 4.7 Diffraction data statistics of full-length WT HsFEN1:DNA crystals diffracted at ≤2.5 Å resolution.[1,8]

	Crystallised complex		
	HsFEN1:JT2+5	hsFEN1:JT2+2_5P	HsFEN1:JT2
Total observations	243935 (11236)	110421 (5663)	256302 (12356)
Unique observations	72940 (3601)	36503 (1797)	39288 (1966)
Resolution (Å)	2.1-71.7 (2.1)	2.5-99.3 (2.5-2.6)	2.5-101.15 (2.5-2.6)
$CC_{1/2}$[2]	1 (0.3)	0.9 (0.1)	1 (0.4)
Completeness (%)[3]	99.7 (99.3)	98.9 (99.3)	99.9 (99.9)
Multiplicity[4]	3.3 (3.1)	3 (3.2)	6.5 (6.3)
Space group	P 1 2 1	P 1 21 1	P 1 2 1
Mean I/sig(I)[5]	8.5 (0.1)	2 (-0.2)	7.2 (0)
R_{meas}[6]	0.094 (4.307)	0.404 (-2.527)	0.112 (20.452)
Unit cell a, b, c (Å) α, β, γ (°)	61.05, 101.56, 105.79, 90, 106.71, 90	60.47, 93.45, 109.81, 90, 105.68, 90	60.89, 97.97, 105.48, 90, 106.36, 90
Cryoprotectant (v/v)[7]	20% Glycerol	Saturated paraffin oil	20% DMSO

[1]Diffraction data statistics collected at 0.9119 Å (iØ3 beamline) and autoprocessed using XDS in xia2 DIALS pipeline (Kabsch, 2010b; Winter, 2010). AIMLESS was used for data scaling and merging (Evans and Murshudov, 2013; Waterman et al., 2016). Values in parenthesis are representative of the outer/highest resolution shell.

[2]$CC_{1/2}$ is the calculated Pearson correlation coefficient between two half-split sets of reflections.

[3]Completeness is indicative of the number of all theoretically possible unique observations measured in the collected dataset.

[4]Multiplicity is representative of the average number of observations per individual unique reflection.

[5]Mean I/sig(I) is representative of the average signal-to-noise ratio of the measured intensities in the collected dataset.

[6]R_{meas} is the multiplicity-corrected R value version (Karplus and Diederichs, 2012; Wlodawer et al., 2008).

[7]Cryoprotectant, refers to the compound used to prevent ice formation during flash-cooling and data collection.

[8]DISTL plots, diffraction pattern typologies and crystal morphologies are shown in Figure 4.12.

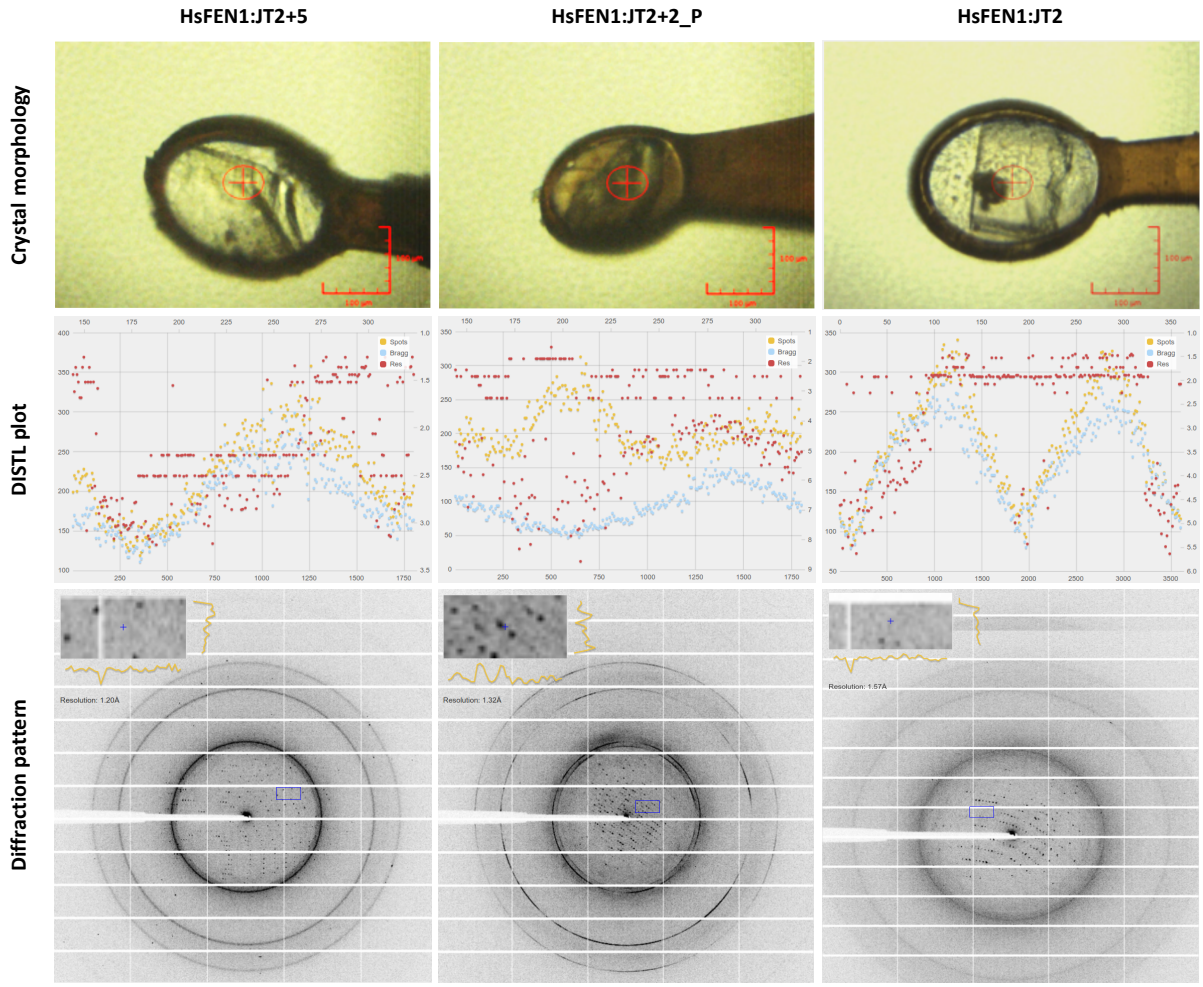


Figure 4.12 Crystal morphologies, DISTL plots and diffraction maps of full-length WT HsFEN1:DNA crystals. Full datasets of HsFEN1:JT2+5, HsFEN1:JT2+2_5P and HsFEN1:JT2 collected using iØ3 beamline at Diamond light source, UK (Resolution: 1.8-2.2 Å; Exposure: 0.050-0.080 s; Beamsize: 60x50 µm; Ω Osc: 0.10˚; No. Images: 1800-3600; Wavelength: 0.9119 Å; Transmission:100%; Type: SAD). Images of mounted crystals are indicative of the crystal plate morphology and size. DISTL plots, indicative of the number of images collected and their respective resolutions throughout data collection, and diffraction pattern typologies of each HsFEN1:DNA complex.

4.3.5 Calculation of Matthews coefficient (V_M)

Asymmetric unit (AU) content was estimated in all cases automatically through CCP4 to facilitate subsequent structure determination. Based on the theoretical MW of full-length WT HsFEN1 (Input FASTA sequence NP_004102.1) and the respective cell parameters of each of the collected diffraction datasets, cell volume, solvent content and V_M were calculated (Kantardjieff and Rupp, 2003; Matthews, 1968). The probability (N) of each particular V_M, indicative of the number of independent protein copies per crystallographic crystallographic asymmetric unit, was also estimated. Based on this, the existence of two protein molecules per asymmetric unit was predicted as the most likely solution for full-length WT HsFEN1:JT2+5,

full-length WT HsFEN1:JT2+2_5P and full-length WT hsFEN:JT2 complexes with solvent content (V_S) at ~60%. In all cases, the calculated cell volumes were analogous and the possibility for the existence of one, four or five protein molecules was extremely unlikely with *N* values as low as zero.

4.3.6 Structure determination, refinement and re-build of full-length WT HsFEN1-DNA complexes

Given the availability of HsFEN1 3D structures in PDB, molecular replacement was the selected method for rapid structure determination and generation of an initial structure solution, which can manually be refined and rebuilt. Using DIMPLE's integrated PHASER-MR software (Difference Map PipeLinE), a jointly developed pipeline by Diamond light source and CCP4, automated molecular replacement was carried out immediately after data collection and reduction. Due to the lack of full-length WT HsFEN1 crystal structures with/without substrate DNA, the structure of a truncated R192F HsFEN1 mutant complexed with DNA was input as the reference model (PDB ID: 5ZOG; Resolution: 2.3 Å).

Consistent to Matthews probability, molecular replacement provided an initial common solution with two protein molecules per asymmetric unit for each full-length WT HsFEN1:JT2+5, full-length WT HsFEN1:JT2+2_5P and full-length WT HsFEN1:JT2 crystal (Figure 4.13). Given the shared 99.7% sequence identity between truncated R192F HsFEN1 (reference model) and the crystallised full-length WT HsFEN1, the resulting models were, in their majority, well-fitted into electron density (Figure 4.14). The structure of truncated R192F HsFEN1 mutant used as a reference model for molecular replacement was in complex with a product DNA substrate, analogous to that seen after processing the JT2+5 DNA substrate by WT HsFEN1 in our crystallization trials (PDB ID: 5ZOG; Xu et al., 2018). However, despite the common use of double flap DNA substrates for crystallization, bearing a 1-nt long 3′-flap and either 3- or 5-nt long 5′-flaps, the differential aa sequence of the oligonucleotides used was likely responsible for the unmodelled DNA regions after molecular replacement. Such regions could easily be located in their majority by visual inspection due to the presence of clear positive electron density, setting the requirement for a series of subsequent refinement and building cycles to complete the final structures.

Despite the three different JT oligonucleotides (JT2+5, JT2+2_P5 and JT2) expected to be seen in the corresponding WT HsFEN1: DNA structures, cleavage of DNA substrates 1-nt into the ss-ds junctions on the 5′-flap strands seemed to have led to three identical WT HsFEN1 structures complexed with cleaved product substrates (Figure 4.13). Given these, the structure of HsFEN1: processed JT2+5 complex with the highest resolution among the three underwent further refinement.

Using COOT, residue-by-residue inspection was carried out for WT HsFEN1: processed JT2+5 complex for the identification of those problematic residues/regions requiring manual re-building and refinement. Among these and despite the initial well-fitted model, residues

Ser101-Thr127, Ala196-Pro203 and Gln333-Lys380 composing the flexible/disordered helical gateway, β-pin loop and C-terminal region, respectively, were the only completely unmodeled protein regions requiring manual fixing. These, in addition to the terminal DNA residues at both ends of the duplex substrate region, were modified to improve the respective correlation coefficients (CC) to the existing electron density maps and overall fitting. C-terminus was only build up to Arg332, because of the absence of electron density map for the downstream residues. Finally, molecules corresponding to water were added based on the unmodeled electron density blobs mainly around the active site. After several rounds of model-fit-refine in Refmac5, the qualitative R_{work}/R_{free} factors, indicative of the agreement between the collected and refinement data, were improved from 0.31/0.36 to 0.27/0.30 (Karplus and Diederichs, 2012; Wlodawer et al., 2008). Although the resolution of the collected dataset suggests that a lower R_{free} value can perhaps be achieved with the building of more water molecules and/or metal ions outside the active site, therefore future improvements are required, this is also constrained by the C-terminal truncation.

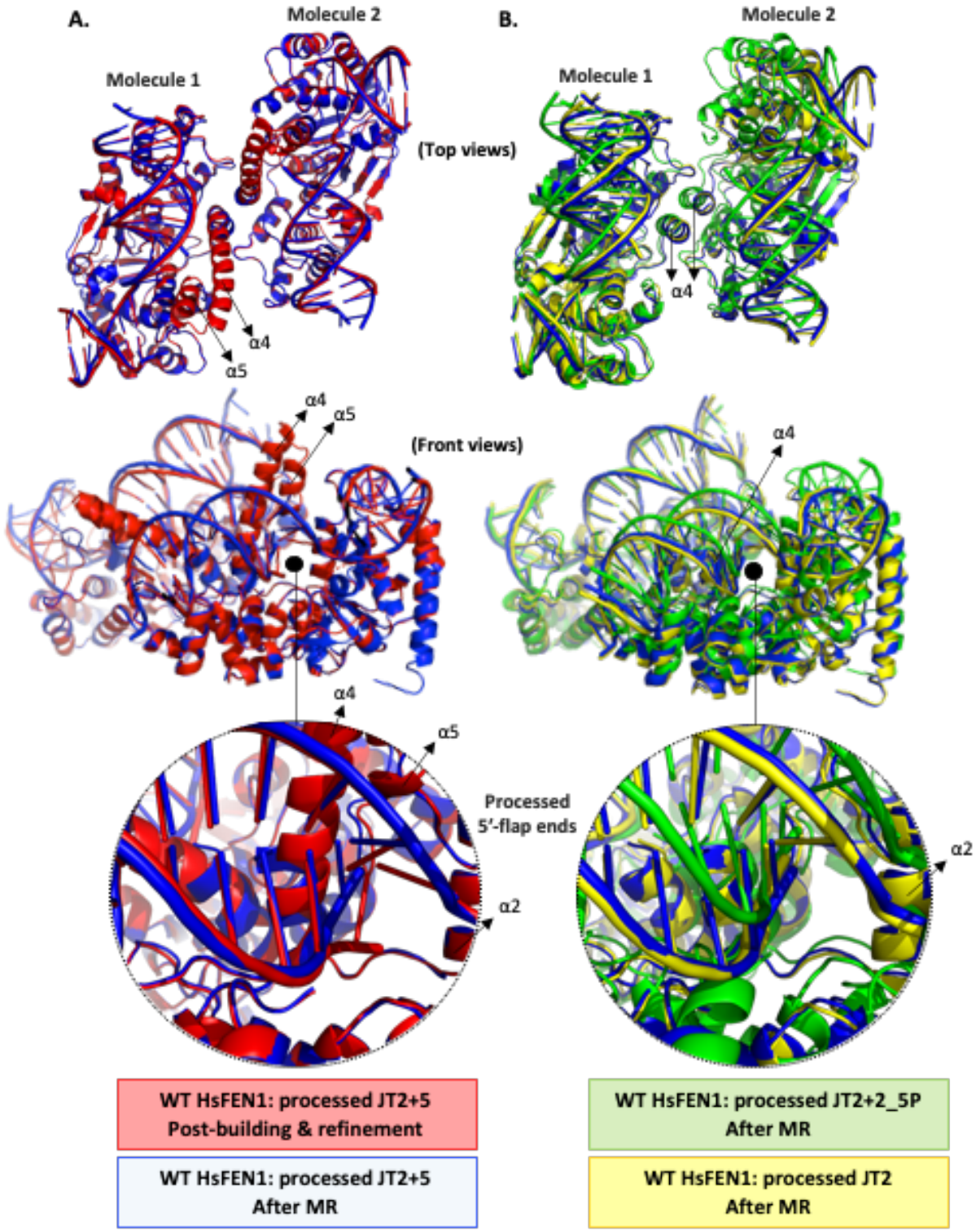


Figure 4.13 Comparison of HsFEN1: processed JT2+5 structure with its refined/(re)built and other HsFEN1: DNA structures. (A) Superimposition of the resulting after MR and post-refined and manual (re)built WT HsFEN1: processed JT2+5 complex structures. Absent after MR and later manually built cap and gateway helices are also indicated (α4 and α5). (B) Superimposition of WT HsFEN1 structures in complex with the identical JT2+5, JT2+2_5P and JT2 product DNA substrates after MR. (All panels) Superimpositions were carried out using both copies of WT HsFEN1 molecules in the asymmetric unit.

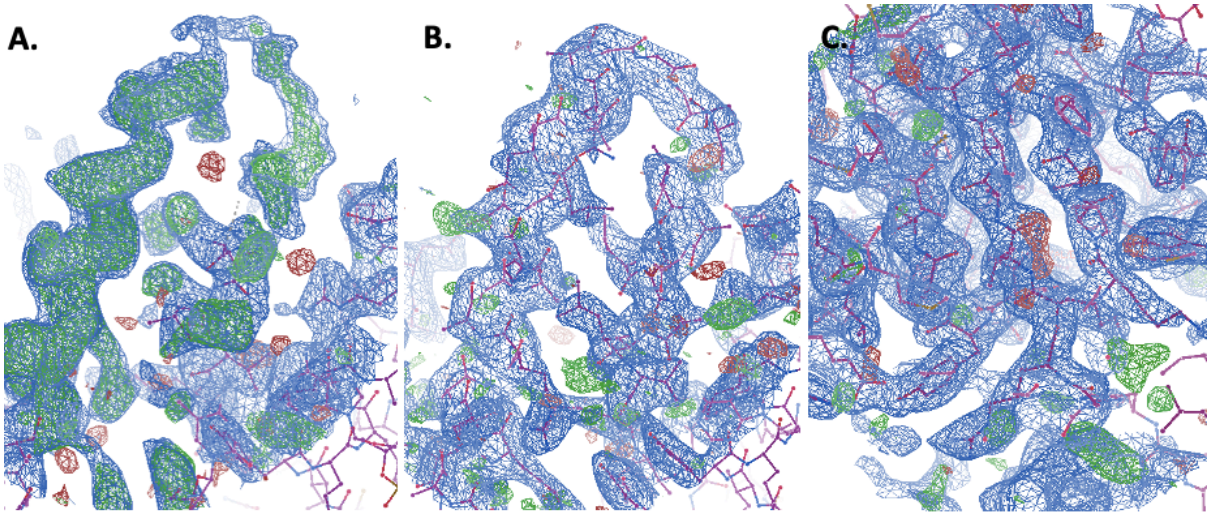


Figure 4.14 Electron density maps of WT HsFEN1: processed JT2+5 DNA complex pre- and post-refinement and rebuilding. Models are representative of the overall fitting in the electron density maps of WT HsFEN1 complexed with product JT2+5 substrate. (A) Electron density of the unmodelled helical gateway and cap regions after MR. (B) Electron density of the manually built and refined helical gateway and cap regions in COOT. (C) Example of a well-fitted region after MR, prior to any manual building and refinement. Positive (green) and negative (red) electron density maps are representative of accurately modelled and unmodelled regions, respectively (1σ with about 3σ positive map).

4.3.7 Structural analysis of WT HsFEN1: processed JT2+5 complex

Despite the use of native full-length WT HsFEN1 for co-crystallisation with JT2+5 oligonucleotide in the resulting HsFEN1:DNA complex structures, the flexible C-terminus was found disordered due to the absence of a HsFEN1-interacting protein partner. Given the complete lack of electron density corresponding to the C-terminus, HsFEN1 molecules in the resulting complex structures were build and refined up to the C-terminal Arg332.

Overall, virtual superimposition of the resulting WT HsFEN1: processed JT2+5 structure onto previously published HsFEN1:DNA complex structures revealed identical conformations with α2, α4 and α5 helices surrounding a mixture of parallel and antiparallel β-sheet-packed core (Tsutakawa *et al*., 2011; Tsutakawa *et al*., 2017; Xu *et al*., 2018) (Figures 4.14 and 4.15). Consistently to all previous HsFEN1:DNA complex structures, HsFEN1 interactions with DNA were primarily found with the upstream and ~100˚ bent downstream dsDNA on each side of helical gateway (Figures 4.14 and 4.15). As previously described, four main protein regions are mediating such interactions with the dsDNA, including the hydrophobic wedge (composed from α2 helix and α2- α3 loop), the β-pin region (composed from β8 and β9 strands), the directly and indirectly interacting through K^+ salt bridges H2TH motif (composed from α10-α11 loop and the corresponding α-helices) and the absent from HsEXO1 structure helix-hairpin-helix motif (composed from α14-α15 loop and their corresponding α-helices).

In addition, the conserved one helical turn distanced hydrophobic wedge and H2TH regions on each side of the active site further supported the reports of Tsutakawa *et al*., (2017) for an enforced ssDNA-dsDNA junction and their protruding 5′-flap selection of HsFEN1. This was not apparent by the later HsFEN1 structures of Xu *et al*., (2018) lacking the flexible gateway and cap regions (Figures 4.14 and 4.15).. Although the WT HsFEN1: processed JT2+5 structure solved within the context of the current thesis had its individual flexible helical gateway and cap completely unmodelled following molecular replacement, and with the only exception of α4-α5 loops, the clearly mapped electron density enabled their accurate determination through manual building for direct comparison (Figure 4.14).

Specificity for ssDNA-dsDNA junctions is also reinforced by active site interactions with the dsDNA next to the branched 5′-flap strand and the complementary to 5′-flap strand. In WT HsFEN1: processed JT2+5, cleavage occurred 1-nt into the ssDNA-dsDNA junction on the 5′-flap, consistent to all HsFEN1:product DNA structures, including D233N HsFEN1: product DNA shown in Figure 4.16. The resulting terminal nt (-1) on the processed 5′-flap strand remained unpaired with the scissile phosphate oriented towards the active site. The released ss product from 5′-flap processing was not retained in the active site for crystallization following cleavage, in contrast to the structure of truncated D233N HsFEN1:product DNA complex (PDB ID: 5K97) published by Tsutakawa *et al*., (2017). The overall orientation of the seven conserved active site carboxylates composing metal ion binding sites A (Asp34, Asp86, Glu158 and Glu160) and B (Asp179, Asp181 and Asp233) were found to nicely align with those of superimposed HsFEN1:DNA structures published by Tsutakawa *et al*., (2017). Residues Asp86 and Asp233 are the only exemptions for D86N and D233N HsFEN1 mutant structures, respectively. The structure of HsFEN1: processed JT2+5 complex contained a bulb-like electron density associated with the coordinated by the seven active site carboxylates, catalytic metals and water molecules (Supplementary Information; Figure SI.8). These correspond to the presented in the crystallisation conditions, Ca^{2+} ions.

The presence of Ca^{2+} ions in the both the prepared protein:DNA mixtures (5 mM) and crystallisation conditions (~2.5 mM in the crystallisation drop) was expected to inhibit 5′-flap cleavage by HsFEN1, based on the previously reported 83% reduction of HsFEN1 activity at 1 mM $CaCl_2$ (Harrington and Lieber, 1994). However, despite the presence of higher concentrations of $CaCl_2$, the resulting HsFEN1:product DNA structures suggest that Ca^{2+} ions enabled some enzymatic activity to take place allowing 5′-flap cleavage by HsFEN1. Although overall the resulting complex conformations are identical to the already existing structures, this is the only to our knowledge HsFEN1:DNA structure with Ca^{2+} ions in HsFEN1 catalytic site, expanding the currently available range of HsFEN1:DNA structures and their crystallised co-factors (Table 4.3). $CaCl_2$ was previously used for crystallisation of a truncated HsFEN1 mutant, without any Ca^{2+} ions in HsFEN1 catalytic site possibly due to interference of the active site mutation (Tsutakawa *et al*., 2011).

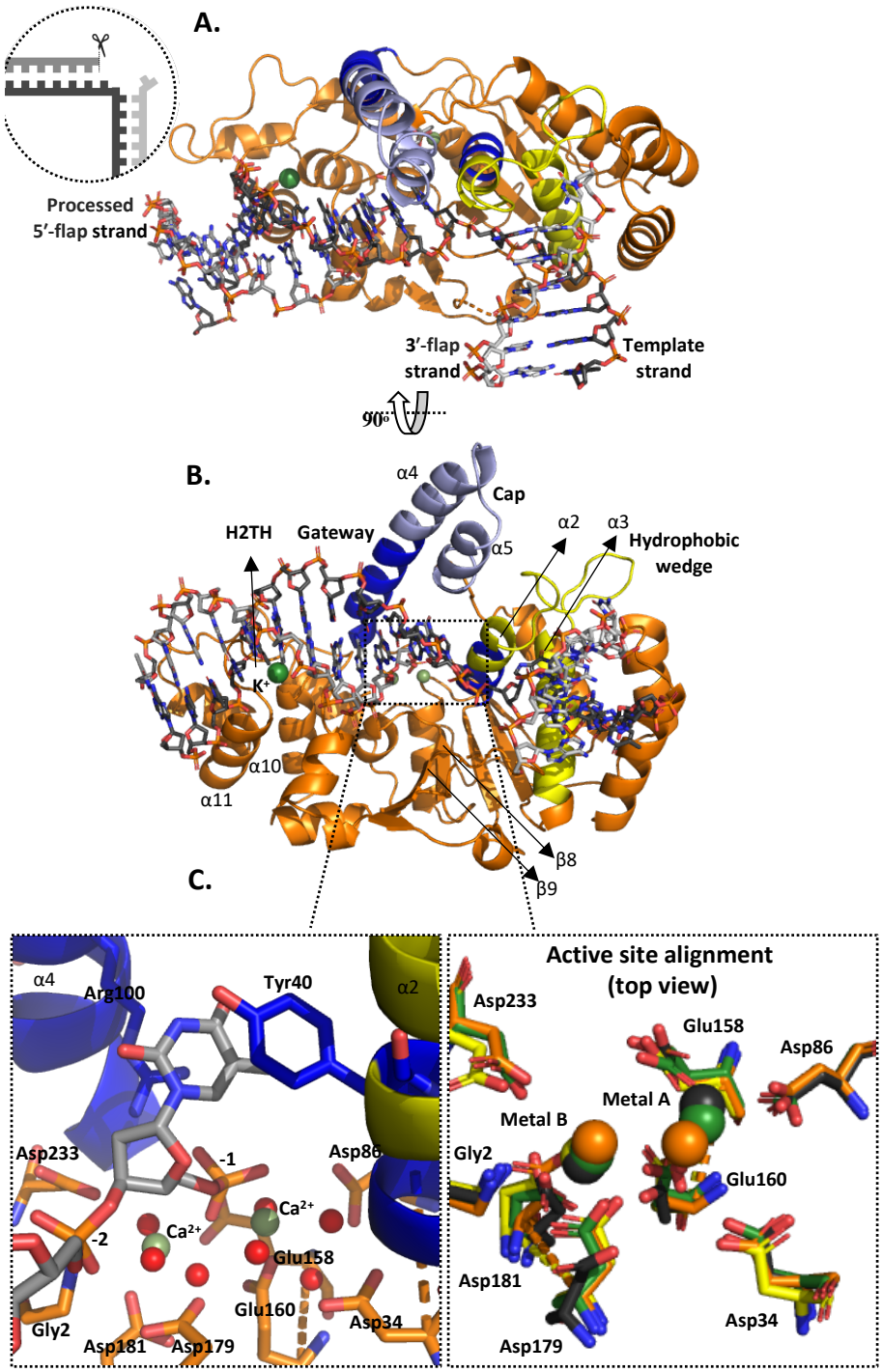


Figure 4.15 Structure of WT HsFEN1: processed JT2+5 DNA complex. (A) Top and (B) front views of WT HsFEN1 in complex with the processed JT2+5 DNA substrate, indicating binding of the resulting product DNA substrate after cleavage of the 5-nt long 5′-flap. (C) Left, close view of active site, showing α4 and α2 gateway helices along with the newly formed 5′-end phosphate oriented towards the active site Ca^{2+} metals. N-terminal Gly2, active site and key gateway residues are shown in sticks. Coordinating water molecules surrounding the individual active site metals are shown in red spheres. Right, alignment of WT HsFEN1: processed JT2+5 DNA active site (orange) with those of 5K97 (grey), 5UM9 (yellow) and 5KSE (green). (All panels) The overall WT HsFEN1 structure is shown in orange cartoon with some differentially coloured key structural elements: cap (light blue), helical gateway (blue) and hydrophobic wedge (yellow). The three strands of the product JT2+5 DNA: 3′-flap, 5′-flap and template strands are coloured in grey shades, as shown on the top left corner panel. H2TH-bridging K^+ and active site Ca^{2+} ions are shown in green and smudge green spheres, respectively.

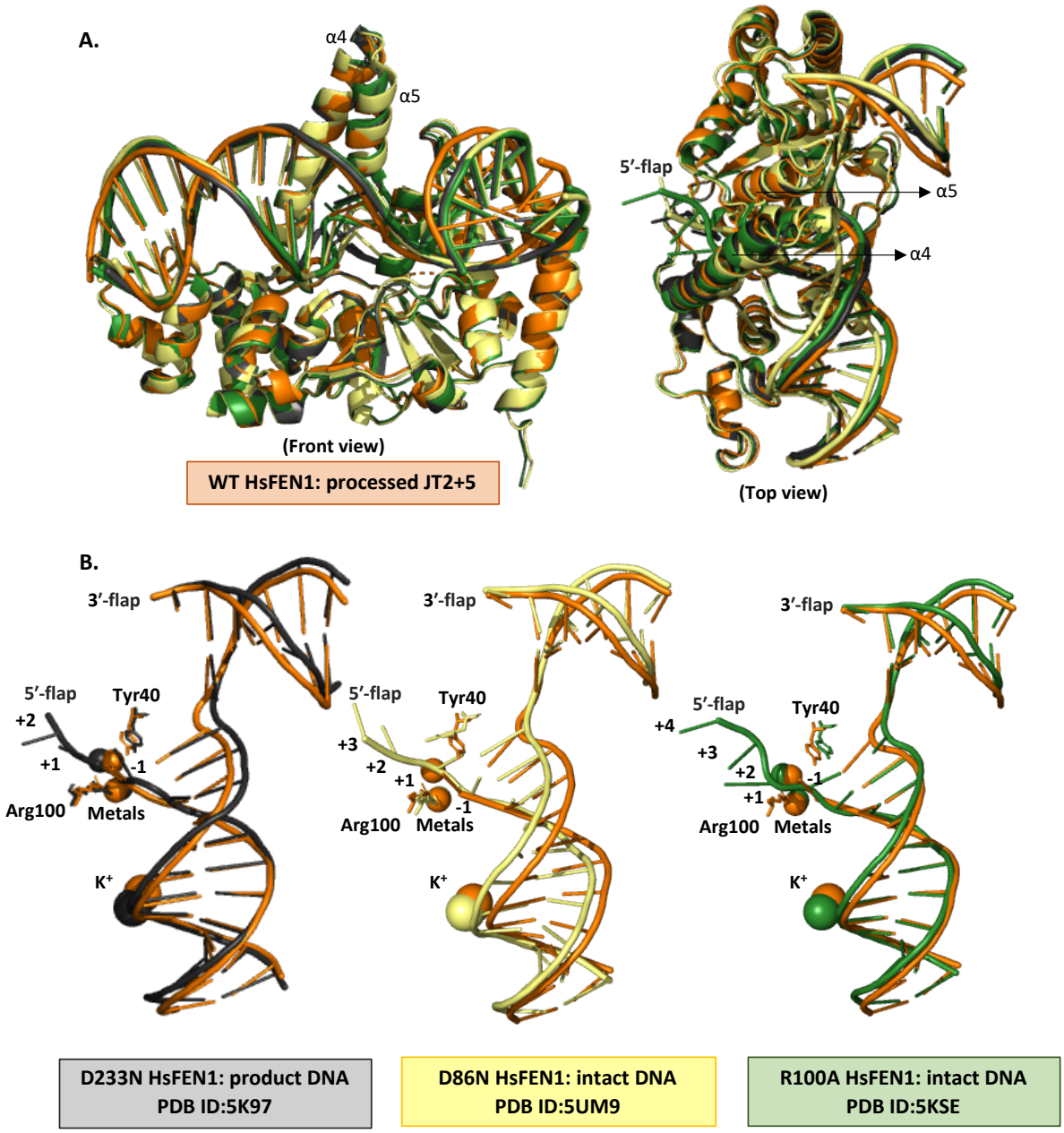


Figure 4.16 Structural comparison between the WT HsFEN1: processed JT2+5 and published HsFEN1: DNA complexes. (A) Front and top views of WT HsFEN1: processed JT2+5 structure, superimposed on the structures of previously published 5K97, 5UM9 and 5KSE mutant HsFEN1:DNA complexes. (B) Superimposition of Product JT2+5 oligonucleotide on the intact and product DNA substrates from the corresponding mutant HsFEN1 structures. The 5′-flap strands are numbered relative to the scissile phosphate (-1). Arg100 and Tyr40 rotamers are shown in sticks. (All panels) The individual HsFEN1 structures, their corresponding complexed DNA molecules, K^+ and active site metal ion(s) are coloured as shown on the panels above e.g. this project's WT HsFEN1:processed JT2+5:Ca^{2+}:K^+ complex is shown in orange.

4.4 Conclusion

Collectively, the structure of WT HsFEN1 in complex with processed JT2+5 substrate DNA, solved at 2.3 Å resolution, was found to have identical conformations to the already existing HsFEN1:product DNA structures in PDB. The observations of the dsDNA substrate's ability to bend 100°, "trapping" of the 1-nt long 3′-flap and positioning of 5′-flap towards the active site as a result of HsFEN1 binding at the DNA bend, provided further support to the unified for the members of FEN1 superfamily specificity requirements for substrate recognition and binding. Finally, co-crystallisation in the presence of Ca^{2+} ions in HsFEN1 catalytic site represents a unique observations of WT HsFEN1: processed JT2+5 complex, which despite the inhibitory effect of Ca^{2+} ions in HsFEN1 catalysis allowed some enzymatic activity to take place.

Chapter 5 – Pre-clinical toxicity studies on human and microbial FEN inhibitors

5.1 Introduction

5.1.1 Toxicity-related attrition of drug candidates: Current state and challenges

Following the hit-to-lead phase, short-listing of the identified drug candidates with good potential for development required determining their absorption, distribution, metabolism and excretion (ADME) properties. Despite the tremendous advances in drug discovery and development with the implementation of *in silico* and *in vitro* automated high-throughput screening strategies, accurate determination of the drug-induced toxicity at the early drug discovery stages, remains an important step. Whether a drug candidate derives from academia or pharmaceutical industry, the overall attrition rates in establishing FDA approval and commercialising a drug candidate remain high (Waring *et al*., 2015). In an effort to address the reasons behind the high attrition rates of small-molecule drug candidates, initial critical reviews focused on the insufficient pharmacokinetic profiles formed at the early phase of hit-to-lead optimisation. Toxicity-, as well as efficacy-related issues, are now considered as the main contributors to the high numbers of clinical failures, highlighting the need for the establishment of strong correlation links between the physicochemical properties of small-molecule drug candidates and their related failures (Leeson and Springthorpe, 2007).

However, considering that the toxicity induced by an individual small-molecule drug candidate depends highly on the mechanism of its primary target and/or its off-target effects, meaningful correlations between the physicochemical properties and toxicology-related attritions are unlikely to occur. In line with this, correlation analysis between the attrition and physicochemical characteristics for a substantially large dataset of small-molecule drug candidates, derived from one of the four world-leading pharmaceutical companies (AstraZeneca, Eli Lilly and Company, GlaxoSmithKline and Pfizer), revealed almost no correlations (Waring *et al*., 2015). Although these findings contrast with previous studies, which lacked diversity within compound sets examined, the compiled datasets collected by Waring *et al*., (2015) represent the largest unbiased dataset analysed to date with patented, marketed and under development small-molecule drug candidates.

With the lack of established correlations between the physicochemical properties of drug candidates and their toxicology-related outcomes, determination of drug-induced toxicities is crucial for the selection and design of molecules with increased survival probabilities at the later stages of drug development (e.g. hit expansion). Several toxicity assays have been developed or re-designed to date to estimate drug-induced toxicity. However, in their majority toxicity assays used at the early *in vitro* drug discovery stages are non-mechanism specific. Therefore, failing to address whether the induced toxicity is driven by the physicochemical characteristics of an individual compound *per se* or, for example, by inhibition of its main target or off-target proteins (Astashkina *et al*., 2012). Such assays typically involve end-point

measurements of cell viability and proliferation following treatment with small-molecule drug candidates in *in vitro* cultures. Whether these use primary cells or a cell line, a clear benefit for rapid and cost-effective estimation of dose-induced toxicity, at least at an early preliminary stage, is provided compared to the related animal models.

5.1.2 Development of FEN inhibitors as a novel class of antimicrobial drugs: DeFENition Ltd.

In addition to the anticancer potential of HsFEN1 inhibition, inhibitors of microbial FEN enzymes are today recognised as attractive drug candidates for the development of innovative antimicrobial therapies. Since the mid 1900s the discovery and development of antimicrobial drugs has tremendously advanced human health and modern medicine, their anthropogenic, widespread and uncontrolled use has led to the threatening evolvement of microbial defence and resistance mechanisms. With ≥700,000 deaths on a yearly basis, the estimated risk for a 14-fold increase by 2050 and the implicated socioeconomic cost, the emergence of drug-resistant pathogens and possible reappearance of epidemics are concerning vulnerability indicators for public health future (Garvey *et al*., 2016). Although since November 2019 the world's focus has been, understandably so, accumulated on SARS-CoV-2, WHO urges for immediate action to tackle antimicrobial-resistant pathogens and prevent the staggering predictions on morbidity and mortality.

With a shared (multi)drug-resistance amongst microbial pathogens affecting occasionally every available class of antimicrobials, WHO publishes a top-priority list of hardly treatable pathogens, some of which (*Acinetobacter baumannii*) require the use of last-resort antimicrobials associated with severe kidney and neurological complications (Gilbert *et al*., 2010; Savage, 2020). These, on top of an alarmingly dry clinical development pipeline of antimicrobial drugs, for which the typical attrition rates and lack of novelty with regards to the targeted mechanisms, hold back on the optimistic promises for universal therapeutic approaches preventing and predicting evolutionary drug-resistant mutations (Savage, 2020; Theuretzbacher *et al*., 2020). Based on this opportunity and the reports that FEN enzymes are required by all organisms, a spin-out company DeFENition Ltd was established by the University of Sheffield in 2016 aiming to deliver a novel class of antimicrobials through FEN inhibition.

Inhibition of microbial FEN enzymes is an innovative, unique to DeFENition Ltd, approach for the development of antimicrobials through manipulation of the indispensable for microbial viability mechanisms of DNA replication, repair and recombination. While the classical evolutionary paradigms and their induced resistance-conferring mutations are inevitable, selective inhibition of microbial evolvability factors has the potential of a fundamentally novel strategy to tackle the threatening emergence of antimicrobial-resistant pathogens (Ragheb *et al*., 2019). With an upward trend on studies confirming the essential roles of FEN enzymes, inhibition of the highly conserved microbial FEN enzymes may hold reduced susceptibility to resistant mechanisms and the powerful potential of a universal approach to avert microbial

pathogenicity and antimicrobial-resistant pathogens (Diaz *et al*., 1992; Fukushima *et al*., 2007; Zheng *et al*., 2007). This is due to the expectation that inhibition of bacterial FEN activity would lead to accumulation of stalled replication forks, increased amounts of ssDNA and replication stress, This would likely lead to fragmentation of the genome and thus, a bactericidal mode of action.

With DeFENition Ltd conducting extensive systematic *in silico* and physical screening in collaboration with the Sayers' laboratory, identification of small-molecule inhibitors against a range of microbial FEN enzymes has focused on those with the potential to differentiate between human and microbial FEN enzymes for ongoing development. Although literature on the pharmacological inhibition of microbial FEN enzymes has not been made available yet, ongoing research has led to a short-list of small-molecule inhibitors with strong antimicrobial activity over individual targets (DeFENition Ltd, unpublished). Physicochemical characterisation, prioritisation and optimisation of their most promising FEN inhibitors is now underway to maximise the success potential of downstream inhibitor design and development processes.

5.1.3 Rationale and aims

Whilst the *in vitro* inhibition of microbial and human FEN is the first central indicator for the selection of small-molecule inhibitors with antimicrobial and anticancer drug potential, further compound prioritisation and optimisation for potentially successful drug development also requires determination of their likely cytotoxic effects. Funded by DeFENition Ltd., a selection of leading microbial and human FEN inhibitors were evaluated for their possible haemolytic and cytotoxic effects, aiming to provide a reference dataset that will facilitate downstream compound filtering and design of molecules with improved physicochemical parameters. Given that the structures of FEN1 inhibitors, identified by DeFENition Ltd., are confidential, establishment of structure-toxicity relationships falls outside the scope of the current thesis.

5.2 RESULTS & DISCUSSION: *In vitro* haemolytic activity of FEN inhibitors

5.2.1 Haemolysis assay optimisation

Aiming to evaluate the haemolytic activity of short-listed FEN inhibitors as the first indicator for toxicity assessment and facilitate downstream compound prioritisation and optimisation, initial experiments focused on the in-house optimisation of a haemolysis assay. Several publications have described similar methodologies for the assessment of drug-induced haemolysis. However, variations in the published assay conditions, such as erythrocytes source and/or suspension buffer, may be expected to affect the resulting haemolysis levels determined. Thus, comparisons between haemolysis levels reported using different protocols could be misleading.

Optimisations focused on determination of the relevant RBCs dilution, required for a detectable and reproducible signal upon lysis, and the tolerated solvent (DMSO) concentration for subsequent drug testing. Given the preliminary *in vitro* stage, erythrocytes derived from horse blood were selected for initial haemolysis evaluation. Although the selected FEN inhibitors are intended to be developed for eventual human use, the use of horse erythrocytes with established structural and compositional membrane similarities to that of human erythrocytes, is expected to provide valuable preliminary data on the induced toxicity effects of the tested compounds (Ebbensgaard *et al*., 2018; Zwaal and Vandeene.Ll, 1968). Combinations of 1-6 % (v/v) RBCs suspensions in PBS and 0-10% (v/v) DMSO, were evaluated for haemolysis, following haemoglobin release spectroscopically, after incubating for 1 hr at 37°C. For each combination, haemolysis was also evaluated in the presence of 10% (v/v) Triton X-100 as the maximum/100% lysis control, an established positive control for haemolysis assays with well-studied haemolytic mechanisms (Chernitsky and Senkovich, 1997).

In the absence of 10% (v/v) Triton X-100, no significant difference was detected between the $A_{540\ nm}$ measurements taken at increasing DMSO concentrations, compared to their respective RBCs only (No DMSO) control (Figure 5.1). The $A_{540\ nm}$ measurements remained unaffected at all RBCs suspension concentrations tested, indicating that DMSO *per se* has no haemolytic effect under the conditions and concentration range explored (Figure 5.1). Despite these, the most frequently adopted DMSO concentrations in standardized haemolysis protocols do not exceed 1% (v/v) (Ebbensgaard *et al*., 2018; Greco *et al*., 2020).

On the other hand, in the presence of 10% (v/v) Triton X-100 $A_{540\ nm}$ measurements, corresponding to complete haemoglobin release as a result of RBCs lysis, were found to linearly increase with the increasing RBCs and DMSO concentrations (Figure 5.1). However, given the negligible effect of DMSO on horse RBCs lysis at 1 hr incubation time (vehicle controls), the increased absorbances are attributed to the increasing concentrations of RBCs in the different suspensions tested. Detectable signals were produced at every concentration of RBC suspension tested in the presence of 10% (v/v) Triton X-100, with, as expected, the lowest $A_{540\ nm}$ at 1% (v/v) and the highest at 5 % (v/v) (Figure 5.1). Based on these and the standardized haemolysis assay methodologies, 2% (v/v) RBCs suspension and 1% (v/v) DMSO were selected as the optimal conditions for subsequent drug-testing, limiting DMSO concentration and maximising cost-effectiveness.

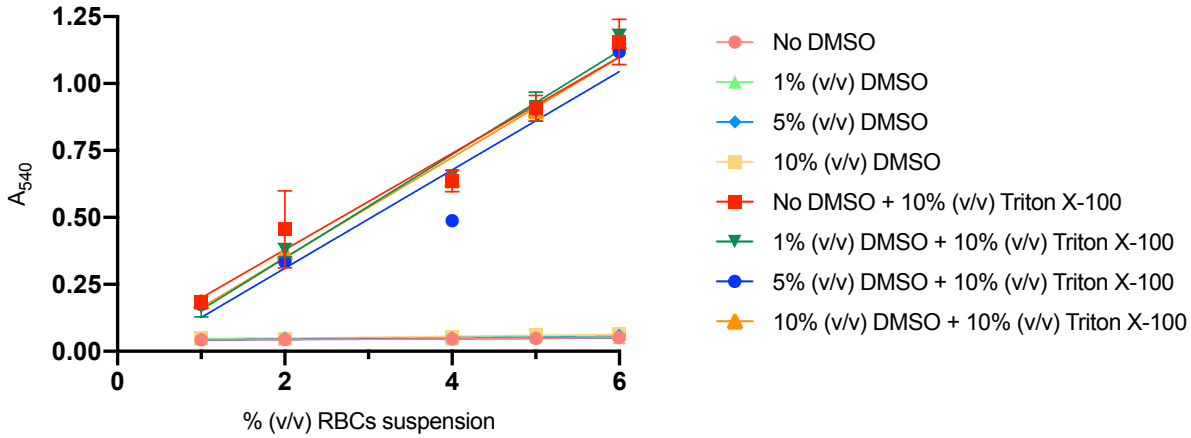


Figure 5.1 Optimisation of haemolysis assay: evaluation of the optimal DMSO and RBCs concentrations. Various RBCs suspensions, prepared in PBS, were incubated for 1hr at 37°C with 0, 1, 5 and 10% (v/v) DMSO, in the presence and absence of 10% (v/v) Triton X-100. Raw A_{540} values, corresponding to the released haemoglobin levels, were plotted against the increasing RBCs suspension concentrations for direct comparison. For each RBCs suspension tested, a no DMSO and no Triton X-100 control, indicative of no/minimum haemolysis (baseline), was carried out. Data were plotted by linear regression using GraphPad Prism 8.2 with R^2 values ≥0.9 for each condition tested in the presence of 10% (v/v) Triton X-100. Error bars represent the SD of the mean $A_{540\ nm}$ values, calculated from 4 technical replicates performed in a single experiment (i.e. $N = 1$, $n = 4$).

5.2.2 Haemolytic activity of FEN inhibitors in 2% (v/v) RBCs suspension

Based on the inhibitory activity patterns of FEN inhibitors identified through the ongoing DeFENition Ltd screening programme, an initial selection of 51 compounds was constructed for determination of their potential cytotoxic effects. Evaluation of inhibitor-induced haemolysis was initially carried out using the optimised 2% (v/v) horse RBCs suspension in PBS and 1% (v/v) DMSO. Horse RBCs were incubated for 1 hr at 37°C with the selected FEN inhibitors at four different concentrations, ranging from 1-1000 µM. Following the optimised standardized methodology for the evaluation of drug-induced haemolysis, the released haemoglobin levels corresponding to the extent of haemolysis were determined in the supernatant spectrophotometrically at 540 nm. Data were normalised against a negative/vehicle and positive/10% Triton X-100 control, indicative of 0% and total haemolysis, respectively. Similar to high-throughput screening and the requirement for multiple in parallel assays to determine the possible haemolytic effect of the selected FEN inhibitor range, consistency between control A_{540} values was maintained (Judson *et al*., 2013; Table 5.1).

The great majority of the evaluated FEN inhibitors were found to cause no detectable haemolysis under the condition used, even at the highest concentration tested (Figure 5.2). In each case, visual inspection of the 96-well experiment plates confirmed their non-haemolytic effect post-1 hr incubation at 37°C (Figure 5.3). Only 17 compounds were found to induce ≥25% haemolysis, from which 16 were shown to induce ≥50% haemolysis at 1000 µM (Figure

5.2). Detectable haemolysis was generally found to lie at concentrations ≥100 μM. However, given the lack of points between 100-1000 μM concentration, accurate estimation of half-maximal responses was not possible for those compounds with significant haemolytic activity at this stage. Although these data did not enable direct comparisons of the respective EC_{50} values, they provided meaningful information for the preliminary discrimination between non-haemolytic and moderately/severely haemolytic FEN1 inhibitors.

Table 5.1 Summary of *Z′*, average $A_{540\ nm}$ values and their respective standard deviations per haemolysis assay performed, using a 2% (v/v) RBCs suspension

Assay no.	Average $A_{540\ nm}$ ±SD per assay		*Z′*[2]
	Positive/ 10% (v/v) Triton X-100[1]	Negative/ vehicle control[1]	
1	0.6±0.04	0.05±0.004	0.8
2	0.6±0.07	0.05±0.009	0.6
3	0.7±0.03	0.05±0.003	0.8
4	0.5±0.05	0.05±0.005	0.6
5	0.6±0.03	0.05±0.004	0.8
6	0.4±0.05	0.04±0.002	0.6
7	0.4±0.01	0.05±0.002	0.9
8	0.4±0.01	0.04±0.008	0.9
9	0.4±0.01	0.05±0.003	0.9
Average $A_{540\ nm}$±SD	0.5±0.1	0.05±0.004	0.7

[1]Positive/ 10% (v/v) Triton X-100 control and negative/ vehicle control are indicative of total and 0% haemolysis, respectively.

[2]*Z′* values were calculated per assay using the formula $Z' = 1-[3(SD_{POS}+SD_{NEG})/(|\bar{x}_{POS}-\bar{x}_{NEG}|)]$ and assay quality was determined as "marginal", "ideal" and "excellent" based on $Z' < 0.5$, $1 > Z' \geq 0.5$ and $= 1$, respectively (Zhang *et al*., 1999).

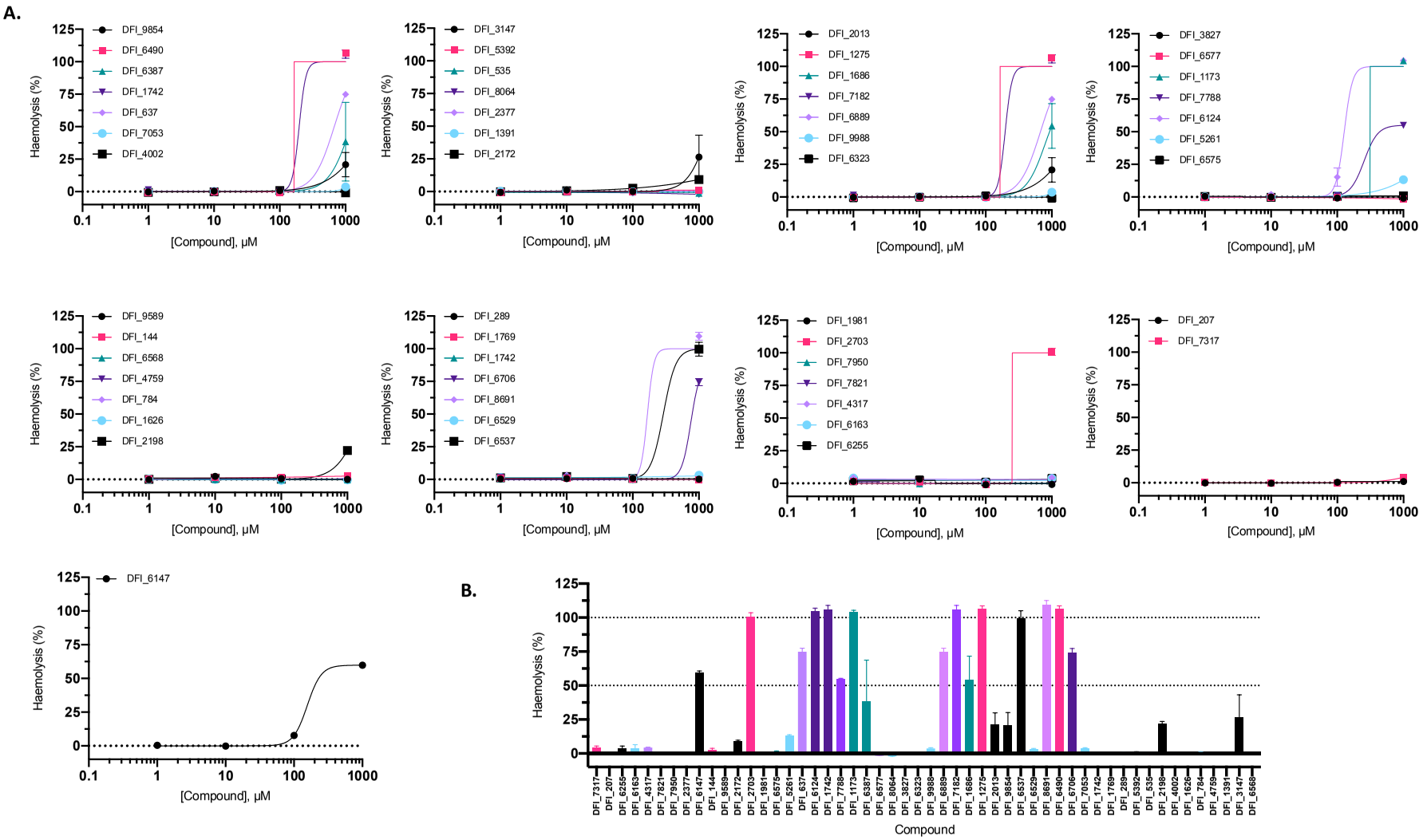


Figure 5.2 Haemolytic effect of selected FEN inhibitors on 2% (v/v) horse RBCs suspension. (a) Dose-response curves of inhibitor-induced haemolysis at concentrations ranging from 1-1000 µM and (b) Summarised haemolytic effect of FEN inhibitors at 1000 µM. Each value represents the normalised mean of three technical repeats (n=3), based on negative/vehicle and positive/10% (v/v) Triton X-100 controls (n=6). Dose-response curves were generated using nonlinear regression [Inhibitor] vs. response -Variable slope (four parameters) model in GraphPad Prism version 8.2. $Z' \geq 0.5$ for all assays performed.

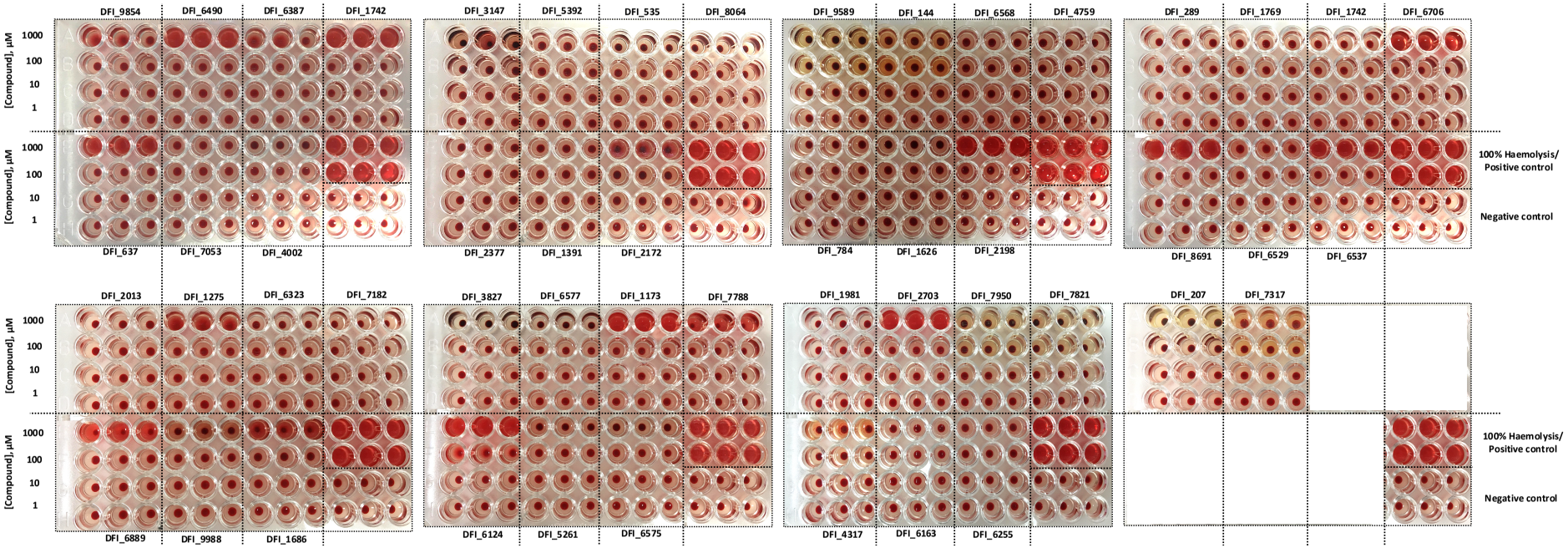


Figure 5.3 Corresponding 96-well experiment plates reflective of FEN inhibitor-induced haemolysis on 2% (v/v) RBCs suspension. Visual assessment of the haemolytic effect of selected FEN inhibitors at 1-1000 µM concentrations, compared to negative/vehicle and positive/10% (v/v) Triton X-100 controls. Clear RBCs pellets are indicative of non-lysed RBCs, whereas reduction of RBCs pellet size and supernatant discoloration are indicative of haemolysis induction. Three technical repeats per concentration tested were carried for every selected FEN inhibitor over a single experiment (i.e. $N = 1$, $n = 3$). For controls, six technical repeats were performed per experiment (i.e. $N = 1$, $n = 6$).

5.2.3 Haemolytic activity of FEN inhibitors in whole horse blood

Aiming to assess whether the observed haemolytic profiles of short-listed FEN inhibitors (Section 5.2.2) are replicated when incubated in an environment that better recapitulates mammalian blood conditions *in vivo*, a second round of haemolysis assays was carried out. Following the pre-described standardized methodology for the evaluation of drug-induced haemolysis, the haemolytic activity of short-listed FEN inhibitors was also evaluated using whole horse blood, instead of the 2% (v/v) RBCs suspension in PBS. Released haemoglobin levels were determined spectrophotometrically using, this time, the appropriately diluted supernatants, taking into account the instrument's dynamic range at which detectable and reproducible signals can be measured.

Encouragingly, none of the FEN inhibitors tested was found to exhibit any significant haemolytic activity (Figures 5.4-5.5). For the great majority of the evaluated FEN inhibitors (64%; Section 5.2.2), this is expected and in accordance with the previously observed haemolytic profiles. However, for the remaining 36%, which exhibited some level of haemolytic effect in assays performed in 2% (v/v) RBCs suspension, it is contradictory. Given the appropriate use of negative/vehicle and positive/10% Triton X-100 controls, the latter of which is reflective of total haemolysis, the opposing observations are more likely attributed to respective microenvironment tested. Potential interactions between FEN inhibitors and the abundant plasma proteins, such as albumin, are likely to eliminate their haemolytic effect, masking a non-toxic profile. Such proteins were washed off during preparation of the 2% (v/v) RBCs suspension, eliminating their interference and enabling FEN inhibitors to exhibit their possible haemolytic effect.

Although this is a commonly seen phenomenon at the early drug discovery stages, it can have a dramatic effect on the biological activity, mode and rate of action of an investigated drug candidate (Wani *et al*., 2018; Zhang *et al*., 2012). Given the data presented here, and the potential of the short-listed FEN1 inhibitors for the development of novel antimicrobial and anticancer drugs, further investigation of their interaction capabilities with plasma proteins might be critical for downstream drug design and development.

Table 5.2 Summary of *Z*, average $A_{540\ nm}$ values and their respective standard deviations per haemolysis assay performed, using whole horse blood erythrocytes

Assay no.	Average $A_{540\ nm}$±SD per assay		Z'[2]
	Positive/ 10% (v/v) Triton X-100[1]	Negative/ vehicle control[1]	
1	1.02±0.15	0.09±0.001	0.5
2	1.18±0.11	0.09±0.004	0.5
3	1.04±0.08	0.11±0.001	0.7
4	1.01±0.12	0.12±0.006	0.6
5	1.09±0.08	0.11±0.002	0.8
6	0.96±0.14	0.11±0.003	0.9
7	1.11±0.11	0.13±0.004	0.8
8	1.4±0.2	0.16±0.009	0.5
9	1.02±0.15	0.04±0.004	0.5
Average $A_{540\ nm}$±SD	1.09±0.13	0.1±0.03	0.6

[1]Positive/ 10% (v/v) Triton X-100 control and negative/ vehicle control are indicative of total and 0% haemolysis, respectively.

[2]Z' values were calculated per assay using the formula $Z' = 1-[3(SD_{POS}+SD_{NEG})/(|\bar{x}_{POS}-\bar{x}_{NEG}|)]$ and assay quality was determined as "marginal", "ideal" and "excellent" based on $Z' < 0.5$, $1 > Z' \geq 0.5$ and $= 1$, respectively (Zhang *et al*., 1999).

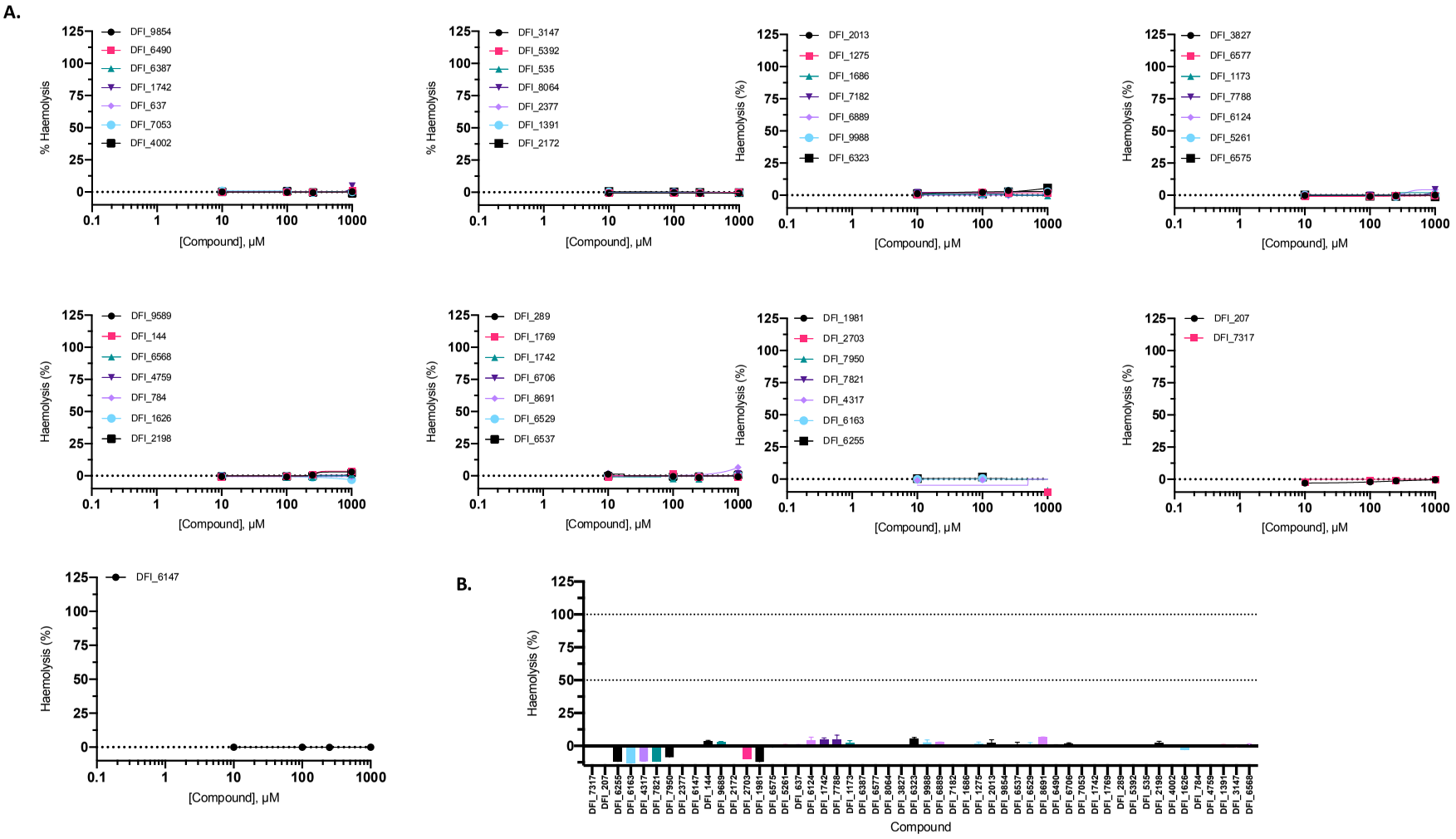


Figure 5.4 Haemolytic effect of selected FEN inhibitors on whole horse blood erythrocytes. (a) Dose-response curves of inhibitor-induced haemolysis at concentrations ranging from 1-1000 µM and (b) Summarised haemolytic effects of FEN inhibitors at 1000 µM. Each value represents the normalised mean of three technical repeats (n=3), based on negative/vehicle and positive/10% (v/v) Triton X-100 controls (n=6). Dose-response curves were generated using nonlinear regression [Inhibitor] vs. response -Variable slope (four parameters) model in GraphPad Prism version 8.2. $Z' \geq 0.5$ for all assays performed.

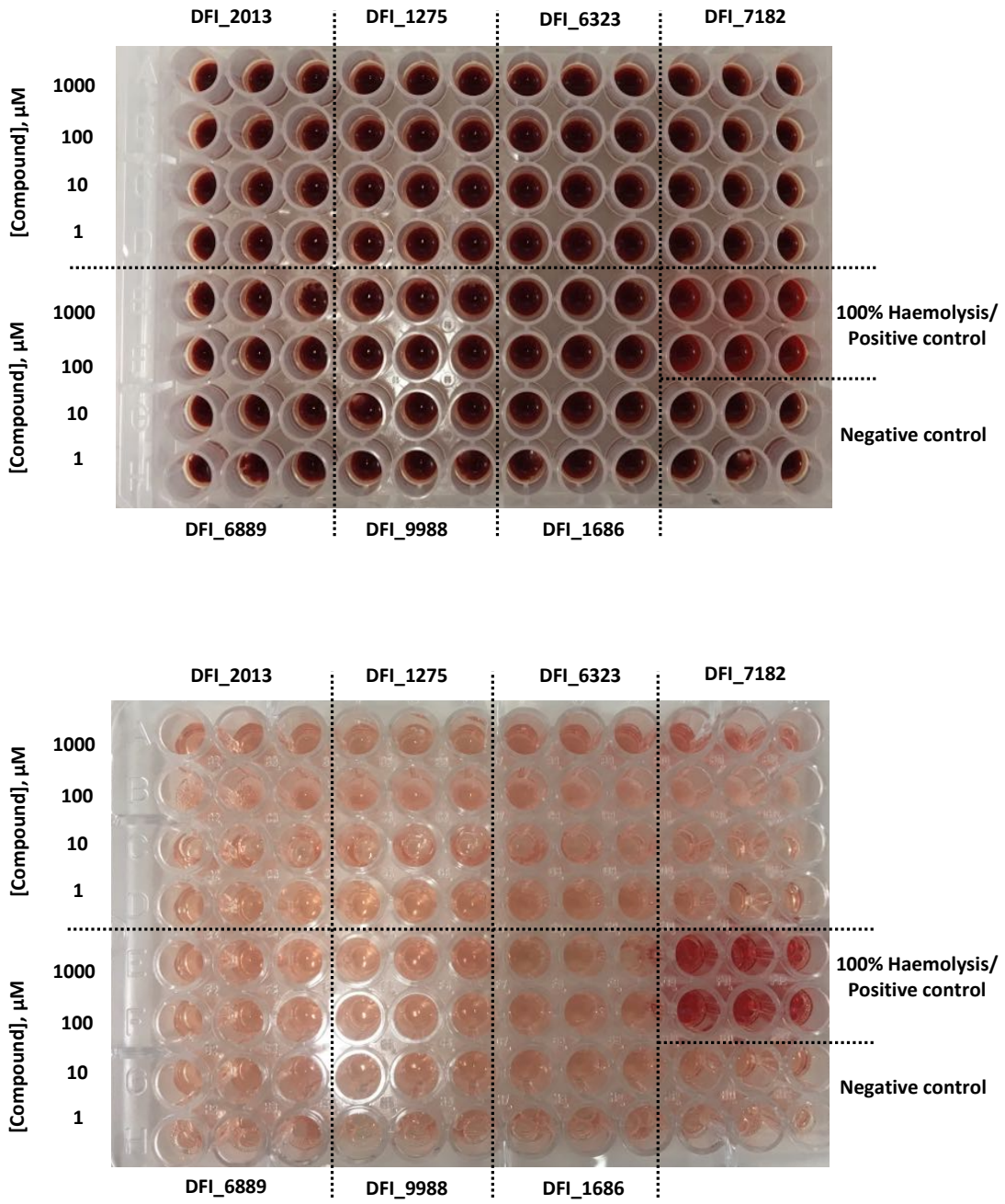


Figure 5.5 Representative example of FEN inhibitor-induced haemolysis in whole horse blood erythrocytes. Visual assessment of the haemolytic effect of selected FEN inhibitors at 1-1000 µM concentrations, compared to negative/vehicle and positive/10% (v/v) Triton X-100 controls. (a) Whole horse blood plate post-1 hr incubation and (b) experiment plate containing the undiluted supernatant. RBCs pellets are indicative of non-lysed RBCs. Reduced size of RBCs pellet sand supernatant discoloration are indicative of haemolysis. Three technical repeats per concentration tested were carried for every selected FEN inhibitor over a single experiment (i.e. $N = 1, n = 3$). For controls, six technical repeats were performed per experiment (i.e. $N = 1, n = 6$).

5.2.4 Suitability of 10% (v/v) Triton X-100 as positive/total haemolysis control

The non-ionic detergent Triton X-100 is an established positive control in standard haemolysis assays. Although reviewing previous haemolysis assay protocols suggested the wide use of 10% (v/v) concentration for maximum/100% haemolysis, the used concentrations of Triton X-100 were found to vary from 1-10% (v/v) in published toxicity studies (Greco *et al*., 2020; Rodriguez *et al*., 2014). Aiming to examine whether exposure of RBCs to different Triton X-100 concentrations induce significantly different levels of haemolysis, non-reflective of the 100% haemolysis for positive control, Triton X-100 titration was carried out.

Using the optimised 2% (v/v) horse RBCs suspension, the haemolytic effect of Triton X-100 at 0, 1, 2, 3, 4, 5, 10 and 20 % (v/v) concentrations was evaluated, following the exact same methodology described in Materials and Methods (Section 2.14.1; 1 hr incubation at 37°C). Generally, comparison of the $A_{540\ nm}$ measurements revealed analogous levels of induced haemolysis between the different Triton X-100 concentrations compared to the 10% (v/v) Triton X-100, used as positive control. The A_{540} values were also analogous to those previously observed for the positive/100% haemolysis controls in the presence of 2% (v/v) RBCs suspension (Figures 5.6-5.7). Whilst these data confirm the reproducibility of the positive/100% haemolysis control used throughout the reported haemolysis assays, they also suggest that the use of lower Triton X-100 concentrations (1-10%), as positive controls, would have no/insignificant effect over the observed drug-induced toxicities (*P* values>0.05. Of course, these data are dependent on the respective microenvironment tested and therefore only reflective of the susceptibility of horse RBCs at 2% (v/v) suspension after 1 hr exposure to Triton X-100 at 37°C.

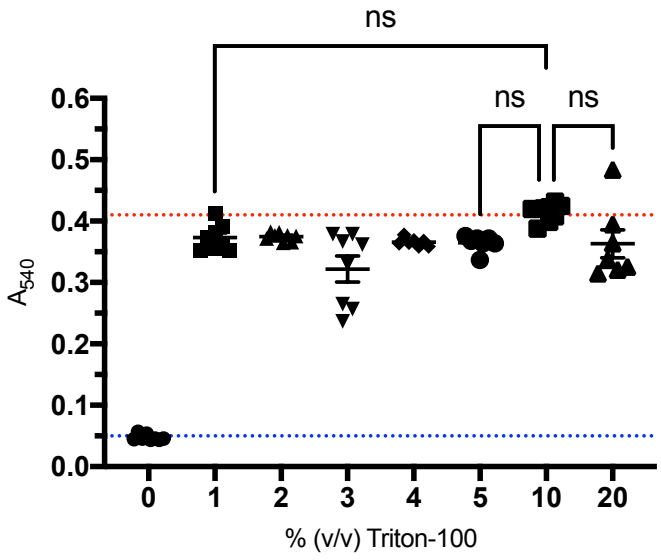


Figure 5.6 Susceptibility of horse RBCs to the haemolytic action of various Triton X-100 concentrations. A_{540} measurements, corresponding to the respective haemoglobin release levels, were taken in the presence of 1, 2, 3, 4, 5, 10 and 20 (v/v) Triton X-100 concentration, along with a no Triton X-100/negative control. Blue and red dotted lines are representative of the mean $A_{540\ nm}$ values of no Triton X-100/negative control and 10% (v/v) Triton X-100, typically used to reflect the 100% haemolysis control in drug-testing. Raw $A_{540\ nm}$ values were plotted against the tested surfactant concentrations. Data analysis was carried out by ordinary one-way ANOVA using GraphPad Prism 8.2 (*P* values>0.05). Error bars represent the SD of the mean A_{540} values, calculated from 8 technical repeats performed in a single experiment (i.e. $N = 1$, $n = 8$).

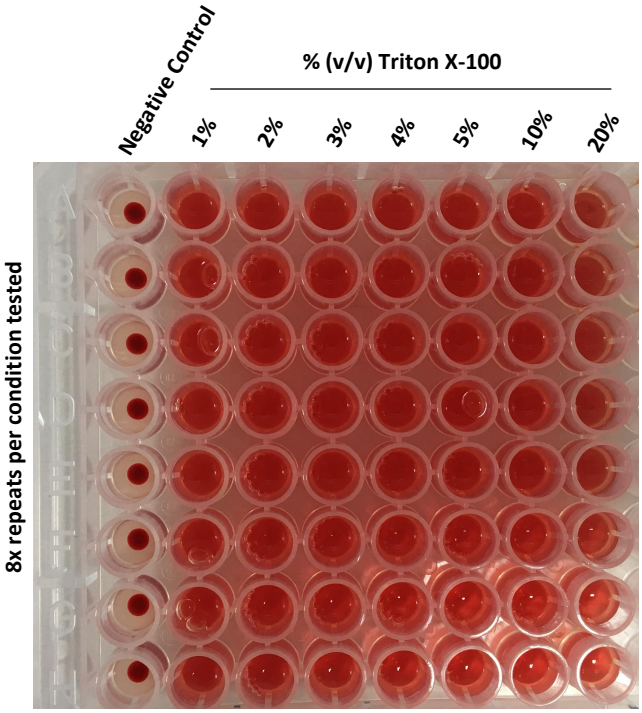


Figure 5.7 Haemolytic activity of various Triton X-100. Dilutions of detergent were examined in 96-well experiment plate after 1 hr incubation at 37°C. Each column is representative of a single Triton X-100 concentration alongside the no Triton X-100/negative control. Clear RBCs pellets are indicative of non-lysed RBCs, whereas reduction of RBCs pellet size and supernatant discoloration are indicative of haemolysis induction. Rows represent the 8x technical repeats performed at each condition tested (i.e. $N = 1$, $n = 8$).

5.2.5 Evaluation of Triton X-100 haemolysis as a reference (control) lytic molecule

Given the low versus negligible FEN inhibitor-induced toxicities observed in the presence of 2% (v/v) horse RBCs and whole horse blood erythrocytes, a Triton X-100 titration was also carried out at concentrations analogous to that used for drug testing (1-1000 μM). Although a Triton X-100 titration was previously carried out to evaluate the level of Triton-induced haemolysis, reflective of a total 100% haemolysis and suitable for the use as positive control, the required concentration range to generate a comparable dose-response curve of a known haemolytic agent must reflect that primarily used for FEN inhibitor testing.

Whole horse blood erythrocytes and two different horse RBCs suspensions, including the previously used 2% (v/v) and a higher one at 5% (v/v), were incubated for 1 hr at 37°C with increasing Triton X-100 concentrations, ranging from 1 to 1000 μM. In each case, a negative/vehicle and positive/100% haemolysis control was carried out, similar to those used during FEN inhibitor testing. Released haemoglobin levels were quantified and normalised for direct comparison, following the standard methodology described in Materials and Methods (Section 2.14.1).

Although the lack of points between 100-1000 μM Triton X-100 concentration, did not enable accurate determination of IC_{50}/EC_{50} values, the observed haemolytic profiles were remarkably similar, without any detectable haemolysis at concentrations ≤100 μM (Figure 5.8-5.9). In all cases, Triton X-100 seemed to exhibit its haemolytic activity between 100-1000 μM concentration and reach total haemolysis at about 1000 μM, which was also detectable by simple visual inspection of the assay plate (Figure 5.8-5.9). Given the use of the pre-described standardized methodology, the resulting haemolytic profiles of Triton X-100, an agent with

established haemolytic effect, could serve as positive control assays to those of FEN inhibitor testing. In addition, the observed consistency between the three different RBC conditions examined, suggest that the use of the pre-described methodology can lead to reproducible haemolytic profiles. Haemolytic profiles, which, in the case of the evaluated FEN inhibitors seemed to change from mild to negligible. Possible binding interactions between the selected FEN1 inhibitors with abundant plasma proteins, can perhaps explain the abolishment of their induced haemolytic activity in whole horse blood erythrocyte assays.

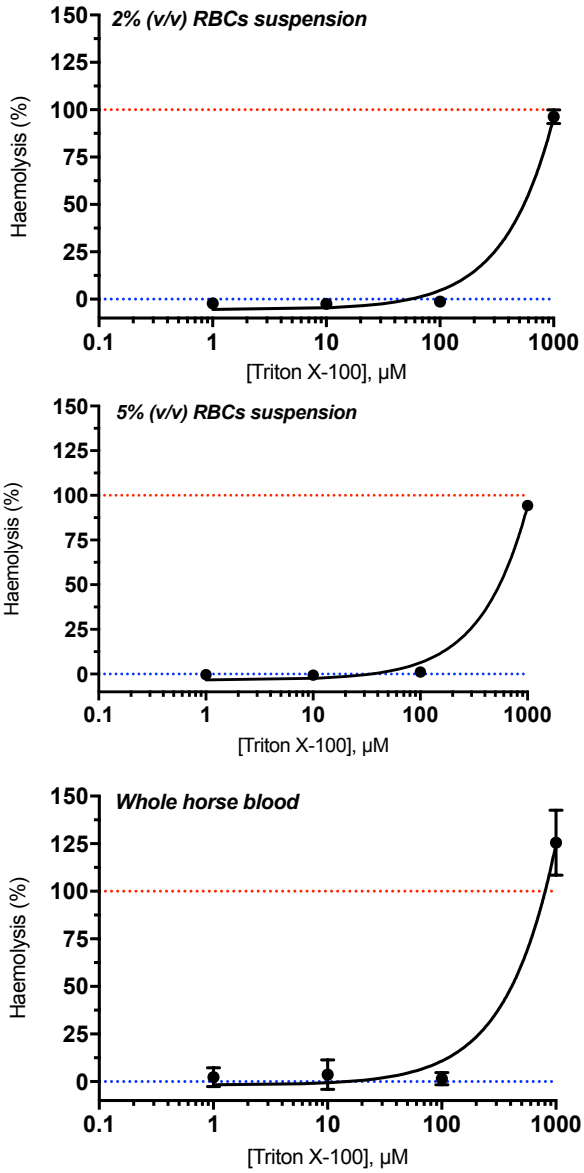


Figure 5.8 Haemolytic activity of Triton X-100 on 2-5% (v/v) RBCs suspensions and whole blood erythrocytes. Triton X-100 haemolytic effect over whole horse blood erythrocytes, 2% and 5% (v/v) RBCs suspension, at 1-1000 µM concentrations. Each value represents the normalised mean of three technical repeats (i.e. $N = 1$, $n = 8$), based on negative/vehicle and positive/10% (v/v) Triton X-100 controls (i.e. $N = 1$, $n = 8$). Data were plotted using nonlinear regression [Inhibitor] vs. response -Variable slope (four parameters) model in GraphPad Prism version 8.2. $Z' \geq 0.5$ for all assays performed.

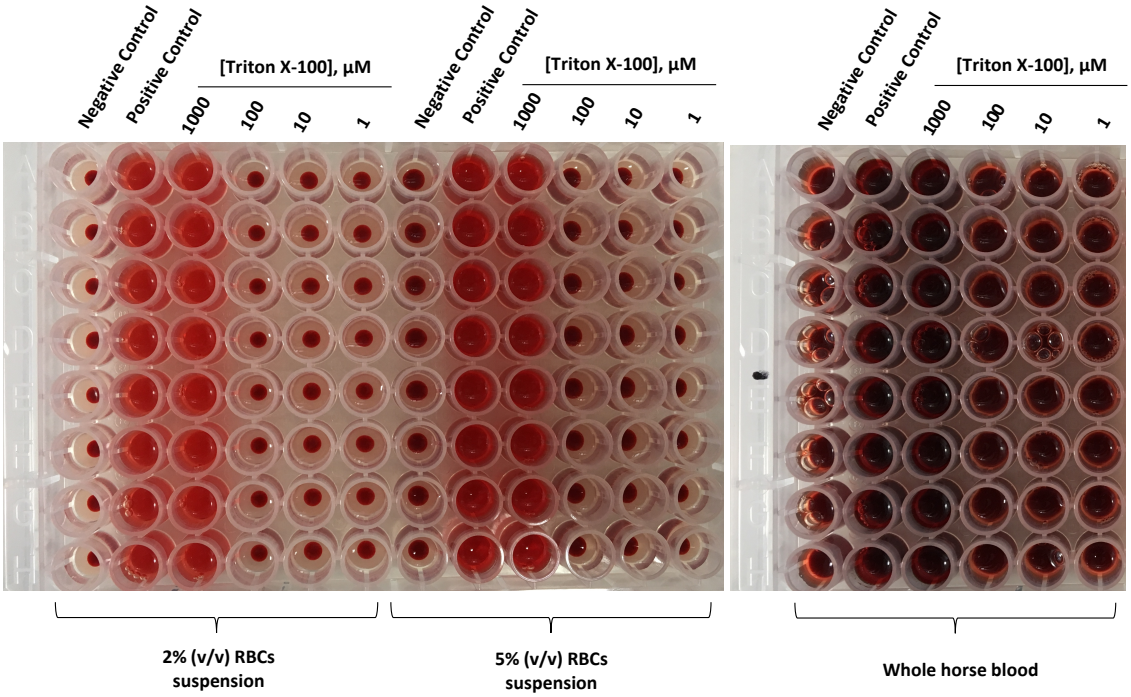


Figure 5.9. Haemolytic activity of Triton X-100 on 2-5% (v/v) RBCs suspensions and whole blood erythrocytes. Representative 96-well experiment plates post-1 hr incubation at 37°C with FEN inhibitors. Each column is representative of a single Triton X-100 concentration alongside the no Triton X-100/negative control. Clear RBCs pellets are indicative of non-lysed RBCs, whereas reduction of RBCs pellet size and supernatant discoloration are indicative of haemolysis induction. Rows represent the 8x technical repeats performed at each condition tested (i.e. $N = 1, n = 8$).

5.3 *In vitro* cytotoxicity of FEN inhibitors

5.3.1 Evaluation of FEN inhibitor-induced cytotoxicity in epithelial HCT-116 cell line

In parallel to the evaluation of FEN inhibitor-induced haemolysis, *in vitro* cytotoxicity was also evaluated in human HCT-116 colorectal cancer cell line with established selective sensitivity to FEN inhibition (Ward *et al.*, 2017). Given the ongoing systematic inhibitor screening in Sayers' laboratory, an updated list of selected FEN inhibitors was evaluated for the potential cytotoxic effect of individual compounds using Resazurin cell viability assay. Adherent cells were typically treated with increasing drug concentrations for 24 hr at 37°C, and cell viability was estimated based on the relative fluorescence intensity of Resazurin's metabolic product (resorufin; $\lambda_{ex}/\lambda_{em}$=540/590 nm). Dose-response curves were generated using the normalised RFU values, against a positive/vehicle (100% cell viability) and negative/1% (v/v) Triton X-100 controls (100% cell death).

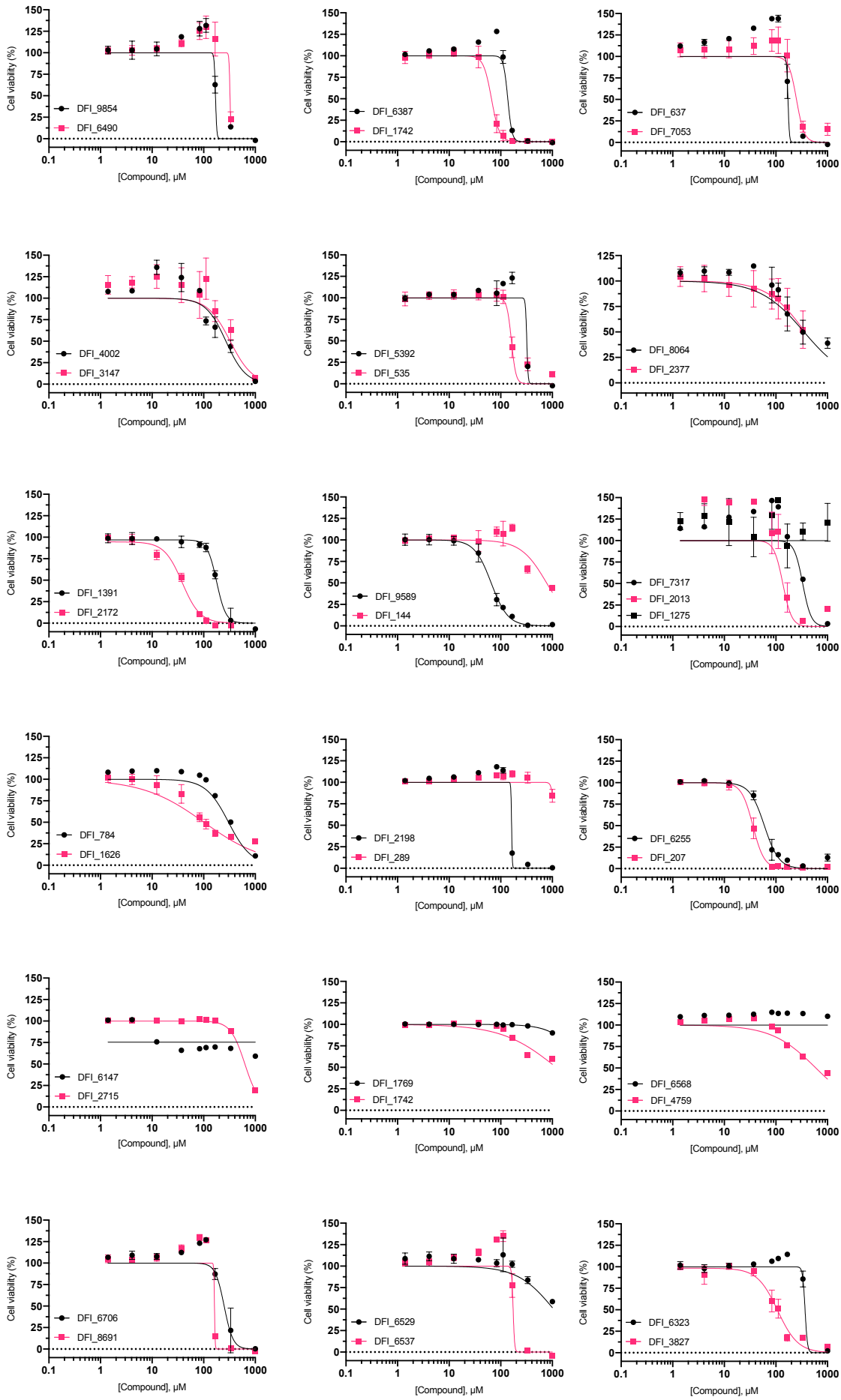
Whilst the DMSO solvent alone had no detectable cytotoxic effect over the tested cell lines at 1% (v/v) concentration, raw fluorescence values were also maintained throughout the number

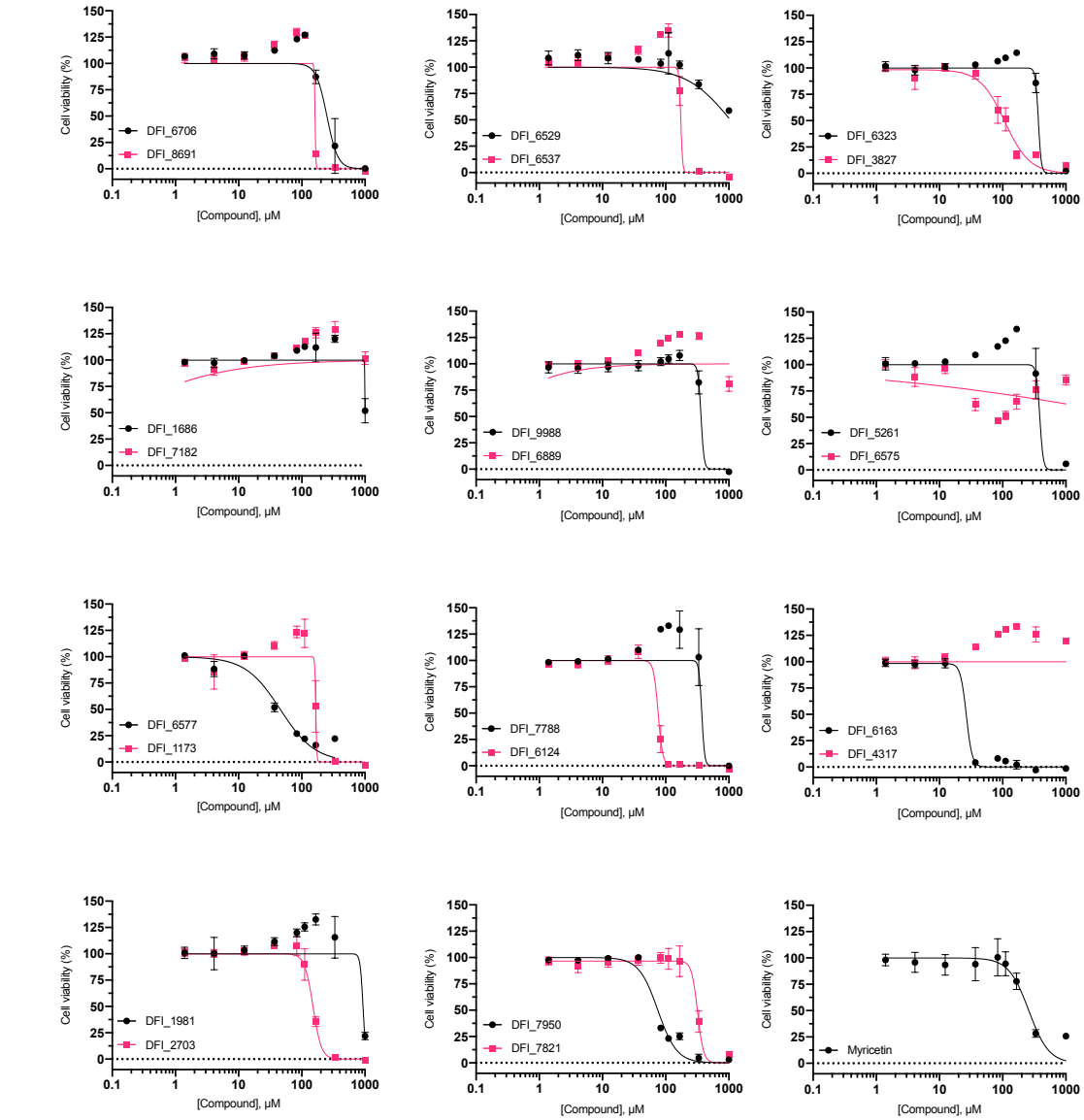
of assays performed to ensure comparability and reproducibility (data not shown). Overall, the cytotoxic activities of the evaluated FEN1 inhibitors over HCT-116 cells seemed to vary from no/negligible (GI_{50} ≥1000 μM) to highly cytotoxic (lower GI_{50} observed = 26.5 μM; DFI_6163) (Figure 5.10; Table 5.3). Out of the 53 tested FEN1 inhibitors, short-listed for cytotoxicity evaluation based on their respective IC_{50} in FRET assay (DeFENition Ltd, unpublished; Table 6.3), only 24 were found to have a remaining cell viability ≥50% posttreatment with 333.3 μM inhibitor concentration. The other 29 compounds were found to have increased cytotoxic effects, resulting in cell viabilities ≤50%. Among these, 15 compounds seemed to completely eliminate (≤5%) cell viability percentages, suggesting a strongly cytotoxic activity over HCT-116 cells.

Along with the short-listed FEN inhibitors, myricetin was also evaluated for its cytotoxic activity over HCT-116 cells. Myricetin is an established HsFEN1 inhibitor with the highest inhibition potency among the bioflavonoid class of HsFEN1 inhibitors reported to date (Ma *et al*., 2019). With previous studies reporting an *in vitro* IC_{50} of around 0.7-12 μM and 10% cytotoxic activity over HT-29 cells (also colorectal cancer cell line) at 64 μM post-24 hr treatment, myricetin cytotoxicity was evaluated alongside the short-listed FEN1 inhibitors for comparison (Ma *et al*., 2019; Dr Sarah L. Oates, unpublished). Using the exact same conditions used for the evaluation of candidate FEN inhibitors (24 hr treatment and 1-1000 μM test range), myricetin was found to have a similar toxicity to that previously determined by Ma *et al*., (2019) with GI_{50} of ∽64 μM. However, the expanded drug concentration range with which cells were treated and the 16-fold higher maximum concentration tested, led to induced cytotoxicity as high as 70±3.6 % at 333.3 μM (Figure 5.10).

The respective GI_{50} values of short-listed FEN inhibitors and myricetin, corresponding to their cytotoxic activities over HCT-116 cells are summarised in Table 5.3. Despite the evaluation of multi-point concentration responses, double than what was previously used in haemolysis assays, accurate quantification of GI_{50} values was not always possible. For example, compounds with no/negligible cytotoxic activity, even at the highest concentration tested (1000 μM), the resulting sigmoidal curves were "incomplete" enabling only a crude estimation of GI_{50} values. Further discussion and comparable analysis of FEN inhibitor-induced cytotoxicity, with respects to GI_{50} values, will be carried out in Section 5.4, aiming to establish some meaningful relationships for the selection of compounds for downstream optimisation.

A.


Cell viability (%)
[Compound], μM
DFI_9854
DFI_6490
DFI_6387
DFI_1742
DFI_637
DFI_7053
DFI_4002
DFI_3147
DFI_5392
DFI_535
DFI_8064
DFI_2377
DFI_1391
DFI_2172
DFI_9589
DFI_144
DFI_7317
DFI_2013
DFI_1275
DFI_784
DFI_1626
DFI_2198
DFI_289
DFI_6255
DFI_207
DFI_6147
DFI_2715
DFI_1769
DFI_1742
DFI_6568
DFI_4759
DFI_6706
DFI_8691
DFI_6529
DFI_6537
DFI_6323
DFI_3827



B.

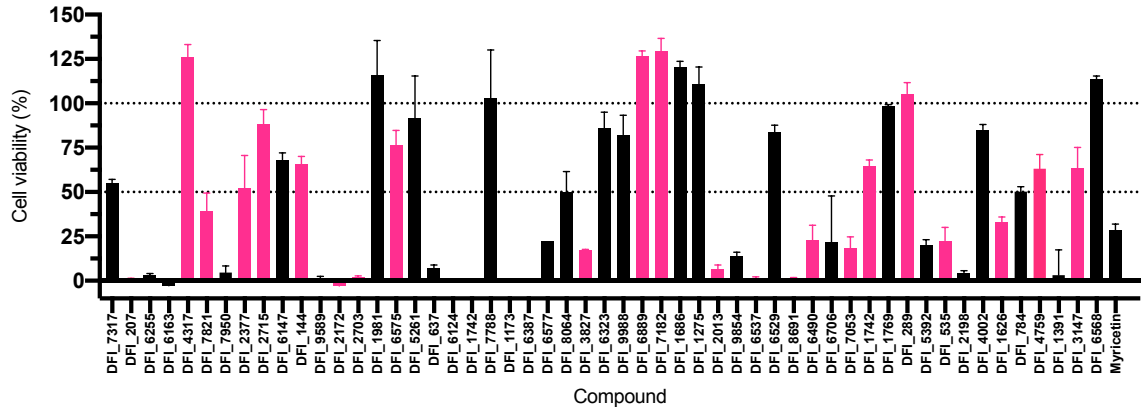


Figure 5.10 Effect of selected FEN inhibitors on the viability of HCT-116 post-24 hr treatment. (a) Dose-response curves of inhibitor-induced haemolysis at concentrations ranging from 1-1000 µM. (b) Calculated cell viability percentages at 333.3 µM drug concentration. Each value represents the normalised mean of three technical repeats (i.e. $N = 1, n = 4$), based on vehicle and 1% (v/v) Triton X-100 controls (i.e. $N = 1, n = 8$). Data analysis and GI_{50} estimates were carried out using nonlinear regression [Inhibitor] vs. response -Variable slope (four parameters) model in GraphPad Prism version 8.2. $Z' \geq 0.5$ for all assays performed.

5.3.2 Evaluation of FEN inhibitor-induced cytotoxicity in epithelial HEK-293 cell line

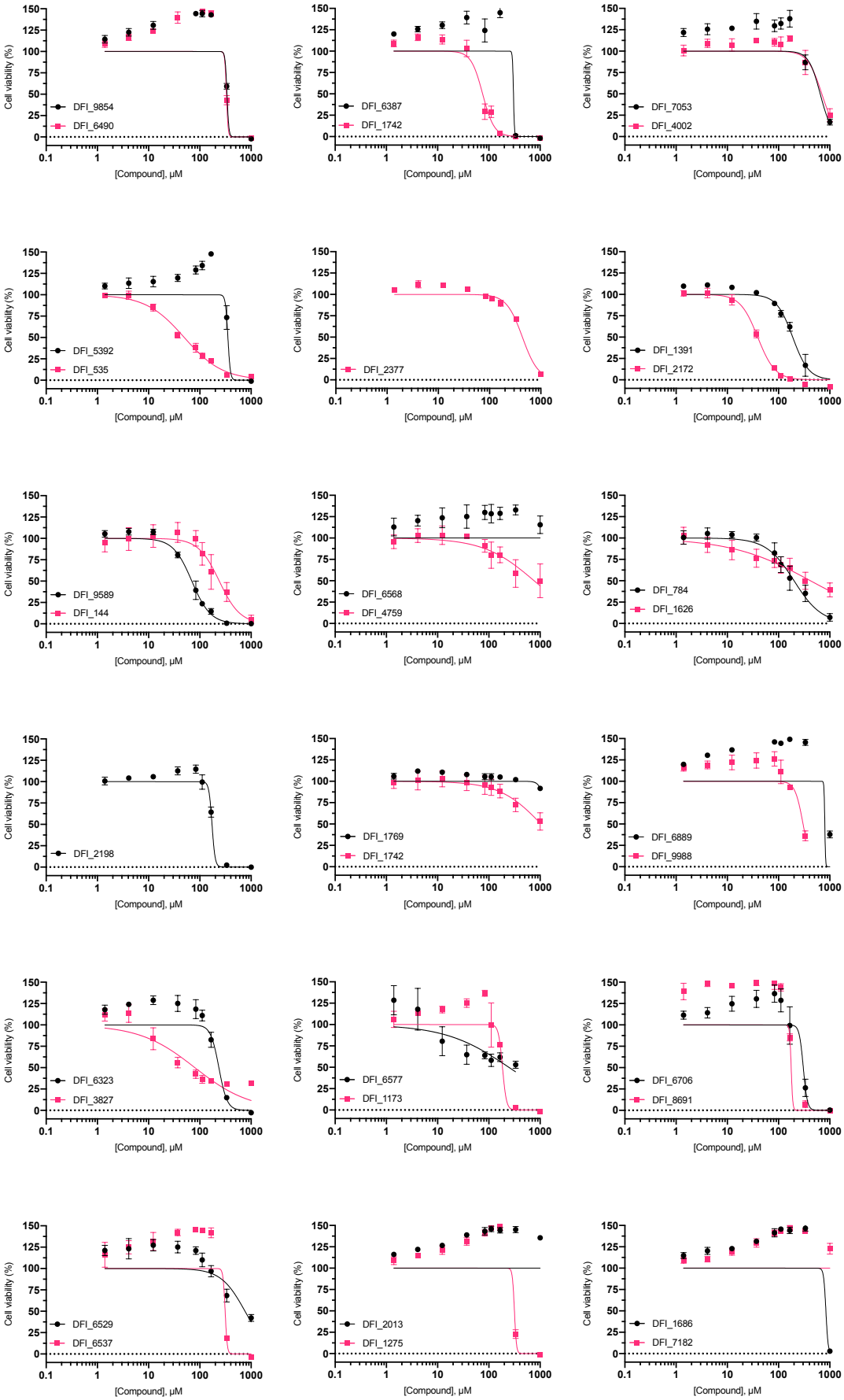
To further investigate the cytotoxic effect of short-listed FEN inhibitors, experiments were extended from a cancerous cell-line with established sensitivity to FEN inhibition, to a non-cancerous cell line, without any established sensitivity or resistance to FEN inhibition. Human embryonic kidney-derived HEK-293 cell line, commonly used as an example of "normal" cell line in cell viability assays for the evaluation of drug-induced toxicities, was selected for another and final round of cytotoxicity assays (Amelian *et al.*, 2017; Stepanenko and Dmitrenko, 2015). Following the pre-described methodology used with HCT-116 cell line, an updated list of 58 FEN inhibitors, including myricetin, undergone evaluation for their induced cytotoxic effect.

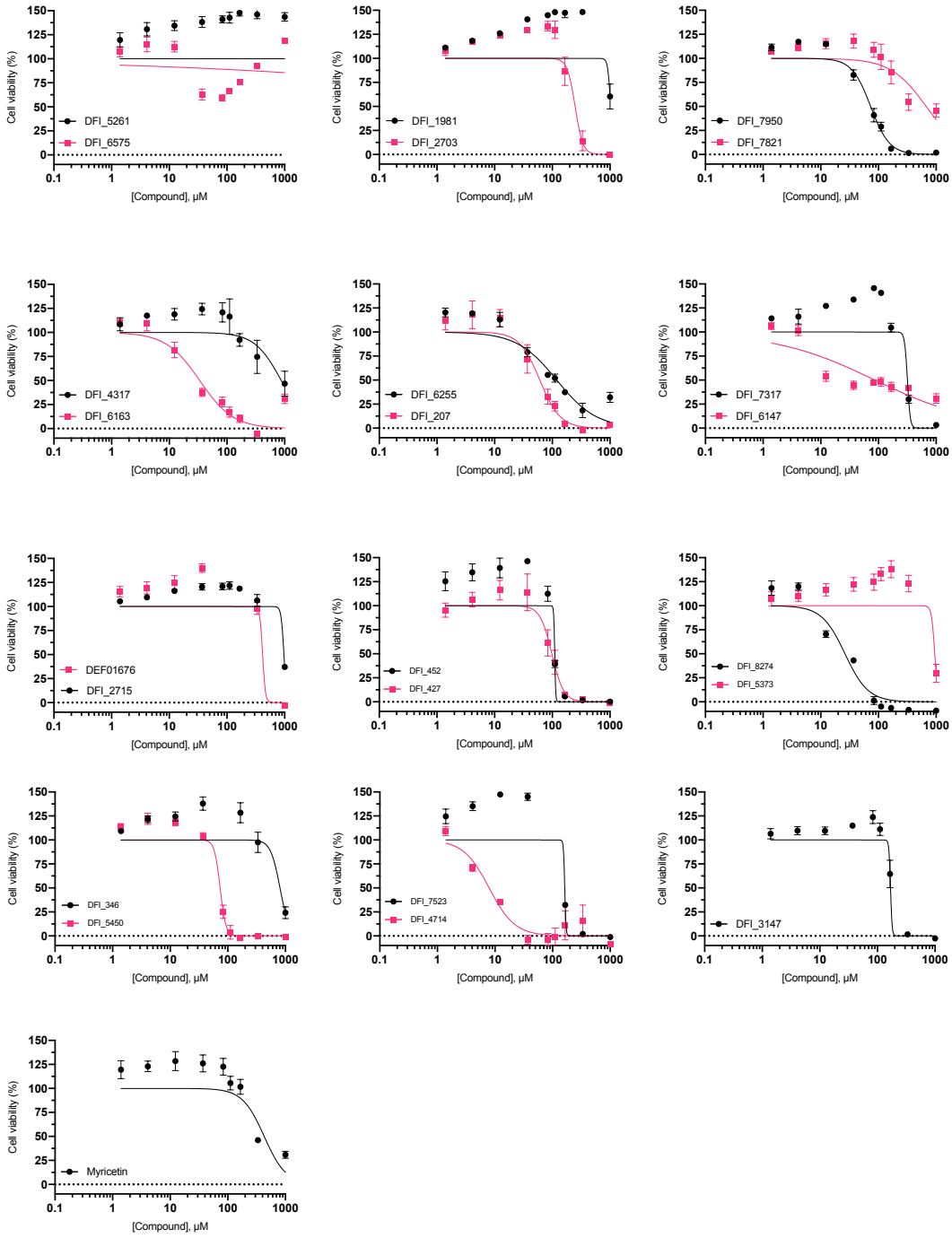
Once again, the resulting dose-response curves revealed varying cytotoxic activities among the tested FEN inhibitors. Whilst some compounds showed no detectable cytotoxic activity, some others were found to be strongly cytotoxic. Compound DEF01806 was identified as the most cytotoxic with a GI_{50} of 8 μM (Figure 5.11; Table 5.3). However, this was among the last set of compounds added on our short-listing and therefore was not evaluated in HCT-116 for direct comparison. DFI_06163 on the other hand, the most potent compound in HCT-116 with GI_{50} of 26.5 μM, showed a GI_{50} of 33.6 μM possibly due to the increased sensitivity of HCT-116 in FEN1 inhibition (Figure 5.11; Table 5.3).

Similar to cytotoxicity assays performed in HCT-116, the remaining cell viability at 333.3 μM was maintained ≥50% for 24 out of the 57 compounds tested (excluding myricetin). For the rest 33 compounds, cell viability was maintained <50%, indicating a higher extent of induced cytotoxicity (Figure 5.11). Out of these, 15 compounds were classified among the most cytotoxic resulting in ≤5% cell viability, consistent to the numbers observed using HCT-116 cells. Myricetin was also found to have an analogous cytotoxic profile to that observed in HCT-116 cells (Figure 5.11).

Whilst the calculated GI_{50} values will be further discussed next, overall the numbers of compounds with no/negligible to highly cytotoxic effects seem to be analogous between experiments carried out in HCT-116 and HEK-293 cells.

A.


Cell viability (%)
[Compound], µM
DFI_9854
DFI_6490
DFI_6387
DFI_1742
DFI_7053
DFI_4002
DFI_5392
DFI_535
DFI_2377
DFI_1391
DFI_2172
DFI_9589
DFI_144
DFI_6568
DFI_4759
DFI_784
DFI_1626
DFI_2198
DFI_1769
DFI_1742
DFI_6889
DFI_9988
DFI_6323
DFI_3827
DFI_6577
DFI_1173
DFI_6706
DFI_8691
DFI_6529
DFI_6537
DFI_2013
DFI_1275
DFI_1686
DFI_7182


Cell viability (%)
[Compound], μM
DFI_5261
DFI_6575
DFI_1981
DFI_2703
DFI_7950
DFI_7821
DFI_4317
DFI_6163
DFI_6255
DFI_207
DFI_7317
DFI_6147
DEF01676
DFI_2715
DFI_452
DFI_427
DFI_8274
DFI_5373
DFI_346
DFI_5450
DFI_7523
DFI_4714
DFI_3147
Myricetin

B.

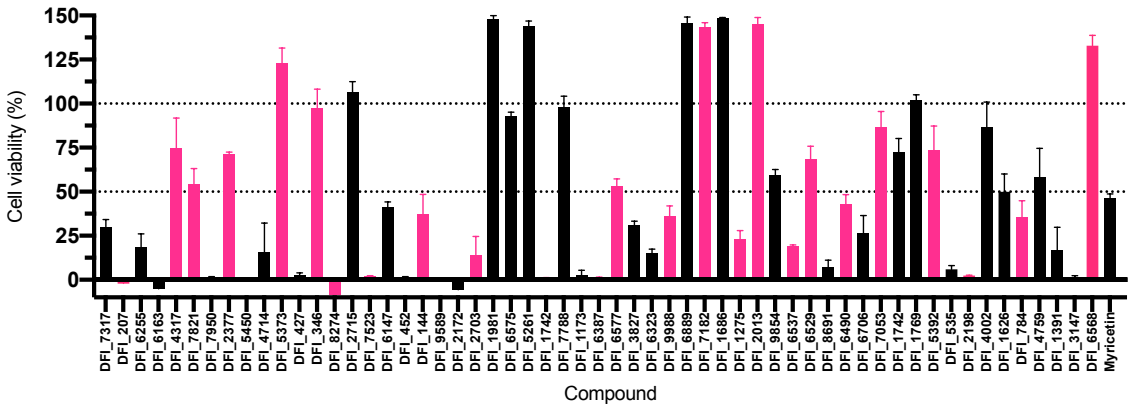

Cell viability (%)
Compound
DFI_7317
DFI_207
DFI_6255
DFI_6163
DFI_4317
DFI_7821
DFI_7950
DFI_2377
DFI_5450
DFI_4714
DFI_5373
DFI_427
DFI_346
DFI_8274
DFI_2715
DFI_7523
DFI_6147
DFI_452
DFI_144
DFI_9589
DFI_2172
DFI_2703
DFI_1981
DFI_6575
DFI_5261
DFI_1742
DFI_7788
DFI_1173
DFI_6387
DFI_6577
DFI_3827
DFI_6323
DFI_9988
DFI_6889
DFI_7182
DFI_1686
DFI_1275
DFI_2013
DFI_9854
DFI_6537
DFI_6529
DFI_8691
DFI_6490
DFI_6706
DFI_7053
DFI_1742
DFI_1769
DFI_5392
DFI_535
DFI_2198
DFI_4002
DFI_1626
DFI_784
DFI_4759
DFI_1391
DFI_3147
DFI_6568
Myricetin

Figure 5.11 Effect of selected FEN inhibitors on the viability of HEK-293 post-24 hr treatment. (a) Dose-response curves of inhibitor-induced haemolysis at concentrations ranging from 1-1000 μM. (b) Calculated cell viability percentages at 333.3 μM drug concentration. Each value represents the normalised mean of three technical repeats (i.e. $N = 1, n = 4$), based on vehicle and 1% (v/v) Triton X-100 controls (i.e. $N = 1, n = 8$). Data analysis and GI_{50} estimates were carried out using nonlinear regression [Inhibitor] vs. response - Variable slope (four parameters) model in GraphPad Prism version 8.2. $Z' \geq 0.5$ for all assays performed.

5.4 Comparative analysis of FEN inhibitor-induced cytotoxicity

Aiming to probe whether the selected FEN inhibitors are suitable, safety wise, for further development, the collected datasets on horse blood erythrocytes, HCT-116 and HEK-293 cells were cross-correlated. Overall, the toxicity profiles of a total of 62 FEN inhibitors, including myricetin, were evaluated and compared for the identification of any meaningful correlations regarding their potential *in vivo* haemolytic and adverse side effects. Maximum haemolytic effects and GI_{50} values over HCT-116 and HEK-293 are all summarised in Table 5.3.

As previously shown, the haemolytic activities displayed by the evaluated FEN inhibitors varied, with significant profiling differences between experiments performed using a 2% (v/v) buffered RBCs suspension and whole blood erythrocytes. Given that haemolysis assays were carried out using an only 4-point concentration range, accurate determination of the respective EC_{50} values was not possible. Comparisons are therefore, restricted to the maximum haemolytic effect observed at the highest FEN inhibitor concentration tested. Although the majority of the tested FEN inhibitors appeared to have no/negligible haemolytic activity (≤5% haemolysis) under both conditions, those with increased haemolytic activity over 2% (v/v) buffered RBCs suspension were found to be totally inactive in whole blood experiments (Tables 5.3-5.4). Given that FEN inhibitors are highly anionic, possible interactions with major plasma transport proteins when exposed to protein-rich environments such as whole blood, are likely to generate misleading underestimated haemolytic effects.

On the other hand, the cytotoxic effects of the FEN1 inhibitors over human HCT-116 and HEK-293 cells lines, the latter of which is commonly used as a reference cell line for the estimation of drug-induced cytotoxicity, seemed to provide greater comparability. Overall, the highly cytotoxic compounds were found to be conserved between the different cell lines used, indicating a general agreement between the two datasets (Tables 5.3-5.4; Figure 5.12). However, HEK-293 cells were found more tolerant to the effect of FEN1 inhibitors, compared to the established as sensitive to FEN inhibition HCT-116 cell line based on the respective GI_{50} values (Ward *et al*., 2017; Table 5.3). Due to the lack of complete sigmoidal dose-response curves for some of the compounds with limited cytotoxic effects, even at 1000 μM, GI_{50} values only represent crude estimates for those cases.

Cross-comparison between the collected haemolysis and cytotoxicity datasets, did not reveal any obvious strong correlations. Whilst the dataset on haemolytic effect of the selected FEN inhibitors was excluded from comparison due to the possibility of being severely misleading, the haemolytic and cytotoxic activities of the individual molecules seemed to vary. Cell viability assays were found significantly more sensitive compared to haemolysis, revealing an increased number of active FEN inhibitors with potent cytotoxic effects (≤50% reduction in cell viability; GI_{50} ≤500 μM) (Tables 5.3-5.4). The great majority of these cytotoxic FEN inhibitors in HEK-293 and HCT-116 cells, had no/negligible haemolytic activity, even when using buffered RBCs suspension. Haemolytic compounds on the other hand, did reveal significant cytotoxic activity against HEK-293 and HCT-116 cells, with compounds DFI_1686-DFI_6889 being the only exemptions (Tables 5.3-5.4). Whether this is specifically related to human membrane, their respective physicochemical properties and/or stereochemistry is unclear.

Although history have shown that both cell viability (e.g. Resazurin) and haemolysis assays are powerful tools for estimating drug-induced toxicity, at least at the early drug discovery stages, the data presented here should be taken with caution. Drug-induced toxicity is often organ- and organism-specific with complex multi-factorial mechanisms. Therefore, the *in vitro* evaluation models used here, do not necessarily mimic the *in vivo* physiologic and pathologic environments expected to potentially be encountered at later clinical stages. These, in addition to the differing drug concentration ranges and slight variations on the experimental conditions, which, especially on high-throughput fashion experiments, are likely to lead to misleading cytotoxic effects. With the inaccurate prediction of drug-induced toxicity counting as one of the leading causes for the high drug attrition and withdrawn rates, conclusions with regards to "safety" of the evaluated FEN inhibitors are inappropriate at this stage (Astashkina *et al.*, 2012). Animal model studies evaluating complex systemic drug-induced toxicity and the establishment of strong correlations between the *in vivo* and *in vitro* cell-based models, are required to sufficiently back up therapeutic index and "safety" conclusions.

Table 5.3 Summary of FEN inhibitor-induced cytotoxicity in horse blood erythrocytes, HEK-293 and HCT-116 cell lines[1,2]

Compound DFI no.	NgFEN IC_{50} (µM)	HsFEN IC_{50} (µM)	*Neisseria gonorrhoea* MIC (µM)	Max (%) haemolysis at 1 mM 2% (v/v) RBC suspension	Max (%) haemolysis at 1 mM Whole blood erythrocytes	GI_{50}, (µM) HEK-293	GI_{50}, (µM) HCT-116
DFI_7317	32	>50	4	4.3±1.2	≤1	317.7	330.3
DFI_207	79	>100	2	≤1	≤1	60.8	35.9
DFI_6255	>100	>100	>256	3.9±1.6	≤1	119.4	59.7
DFI_6163	>100	>100	>256	3.9±2.6	≤1	33.6	26.5
DFI_4317	34	>100	1	4.4±0.3	≤1	872.3	>1000
DFI_7821	>100	>100	6	≤1	≤1	672.7	317.4
DFI_7950	14	>100	3	≤1	≤1	73.3	75.1
DFI_2377	3	>1000	128	≤1	≤1	436.9	366
DFI_5450	3.2	18	n/a	n/a	n/a	73.7	n/a
DFI_4714	2.9	442	n/a	n/a	n/a	7.9	n/a
DFI_5373	22	365	128	n/a	n/a	945.8	n/a
DFI_427	4.9	31	8	n/a	n/a	98.2	n/a
DFI_346	1.4	23	8	n/a	n/a	805.9	n/a
DFI_8274	3.2	35	128	n/a	n/a	26.3	n/a
DFI_2715	63	>1000	>256	n/a	n/a	971.2	628.2
DFI_7523	4.9	23	32	n/a	n/a	162.5	n/a
DFI_6147	4.4	68	>256	59.8±0.9	≤1	87.2	>1000
DFI_452	2	26	32	n/a	n/a	110.1	n/a
DFI_144	12	140	4	2.6±1.2	≤1	239.4	773.4
DFI_9589	44	>1000	1.5	≤1	≤1	69.7	65.2
DFI_2172	45	542	>256	9.2±0.6	≤1	39.1	36.6
DFI_2703	5	134	3	100±2.8	≤1	248.5	150.6
DFI_1981	115	>1000	192	≤1	≤1	>1000	931.4
DFI_6575	7.4	0.02	128	≤1	≤1	>1000	>1000
DFI_5261	7.9	156	24	13.3±0.6	≤1	>1000	386.9
DFI_637	13	58	16	74.9±2.6	≤1	n/a	171.3
DFI_6124	9.9	39	32	104.5±2.5	4.3±2.3	n/a	76.1
DFI_1742	10	53	0.4	105.8±3.1	≤1	75.5	66.1
DFI_7788	24	159	48	54.8±0.4	4.4±3.7	414.8	373.1
DFI_1173	13	56	4	104.3±1.1	2±1.9	184.3	166.9
DFI_6387	7.9	49	6	54.4±17.2	≤1	302.2	136.8
DFI_6577	7.1	0.07	6	≤1	≤1	244.8	45.6
DFI_8064	0.25	0.16	8	≤1	≤1	n/a	350.8
DFI_3827	5.7	0.54	>256	≤1	≤1	79.9	106.3
DFI_6323	10	107	16	≤1	5.6±0.9	234.9	367
DFI_9988	12	113	8	3.7±0.5	2.7±1.8	300.4	362.7
DFI_6889	52	278	128	74.9±2.6	2.6±0.4	798.3	>1000
DFI_7182	275	>1000	128	105.8±3.2	≤1	>1000	>1000
DFI_1686	>1000	>1000	>256	54.4±17.1	≤1	838.4	>1000
DFI_1275	60	186	8	106.6±2.1	1.6±1.2	315.4	>1000
DFI_2013	890	950	>256	21.2±8.7	2.4±2.3	>1000	136

DFI_9854	2.2	96	3	20.7±9.3	≤1	338.6	169.2
DFI_6537	7.4	59	6	99.6±5.3	≤1	311.3	173.6
DFI_6529	36	3.5	>256	3.2±0.5	≤1	759.9	>1000
DFI_8691	8	33	5	109.5±3.1	6.6±0.3	176.2	161.8
DFI_6490	4.6	64	8	106.6±2.1	≤1	329.9	323.5
DFI_6706	17	71	16	74.5±2.7	1.7±0.6	303.7	249.3
DFI_7053	18	387	8	3.7±0.5	≤1	648.8	250.2
DFI_1742	>1000	28	>256	≤1	≤1	>1000	>1000
DFI_1769	>1000	>1000	>256	≤1	≤1	>1000	>1000
DFI_289	>1000	>1000	>256	≤1	≤1	n/a	>1000
DFI_5392	19	132	8	1±0.2	≤1	352.4	319.6
DFI_535	2.2	0.03	6	≤1	≤1	49.6	160.9
DFI_2198	17	68	4	22.2±1.4	2.2±1.2	172.6	162.4
DFI_4002	15	379	32	≤1	≤1	698.1	458.1
DFI_1626	20	0.06	>256	≤1	≤1	418.7	100.8
DFI_784	35	0.04	128	≤1	≤1	200.3	303
DFI_4759	237	0.5	>256	≤1	≤1	763.6	594.7
DFI_1391	75	>1000	64	≤1	≤1	188.5	175.4
DFI_3147	14	124	0.5	26.4±16.7	≤1	168.9	326.6
DFI_6568	>1000	>1000	>256	≤1	≤1	>1000	>1000
Myricetin	n/a	n/a	n/a	n/a	n/a	428.8	395.2

[1]Potency indicator (IC_{50}), corresponding to the half-maximal responses of the short-listed FEN1 inhibitors, and minimum inhibitory concentration (MIC) values were acquired by Dr Sarbendra L. Pradhananga and DeFENition Ltd.

[2]"n/a" refers to non-evaluated FEN inhibitors under the stated conditions. Due to the ongoing systematic screening, these were either not included in the original short-list or there were insufficient quantities at the time.

Table 5.4 Numbers of highly to no/moderately cytotoxic FEN inhibitors

Experiment	**Total tested**	**Number of DFIx compounds[1,2]**		
		<25% Haemolysis/ GI_{50} ≥750 µM	**<50% Haemolysis/ GI_{50} ≥500 µM**	**≥50% Haemolysis/ GI_{50} <500 µM**
2% (v/v) RBCs suspension	52	35 (67.3%)	36 (69.2%)	16 (30.8%)
Whole blood	52	52 (100%)	52 (100%)	0 (0%)
HCT-116	53	14 (26.4%)	16 (30.2%)	37 (69.8%)
HEK-293	57	16 (28.2%)	19 (33.3%)	38 (66.7%)

[1]Number of DFIx compounds classified based on their induced percentage haemolysis for haemolysis experiments, and GI_{50} values for cell viability experiments. This table only refers to numbers of DFIx compounds and excludes myricetin.

[2]Parenthesis reflect the percentage of the number of DFIx compounds classified in a particular group, over the total number of compounds tested on that particular experiment.

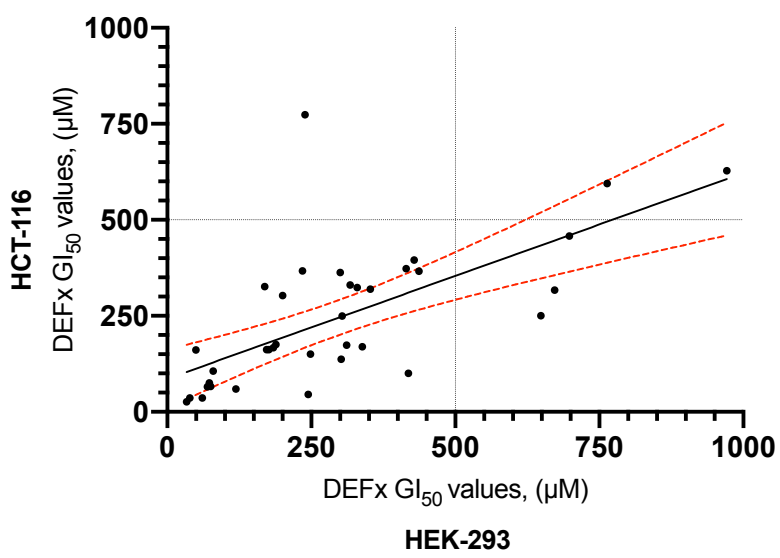


Figure 5.12 Correlation between the calculated DFIx inhibitor cytotoxic potencies in HEK-293 and HCT-116 cell lines. GI_{50} values for each of the evaluated DFIx inhibitor in both HEK-293 and HCT-116 were plotted in GraphPad Prism 8.2. Data were analysed by Bravais-Pearson correlation analysis with R^2 and r values of 0.45 and 0.67, respectively. Inhibitors with GI_{50} values >1000 μM were excluded from analysis. Red dotted lines are representative of the 95% confidence bands of the best-fitted line.

5.5 Conclusion

Collectively, the haemolytic and cytotoxic profiles of a series of potent structurally related FEN inhibitors were determined with respect to the tested microenvironments. Although the majority of the evaluated FEN inhibitors with increased haemolytic activity was found to have an analogous increased cytotoxic activity, the results presented here suggest that this is not necessarily the rule. Cell viability assays revealed a significantly increased number of cytotoxic compounds compared to the number of haemolytic molecules. Due to restrictions in presentation and discussion of DFIx inhibitor structures, structure-toxicity relationships are not provided within the context of the present thesis. Despite the limitations and the requirement for further *in vivo* studies, the current toxicity data are the first ever collected on a series of in-house potent FEN inhibitors aiming to facilitate downstream drug design and development processes.

Chapter 6 – Discussion and Conclusion

6.1 Discussion

With >421 published articles between 2005-2019, FEN1-targeted research has unravelled and repeatedly supported the therapeutic potential of FEN1 inhibition for the development of novel anticancer and, albeit to a much lesser extent, antimicrobial drugs (Wei *et al*., 2021). Based on the available by 2019 literature, the systematic bibliometric analysis performed by Wei *et al*., (2021) indicated a tremendously growing research field around FEN1 enzymes. Studies mainly focusing on the biological roles of FEN1 in maintaining genomic integrity, its catalytic mechanism and links to cancer development and progression have prompted a novel research direction on FEN1 inhibitor development for cancer and microbial infections treatment. Whilst the numbers of these studies continue to increase, literature on FEN1 inhibitors remains limited to the *N*-hydroxyurea class of molecules (Exell *et al*., 2016; Tumey *et al*., 2005). Although *N*-hydroxyurea inhibitors represent the most well-studied class of HsFEN1 inhibitors to date, even 16-years after their initial introduction, none of the proposed FEN1 inhibitors has been effectively translated into the next stage of clinical evaluation.

Ever since the approval and successful application of PARP inhibitor Olaparib (Lynparza; AstraZeneca, UK) for *BRCA1/2*-defective cancer treatment, targeted inactivation of DNA repair genes with compensating functions for tumour-specific defective genes prompted a new era for personalised cancer therapeutics (Topatana *et al*., 2020; Yap *et al*., 2019). Despite the controversial "bewildering hodgepodge of genetic oddities" unravelled with the completion of The Cancer Genome Atlas (TCGA) in 2014, tumour-specific genetic alterations in DNA repair and replication genes indicated a common stem-hole for delivering targeted cancer therapies (Akbani *et al*., 2015; Ledford, 2015). Gene-specific modulation of DNA repair pathways for cancer treatment hence provided a unique potential in covering the heterogeneous nature of malignant tumors (Helleday *et al*., 2008).

In vitro studies validating the chemotherapeutic potential of HsFEN1, either as a monotherapy or combination therapy, are constantly increasing. These, in addition to its expanding roles in DNA repair pathways and its potential synthetic lethal interaction pair genes through which FEN1-modulated treatments can be delivered (Helleday *et al*., 2008). However, given the limited literature in FEN1 inhibitors, and without any related to FEN1 inhibition clinical trials registered to date, low-molecular weight inhibitors of HsFEN1 are now in demand to enable further drug design and development.

6.1.1 Development of an *in silico* screening strategy for the identification of novel HsFEN1 inhibitors

Over the past few decades, computational molecular docking methods have played an important role in the successful discovery and later development of novel medicinal products (Meng *et al*., 2011). Such drug discovery successes were underlined by tremendous

advancements in the complexity and efficiency of the available docking tools, computing power and their implemented strategies (Yuriev *et al*., 2015). Typically, the efficiency and performance of an *in silico* molecular docking strategy relies on the interplay between the accuracy of the predicted favourable binding site, ligand pose and affinity, or the so-called search algorithm and scoring function. Although often the linear correlation between the extent of experimental inhibition and *in silico* binding energy is poor, discrimination between strong and no/weak binders is sufficient to maintain time and cost-effectiveness (Englebienne and Moitessier, 2009).

Metalloenzymes with metal ion-dependent catalytic capabilities, such as human and microbial flap endonucleases, do possess an additional challenge for molecular docking. The range of possible metal coordination geometries and hence, the provided inhibitor binding flexibility, require alternative parametrisation to ensure accurate prediction of biologically relevant scoring functions (Irwin *et al*., 2005). Practical approaches to improve binding energy prediction and establish models with increased overall performance were employed since the early 2000s, focusing on optimisation of the active site metal charges and radius in metalloproteinases (Hu *et al*., 2004; Hu and Shelver, 2003). Despite the covalent-like binding interaction between metals and ligands, standard non-covalent docking against metalloenzymes with sterically controlled metal chelation environments was shown to effectively predict strong binders under tweaked metal parameters (Irwin *et al*., 2005).

Given the range of FEN enzymes targeted in Sayers' laboratory for the development of novel antimicrobial drugs, in addition to the anticancer potential of HsFEN1 inhibitors, a constant requirement for robust discrimination between the strongest and not so strong binders at an early *in silico* analysis stage exists. Utilizing the widely used in academic drug discovery open-source docking programs, AutoDock and Vina, varying docking runs of MaybridgeRO3 fragment library against HsFEN1 revealed predictive power dependence on the respective active site metal-charges. Focusing on HsFEN1 active site as the pre-defined ligand search space, which was later shown to be the primary predicted binding site, docking runs of +2 active site metal charges were found to effectively predict 78% of the total number of *in vitro* confirmed hits. This is higher than the average prediction accuracy of ten popular commercial and academic docking programs, which was previously estimated between 47-68% for both top scored and best poses (Wang *et al*., 2016).

Whilst Irwin *et al*., (2005) made clear that this simplified approach of non-covalent docking is not suitable for every metalloenzyme, the results presented here indicate that HsFEN1 is one of those well-performing metal targets. Based on these and the indication that FEN1 homologs, including the targeted for antimicrobial drug development FEN enzymes, are also likely to be well performing at +2 active site metal charge, analogous experiments carried out in parallel projects of Sayers' laboratory have validated these assumptions (data not shown). This is now a widely used *in silico* drug screening approach for the early identification of potentially strong binders against a range of FEN1 enzymes in Sayers' laboratory. However, as suggested by Irwin *et al*., (2005), effective application of the current approach is limited to constrained active

sites and not really "solving" the long-standing poor correlation issues between the predicted binding energies and experimental inhibitions. Alternative docking programs with atom-type-dependent functions (e.g. FITTED 3.1), specifically designed to improve metalloenzyme inhibitor prediction, possess the potential of further prediction power improvement, beyond our 78% (Pottel *et al*., 2014). However, unlike AutoDock suite, these are typically proprietary programs.

Traditionally, utilization of an average docking score/rank for a small-molecule compound, calculated by the individual scores/ranks from multiple differing docking tools, is used to provide an equally simple, outperforming and, perhaps occasionally, a more robust alternative (Ericksen *et al*., 2017; Osguthorpe *et al*., 2012; Tian *et al*., 2014). Despite the multiple consensus scoring/ranking factors proposed to date, focusing at the intersection of the best scoring compounds among multiple docking tools, provides an increased sensitivity to algorithm-dependencies and parameterization of the distinctly different docking tools (Kukol, 2011; Palacio-Rodriguez *et al*., 2019). Low performance of an individual docking program used for averaging the emerging scores/ranks, compromises the overall performance of the followed consensus strategy (Palacio-Rodriguez *et al*., 2019). Based on these, but yet still within the concept of consensus scoring/ranking, Palacio-Rodriguez *et al*., (2019) have proposed a consensus-ranking strategy which has the potential to further increase prediction power. Relying on the sum exponential distribution of individual compound rankings, derived from multiple docking runs using differing structures of the same proteins, this approach aims to eliminate not only docking tool-dependencies, but also any potential structure-dependencies (Palacio-Rodriguez *et al*., 2019). Both enzyme and substrate classes transit into intermediate states, which are often hard to capture and even harder to be accounted for during *in silico* docking. Despite the modelling improvements in protein/ligand flexibility, incorporating the accommodated conformational changes of the adopted intermediate states remains a great challenge (Cavasotto and Singh, 2008; Spyrakis and Cavasotto, 2015; Wang and Zhu, 2016). The currently available HsFEN1 structures, including those solved within the framework of this project, revealed threading intermediates with distinctly differing orientations of key arch and active site residues (Tsutakawa *et al*., 2011; Tsutakawa *et al*., 2017). The characteristic structural plasticity of the helical gateway, its transition from an ordered-to-disordered conformation, in addition to the metal ion-dependent activity of HsFEN1, possess huge challenges for the *in silico* reflection of the "real" environment. A reflection, whose accuracy, strongly determines the effectiveness of the strategy used for early identification of potential HsFEN1 inhibitors. Given the numbers of available HsFEN1 structures and the current resources in Sayers' laboratory, the potential of consensus ranking combining datasets from docking runs performed against a range of apo to complex intermediate HsFEN1 structures is worthy further exploration.

To facilitate the identification HsFEN1 inhibitors and also further validate the proposed for HsFEN1 docking strategy of +2 active-site metal charges, a second library (BIONET) was screened. Experimental evaluation of the *in silico* predicted BIONET hits showed a conserved to MaybridgeRO3 "true" inhibitors:false positives ratio. Out of the total number of *in silico*

predicted MaybridgeRO3 and BIONET hits, 86.8% and 87.6%, respectively, were found to significantly inhibit the catalytic activity of HsFEN1 in FRET-based assays. This indicates that docking against HsFEN1 with +2 active-site metal charges can effectively predict the great majority of molecules with significant inhibitory activity *in vitro* with a potentially analogous performance. However, whether the relation of the *in silico* predicted, but also *in vitro* confirmed, BIONET hits to the overall number of "true" inhibitors in the library correlates the 78% seen in MaybridgeRO3, remains unclear due to the absence of a complete *in vitro* validation of all molecules composing BIONET library.

Out of the 1,000 MaybridgeRO3 fragments examined, 127 were identified to effectively inhibit ≥60% of HsFEN1 activity. Perhaps more interestingly, 59 of these, constituting about 46.5% of the total number of hits and 5.9% of the entire library screened, were found to inhibit ≥90% of HsFEN1 activity suggesting an increased potential. Overall, 80.3% of the total number of hits contained at least one carboxyl group in their structures, indicating a potentially metal coordination-based inhibition mechanism. Given that the structures of every HsFEN1 inhibitor previously proposed are also possible metal chelators (Section 1.7.3), a preliminary similarity analysis of the identified top hits was carried out revealing 13 main clusters with differing structural cores from those published. These differences mainly relied on the existence of carboxyl groups directly attached to a single or combination of aromatic rings (e.g. benzene, pyrrole, pyridine, pyrimidine, furan, thiophene), or through an aliphatic hydrocarbon chain. Based on these and a collaborative pilot HE study, compound HE_20 with an IC_{50} of ~1.7 μM was identified as an interesting alternative to *N*-hydroxyurea inhibitors expanding the current range of reported HsFEN1 inhibitors to date. Compound HE_20 is one of the most potent HsFEN1 inhibitors identified to date through early stage computational/virtual screening approaches (Deshmukh *et al*., 2017; Panda *et al*., 2009; Table 1.7). Although HE_20 is a smaller than the co-crystallised with HsFEN1 *N*-hydroxyurea molecule, with an also lower number of H-acceptors/donors, virtual modelling indicated important interactions for consideration in future inhibitor development studies (Exell et al., 2016).

Future work will be focusing on the investigation of HE_20 mode of action and specificity, whilst screening for the identification of additional small-molecule inhibitors against HsFEN1 is ongoing. Isothermal titration calorimetry and/or the previously optimized in Sayers' laboratory thermal shift assay will be used for measuring inhibitor binding kinetics, aiming to determine the association and dissociation rates of HE_20, its binding thermodynamics and how these may relate to its chemical structure and covalent nature. Given that it remains unclear whether HsFEN1 inhibition by HE_20 is a result of direct interactions between HsFEN1 and HE_20, inhibitor kinetic experiments will be also evaluating the potential of DNA integration and its effect in HsFEN1 inhibition. Co-crystallization experiments of HE_20 and HsFEN1 will be carried out to unravel the "true" binding mode of HE_20, evaluating the predicted binding poses and modelled intermolecular interactions between HE_20 and HsFEN1 obtained through virtual screening. These, in addition to experiments investigating cellular activity and target engagement of HE_20 in live cells, can enable the establishment of structure-activity

relationships and a more targeted structure-based design approach to be followed for small-molecule inhibitor optimization towards a medicinal candidate.

6.1.2 Crystallisation of HsFEN1:DNA complexes

Based on the existing controversies surrounding suggested threading- and tracking-mediated HsFEN1 catalysis mechanisms at the commencement of the current project, crystallisation trials of HsFEN1:DNA complexes were carried out to improve understanding and facilitate future development of potent mechanism-based inhibitors. Two different approaches were followed for crystallisation involving the use of either a catalytically inactive truncated D179K HsFEN1 mutant, believed to retain DNA binding, or the native full-length WT HsFEN1 with calcium ions, which were thought to inhibit HsFEN1 (Harrington and Lieber, 1994). Among these, crystallisation trials and subsequent optimisations resulted in multiple crystals of the native full-length WT HsFEN1 complexed with the exact same cleaved product of the range of DNA substrates examined. This range of DNA substrates (JT oligonucleotides) were composed from the same duplex core, capturing threading of their various length 5′-flaps in the presence of a Ca^{2+}-inhibited HsFEN1 had the potential to unravel the guiding process and conformational changes prior to threading and HsFEN1-mediated catalysis. Despite the presence of $CaCl_2$ at about 2.5 mM concentration in the crystallisation drop, all co-crystallised DNA substrates were found to have been processed at the phosphodiester bond 1-nt into the ss-ds DNA junction, as it would have been expected by the catalytically active HsFEN1 enzyme.

Based on biochemical studies on the restriction endonuclease *Eco*RV, T5FEN and *Trypanosoma brucei* (Tb) FEN, Ca^{2+} ions are considered as catalytically inert while enhancing substrate binding capabilities (Dr Sarah L. Oates, unpublished; Feng *et al*., 2004; Vipond and Halford, 1995). Early experiments characterising the enzymatic activity of HsFEN1 with a range of divalent metal ions showed about 83% inhibition of HsFEN1 activity in the presence of 1 mM $CaCl_2$ concentration (Harrington and Lieber, 1994). Given that this is 2.5-times less than the overall concentration of $CaCl_2$ presented in our crystallisation drops, the resulting product DNA substrates co-crystallised with WT HsFEN1 indicated that some enzymatic activity took place, despite the increased $CaCl_2$ concentration used. Whether the observed substrate processing by WT HsFEN1 was indeed supported by the presence of $CaCl_2$ at this concentration rather than traces of divalent metal ions (e.g. Mg^{2+}) from protein purification or contaminants in other buffer components, is unclear. However, contamination from protein purification procedures seems less likely, given that gel filtration was used as the final purification step, and storage of highly concentrated protein aliquots were carried out in buffer conditions lacking any source of divalent metal salts. Future assays examining the effect of increasing Ca^{2+} concentrations on HsFEN1 cleavage and substrate binding capabilities are required. Previous studies on bacteriophage T5 exonuclease showed that calcium ions were completely inhibitory, at least over periods of hours (Garforth *et al*., 1999). However, it is possible that Ca^{2+} supports a very slow rate of catalysis compared with Mg^{2+} so that hydrolysis

of the substrate only became apparent over the timescales (days) under which crystals were grown.

In addition, the overall crystallisation of WT HsFEN1:DNA complexes was likely facilitated by the enzymatic activity occurred and the presence of Ca^{2+} metal ions (Feng *et al*., 2004; Harrington and Lieber, 1994). These, given previous reports for an enhanced by Ca^{2+} substrate binding and the existence of two different crystal morphologies during screening, from which those suitable for data collection and subsequent optimisation grew slower. The discussed structure of WT HsFEN1:Ca^{2+}:processed JT2+5 complex solved at 2.3 Å, although requires further refinement and validation, is the first to our knowledge HsFEN1: product DNA structure with Ca^{2+} metal ions in the catalytic site of HsFEN1. However, Ca^{2+} was included in the crystallisation conditions of truncated D181A HsFEN1 in previous reports, but was not observed in the final structure (Tsutakawa *et al*., 2011).

Despite the fact that multiple datasets were collected from co-crystallisation of WT HsFEN1 with DNA substrates of varying 5′-flap lengths, molecular replacement used for structure solution indicated that identical HsFFEN1: product DNA structures were obtained. Among these, the structure of WT HsFEN1:processed JT2+5 complex, which had the highest resolution, was selected for further refinement and discussion. Comparison of WT HsFEN1:processed JT2+5 complex with those HsFEN1:DNA complexes deposited in the Protein Data Bank revealed overall identical conformations and binding of product DNA. The observations of the dsDNA substrate's ability to bend 100°, "trapping" of the 1-nt long 3′-flap and positioning of 5′-flap towards the active site as a result of HsFEN1 binding at the DNA bend and position of the H2TH domain, supported the previously reported specificity requirements for substrate recognition and binding of HsFEN1 characteristic of the wider members of FEN1 superfamily. Finally, the reproducibility of the reported crystallisation conditions, given the number of collected datasets with different DNA substrates and the analogies between their crystallisation symmetry, may possess the potential for future co-crystallisation experiments with HsFEN1 inhibitors.

Given the reproducibility of the reported crystallization conditions and the resulting crystal qualities, typically diffracting between 2-2.5 Å, future work will be focusing on utilizing these conditions for crystallization of WT HsFEN1:inhibitor complexes in the presence and/or absence of DNA. In addition, the reported crystallization conditions can facilitate the use of fragment-based drug discovery as an approach not only for optimization of the current HsFEN1 inhibitors, but also for crystallographic fragment screening (e.g. XChem; Diamond Light Source, UK). Crystallographic fragment screening typically involves soaking of high-quality protein crystals in optimal for crystal growth conditions, which however include high concentrations of small-molecule compounds. Such approaches are typically used as an alternative to the conventional spectrophotometric and fluorescence high-throughput assays and aim at the identification of additional binders with low complexity and perhaps affinity. These can provide key information with regards protein-ligand intermolecular interactions and ligand binding "hot spots", enabling further optimizations and growing of these binders into

drug-like lead compounds with increased inhibitory potency over HsFEN1. Finally, the possibility of generating a synthetic construct encoding the truncated WT HsFEN1 for crystallization at resolutions <2 Å, must not be excluded. Such construct will lack the flexible C-terminal domain, which was repeatedly reported to compromise crystallization efficiency and crystal quality (Tsutakawa *et al*., 2011).

6.1.3 Pre-clinical toxicity studies on human and microbial FEN1 inhibitors

According to US FDA's guidance for "Nonclinical Safety Studies for the Conduct of Human Clinical Trials and Marketing Authorization for Pharmaceuticals" (2010), *in vitro* toxicity studies are required and typically employed as an indication of therapeutic index and safety. Based on these, and aiming to, not only evaluate "safety", but also facilitate downstream prioritisation and optimisation of FEN1 inhibitors with the highest possible success potential, a series of standardised "stepwise" haemolysis and cytotoxicity experiments was carried out.

Due to the ease of erythrocyte isolation, haemolysis assays are commonly employed for preliminary toxicity evaluation (Farag and Alagawany, 2018). Although for drugs intended for human administration, the use of mammalian erythrocytes other than human is generally acceptable and commonly seen, it may not necessarily be reflective of the "true" *in vivo* toxicity profile (Greco *et al*., 2020). Animal studies are required to recapitulate the *in vivo* environments' complexity and dynamics, which are not encountered in conventional haemolysis and cell viability assays. During early-phase drug development though, the high-throughput experimentation required to evaluate cytotoxicity of an increased primary "hits" number, is simply not achievable in *in vivo* studies (Judson *et al*., 2013). Employment of *in vitro* drug-induced toxicity assays are therefore a requirement and, depending on the current objectives, must be designed wisely to prevent establishment of misleading structure-toxicity relationships.

With these in mind, selection of horse erythrocytes for the evaluation of FEN1 inhibitor-induced haemolysis, was based on their membrane compositional similarities to human (Table 6.1). Distribution of phospholipid head classes and their respective content within the outer and inner erythrocyte leaflets is a critical contributor to membrane's net charge (Belokoneva *et al*., 2003; Pagano and Faggio, 2015). This affects not only cellular sensitivity, but also selectivity between the negatively charged bacterial and neutral mammalian membranes (Matsuzaki, 2009). Although the latter does not apply for FEN1 inhibitors intended for development of anticancer drugs, it is central for antimicrobials and, thus the majority of DeFENition-derived FEN1 inhibitors.

Table 6.1 Phospholipid composition and distribution across human and horse erythrocyte membrane[1]

Phospholipid group class	Total content (%) (Outer/Inner Leaflet) Human erythrocytes	Total content (%) (Outer/Inner Leaflet) Horse erythrocytes
CPs (Phosphatidylcholine & sphingomyelin)	55.8 (88.9/23.1)	56.3 (79.7/33.2)
APs (Phosphatidylserine, phosphatidylinositol & phosphatidic acid)	16.6 (0/32.9)	15.7 (0/31.2)
PE (Phosphatidylethanolamine)	27.6 (11.1/43.9)	28 (20.3/35.6)

[1]Content percentages of the individual phospholipid classes were found from (Virtanen *et al*., 1998).

Whole blood erythrocytes and erythrocytes washed with and resuspended in PBS were both used for evaluation of FEN1 inhibitor-induced haemolysis. PBS is an established non-haemolytic isotonic buffer used in standard haemolysis assays, which although suitable, can't recapitulate the rich in plasma proteins whole blood condition. The haemolytic profiles of FEN1 inhibitors were found to significantly differ between whole blood and suspended in PBS erythrocytes. With the majority of FEN1 inhibitors exhibiting no/negligible haemolytic activity under both conditions, those with different levels of detectable haemolytic activity appeared completely inactive in whole blood assays. Given that the supplementary experiments performed ruled out the possibilities of this being attributed to blood storage, background or disproportioned positive controls haemolysis, the interaction potential of FEN1 inhibitors with major plasma proteins needs to be further explored.

Unlike haemolysis assays, the employed cell viability assays seemed to provide greater comparability, with the individual toxicity profiles being mostly analogous among the different human cell lines tested. About 60% of the total number of FEN1 inhibitors tested using both the cancerous HCT-116 and normal HEK-293 cells, showed lower GI_{50} values in HCT-116. This is consistent to the previously reported increased sensitivity of HCT-116 cells to pharmacological FEN1 inhibition by Ward *et al*., (2017). Although the majority of DeFENition-derived FEN1 inhibitors have increased potency over microbial FEN1 enzymes (data not shown), the HCT-116 sensitivity is more likely attributed to partial inhibition of HsFEN1 due to structural and sequence similarities among the targeted for inhibition FEN1 sites.

Both HCT-116 and HEK-293 cell lines were cultured under the exact same conditions, providing reliable analogies between the induced cytotoxic effects of individual FEN1 inhibitors against the tested cells. However, given the presence of 10% (v/v) FBS in culture medium, the reported cytotoxic effects are likely to be underestimated for those FEN1

inhibitors able to associate with plasma proteins. In any case, as the toxicity data presented here are the first ever obtained on these particular in-house series of microbial and human FEN1 inhibitors, they provide valuable information with regards to highly cytotoxic compounds. In addition, structure-toxicity analyses, although must be performed with caution, can still indicate potential therapeutic windows based on the individual GI_{50} and their respective MIC concentrations.

Because the structures of DeFENition-derived FEN1 inhibitors remain confidential, discussion of potential structure-toxicity relationships falls outside the scope of the current thesis. Whilst further experiments accounting for multi-species variability and interaction with plasma proteins are required, questions on the potential correlation to the individual compound properties, human erythrocytes and *in vivo* models remain. Are the identified cytotoxic concentrations a red flag for development of any of the short-listed compounds? How might these might be relevant to patients and treatment? Has designing of alternate chemical analogs, with improved for drug development physicochemical characteristics, the potential to circumvent FEN1 inhibitor-induced cytotoxicity? The reliability of the established structure-toxicity relationships is crucial in providing these answers and therefore, shaping how downstream drug design and development is approached.

Toxicity-related cell death or apoptosis can be a result of multi-parametric mechanisms including protein- and cell-drug interactions, or even the formation of their respective drug metabolites (Astashkina *et al*., 2012). The metabolic products of cisplatin for example, a widely used chemotherapeutic agent, have been found to induce a more severe toxicity-related damage than cisplatin *per se* (Sooriyaarachchi *et al*., 2016). Similar, the antibacterial potency of microbial peptides showed a 10-fold reduction in *in vitro* assays performed under physiological serum albumin concentrations (Svenson *et al*., 2007).

Collectively, a range of FEN1 inhibitors short-listed due to their inhibitory potency over human and microbial FEN1 enzymes were evaluated for their induced haemolytic and cytotoxic effects. The datasets presented here are only reflective of the respective microenvironments tested, and the selection criteria of the employed toxicity evaluation methodologies rely on the current objectives of the early inhibitor development stage. This is the first and most comprehensive toxicity study on the identified in-house FEN1 inhibitors to date. Although it sets the requirement for further research, it provides valuable information with regards to highly cytotoxic compounds and those of increased therapeutic potential for downstream development.

Future work will be focusing on both the establishment of structure-activity relationships and characterization of the intermolecular interactions between FEN inhibitors and plasma proteins. According to the free drug hypothesis, excessive plasma protein binding (PPB) can dramatically decrease the concentration of free drug and its overall efficacy. Whilst drug candidates with strong plasma protein binding capabilities are often abandoned from downstream development, modern history has shown that occasionally PPB can be optimized

to effectively increase half-life in humans (Gardiner *et al*., 2019). However, a thorough pharmacokinetic characterization, including steady-state distribution, transport rate and clearance, in addition to PPB characterization, is required for the implementation of a rigid prioritization and optimization strategy. Experimental work will be focusing on the binding capabilities of FEN inhibitors to plasma proteins (e.g. albumin), aiming at the pharmacokinetic profiling for some of these FEN inhibitors. The extent of these interactions, their effect on inhibition potency and half-life, as well as their association and dissociation constants to albumin will be also addressed. Collectively, these will facilitate establishment of strong structure-activity relationships and optimization of small-molecule inhibitors towards medicinal candidates.

6.2 Final conclusion

Take together, manipulation of active site metal charges indicated HsFEN1 with active site metals of +2 charge as a well-performing target for non-covalent metal docking. Whilst such strategy was found to effectively predict the great majority of *in vitro* confirmed HsFEN1 inhibitors at an early *in silico* stage, it also proposed to have the potential to be applied against FEN homologs for the identification of inhibitors with antimicrobial potential in a cost-effective manner. As part of a collaborative HE study, a HsFEN1 inhibitor with IC_{50} of ~1.7 μM was also identified for future inhibitor development studies, expanding the current range of published FEN1 inhibitors. In addition, the likely facilitated by the presence of $CaCl_2$ crystallization of HsFEN1:product DNA complex solved at 2.3 Å supported the unified for the members of FEN1 superfamily specificity requirements for substrate recognition and binding. These involved the dsDNA substrate's ability to bend 100°, "trapping" of the 1-nt long 3′-flap and positioning of 5′-flap towards the active site as a result of HsFEN1 binding at the DNA bend. Despite this project's main focus on HsFEN1, the cytotoxic effect of a series of both human and microbial FEN inhibitors, identified by the UoS spin-out DeFENition Ltd, was investigated for the first time. Although restrictions in structure presentation and discussion set the establishment of structure-activity relationships outside the scope of the present thesis, the collected data provided valuable insights for the direction of future toxicity and inhibitor development studies.

References

1. Akbani, R., Ng, P.K.S., Werner, H.M.J., Shahmoradgoli, M., Zhang, F., Ju, Z., Liu, W., Yang, J.Y., Yoshihara, K., Li, J., *et al*. (2015). A pan-cancer proteomic perspective on The Cancer Genome Atlas. Nature Communications *6*, 4852. 10.1038/ncomms5852.

2. Algasaier, S.I., Exell, J.C., Bennet, I.A., Thompson, M.J., Gotham, V.J.B., Shaw, S.J., Craggs, T.D., Finger, L.D., and Grasby, J.A. (2016). DNA and Protein Requirements for Substrate Conformational Changes Necessary for Human Flap Endonuclease-1-catalyzed Reaction. Journal of Biological Chemistry *291*, 8258-8268. 10.1074/jbc.M115.698993.

3. AlMalki, F.A., Flemming, C.S., Zhang, J., Feng, M., Sedelnikova, S.E., Ceska, T., Rafferty, J.B., Sayers, J.R., and Artymiuk, P.J. (2016). Direct observation of DNA threading in flap endonuclease complexes. Nature Structural & Molecular Biology *23*, 640-646. 10.1038/nsmb.3241.

4. Amelian, A., Wasilewska, K., Megias, D., and Winnicka, K. (2017). Application of standard cell cultures and 3D in vitro tissue models as an effective tool in drug design and development. Pharmacological Reports *69*, 861-870. 10.1016/j.pharep.2017.03.014.

5. Anderson, A.C. (2003). The process of structure-based drug design. Chemistry & Biology *10*, 787-797. 10.1016/j.chembiol.2003.09.002.

6. 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., and Artymiuk, P.J. (2013). The structure of Escherichia coli ExoIX-implications for DNA binding and catalysis in flap endonucleases. Nucleic Acids Research *41*, 8357-8367. 10.1093/nar/gkt591.

7. Astashkina, A., Mann, B., and Grainger, D.W. (2012). A critical evaluation of in vitro cell culture models for high-throughput drug screening and toxicity. Pharmacology & Therapeutics *134*, 82-106. 10.1016/j.pharmthera.2012.01.001.

8. Balakrishnan, L., and Bambara, R.A. (2013). Flap Endonuclease 1. Annual Review of Biochemistry *82*, 119-138. 10.1146/annurev-biochem-072511-122603.

9. Balakrishnan, L., Stewart, J., Polaczek, P., Campbell, J.L., and Bambara, R.A. (2010). Acetylation of Dna2 Endonuclease/Helicase and Flap Endonuclease 1 by p300 Promotes DNA Stability by Creating Long Flap Intermediates. Journal of Biological Chemistry *285*, 4398-4404. 10.1074/jbc.M109.086397.

10. Batool, M., Ahmad, B., and Choi, S. (2019). A Structure-Based Drug Discovery Paradigm. International Journal of Molecular Sciences *20*, 2783. 10.3390/ijms20112783.

11. Belokoneva, O.S., Villegas, E., Corzo, G., Dai, L., and Nakajima, T. (2003). The hemolytic activity of six arachnid cationic peptides is affected by the phosphatidylcholine-to-sphingomyelin ratio in lipid bilayers. Biochimica Et Biophysica Acta-Biomembranes *1617*, 22-30. 10.1016/j.bbamem.2003.08.010.

12. Bergmans, H.E.N., Vandie, I.M., and Hoekstra, W.P.M. (1981). Transformation in Escherichia coli: Stages in the process. Journal of Bacteriology *146*, 564-570. 10.1128/jb.146.2.564-570.1981

13. Bootorabi, F., Janis, J., Valjakka, J., Isoniemi, S., Vainiotalo, P., Vullo, D., Supuran, C.T., Waheed, A., Sly, W.S., Niemela, O., and Parkkila, S. (2008). Modification of carbonic anhydrase II with acetaldehyde, the first metabolite of ethanol, leads to decreased enzyme activity. BMC Biochemistry *9*, 32. 10.1186/1471-2091-9-32.

14. Bornarth, C.J., Ranalli, T.A., Henricksen, L.A., Wahl, A.F., and Bambara, R.A. (1999). Effect of flap modifications on human FEN1 cleavage. Biochemistry *38*, 13347-13354. 10.1021/bi991321u.

15. Bradford, M.M. (1976). Rapid and sensitive method for quantitation of microgram quantities of protein utilizing principle of protein-dye binding. Analytical Biochemistry *72*, 248-254. 10.1006/abio.1976.9999.

16. Bruning, J.B., and Shamoo, Y. (2004). Structural and thermodynamic analysis of human PCNA with peptides derived from DNA polymerase-delta p66 subunit and flap endonuclease-1. Structure *12*, 2209-2219. 10.1016/j.str.2004.09.018.

17. Bullock, W.O., Fernandez, J.M., and Short, J.M. (1987). XL1-blue: A high efficiency plasmid transforming recA Escherichia coli strain with beta-galactosidase selection. Biotechniques *5*, 376.

18. Burgers, P.M.J. (2009). Polymerase Dynamics at the Eukaryotic DNA Replication Fork. Journal of Biological Chemistry *284*, 4041-4045. 10.1074/jbc.R800062200.

19. Cavasotto, C.N., and Singh, N. (2008). Docking and high throughput docking: Successes and the challenge of protein flexibility. Current Computer-Aided Drug Design *4*, 221-234. 10.2174/157340908785747474.

20. Chang, M.W., Lindstrom, W., Olson, A.J., and Belew, R.K. (2007). Analysis of HIV wild-type and mutant structures via in silico docking against diverse ligand libraries. Journal of Chemical Information and Modeling *47*, 1258-1262. 10.1021/ci700044s.

21. Chapados, B.R., Hosfield, D.J., Han, S., Qiu, J., Yelent, B., Shen, B., and Tainer, J.A. (2004). Structural basis for FEN-1 substrate specificity and PCNA-mediated activation in DNA replication and repair. Cell *116*, 39-50. 10.1016/s0092-8674(03)01036-5.

22. Chen, D., Menche, G., Power, T.D., Sower, L., Peterson, J.W., and Schein, C.H. (2007a). Accounting for ligand-bound metal ions in docking small molecules on adenylyl cyclase toxins. Proteins-Structure Function and Bioinformatics *67*, 593-605. 10.1002/prot.21249.

23. Chen, J.H., Linstead, E., Swamidass, S.J., Wang, D., and Baldi, P. (2007b). ChemDB update - full-text search and virtual chemical space. Bioinformatics *23*, 2348-2351. 10.1093/bioinformatics/btm341.

24. Chernitsky, E.A., and Senkovich, O.A. (1997). Erythrocyte hemolysis by detergents. Membrane Cell Biology *11*, 475-485.

25. Choudhary, C., Kumar, C., Gnad, F., Nielsen, M.L., Rehman, M., Walther, T.C., Olsen, J.V., and Mann, M. (2009). Lysine Acetylation Targets Protein Complexes and Co-Regulates Major Cellular Functions. Science *325*, 834-840. 10.1126/science.1175371.

26. Civati, F., O'Malley, C., Erxleben, A., and McArdle, P. (2021). Factors controlling persistent needle crystal growth: The Importance of dominant one-dimensional secondary bonding, stacked structures, and van der Waals contact. Crystal Growth & Design 21, 3449-3460.

27. Congreve, M., Carr, R., Murray, C., and Jhoti, H. (2003). A rule of three for fragment-based lead discovery? Drug Discovery Today *8*, 876-877. 10.1016/s1359-6446(03)02831-9.

28. Cordes, R.M., Sims, W.B., and Glatz, C.E. (1990). Precipitation of nucleic acids with poly(ethyleneimine). Biotechnology Progress *6*, 283-285. 10.1021/bp00004a009.

29. Cyprotex Ltd. (2021). In vitro hemolysis assay protocol. <https://www.cyprotex.com/toxicology/mechanistic-toxicity/hemolysis/>

30. Deshmukh, A.L., Chandra, S., Singh, D.K., Siddiqi, M.I., and Banerjee, D. (2017). Identification of human flap endonuclease 1 (FEN1) inhibitors using a machine learning based consensus virtual screening. Molecular Biosystems *13*, 1630-1639. 10.1039/c7mb00118e.

31. Devos, J.M., Tomanicek, S.J., Jones, C.E., Nossal, N.G., and 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*, 31713-31724. 10.1074/jbc.M703209200.

32. Diaz, A., Lacks, S.A., and Lopez, P. (1992). The 5′ to 3′ exonuclease activity of DNA polymerase I is essential for Streptococcus pneumoniae. Molecular Microbiology *6*, 3009-3019. 10.1111/j.1365-2958.1992.tb01759.x.

33. Dittrich, C., Greim, G., Borner, M., Weigang-Kohler, K., Huisman, H., Amelsberg, A., Ehret, A., Wanders, J., Hanauske, A., Fumoleau, P., and Grp, E.E.C.S. (2002). Phase I and pharmacokinetic study of BIBX 1382 BS, an epidermal growth factor receptor (EGFR) inhibitor, given in a continuous daily oral administration. European Journal of Cancer *38*, 1072-1080. 10.1016/s0959-8049(02)00020-5.

34. Doherty, A.J., Serpell, L.C., and Ponting, C.P. (1996). The helix-hairpin-helix DNA-binding motif: A structural basis for non-sequence-specific recognition of DNA. Nucleic Acids Research *24*, 2488-2497. 10.1093/nar/24.13.2488.

35. Donati, E., Genna, V., and De Vivo, M. (2020). Recruiting Mechanism and Functional Role of a Third Metal Ion in the Enzymatic Activity of 5 ' Structure-Specific Nucleases. Journal of the American Chemical Society *142*, 2823-2834. 10.1021/jacs.9b10656.

36. Dore, A.S., Kilkenny, M.L., Jones, S.A., Oliver, A.W., Roe, S.M., Bell, S.D., and Pearl, L.H. (2006). Structure of an archaeal PCNA1-PCNA2-FEN1 complex: elucidating PCNA subunit and client enzyme specificity. Nucleic Acids Research *34*, 4515-4526. 10.1093/nar/gkl623.

37. Duffy, C.M., Hilbert, B.J., and Kelch, B.A. (2016). A Disease-Causing Variant in PCNA Disrupts a Promiscuous Protein Binding Site. Journal of Molecular Biology *428*, 1023-1040. 10.1016/j.jmb.2015.11.029.

38. Dupureur, C.M. (2010). One is enough: insights into the two-metal ion nuclease mechanism from global analysis and computational studies. Metallomics *2*, 609-620. 10.1039/c0mt00013b.

39. Ebbensgaard, A., Mordhorst, H., Overgaard, M.T., Aarestrup, F.M., and Hansen, E.B. (2018). Dissection of the antimicrobial and hemolytic activity of Cap18: Generation of Cap18 derivatives with enhanced specificity. Plos One *13*, e0197742. 10.1371/journal.pone.0197742.

40. Englebienne, P., and Moitessier, N. (2009). Docking Ligands into Flexible and Solvated Macromolecules. 4. Are Popular Scoring Functions Accurate for this Class of Proteins? Journal of Chemical Information and Modeling *49*, 1568-1580. 10.1021/ci8004308.

41. Ericksen, S.S., Wu, H.Z., Zhang, H.K., Michael, L.A., Newton, M.A., Hoffmann, F.M., and Wildman, S.A. (2017). Machine Learning Consensus Scoring Improves Performance Across Targets in Structure-Based Virtual Screening. Journal of Chemical Information and Modeling *57*, 1579-1590. 10.1021/acs.jcim.7b00153.

42. Evans, P. (2006). Scaling and assessment of data quality. Acta Crystallographica Section D-Biological Crystallography *62*, 72-82. 10.1107/s0907444905036693.

43. Evans, P.R., and Murshudov, G.N. (2013). How good are my data and what is the resolution? Acta Crystallographica D-Biological Crystallography *69*, 1204-1214. 10.1107/s0907444913000061.

44. Exell, J.C., Thompson, M.J., Finger, L.D., Shaw, S.J., Debreczeni, J., Ward, T.A., McWhirter, C., Sioberg, C.L.B., Molina, D.M., Abbott, W.M., *et al*. (2016). Cellularly active N-hydroxyurea FEN1 inhibitors block substrate entry to the active site. Nature Chemical Biology *12*, 815-821. 10.1038/nchembio.2148.

45. Farag, M.R., and Alagawany, M. (2018). Erythrocytes as a biological model for screening of xenobiotics toxicity. Chemico-Biological Interactions *279*, 73-83. 10.1016/j.cbi.2017.11.007.

46. FDA guidance – Nonclinical Safety Studies for the Conduct of Human Clinical Trials and Marketing Authorization for Pharmaceuticals. (2010). <https://www.fda.gov/media/71542/download>

47. Feinstein, W.P., and Brylinski, M. (2015). Calculating an optimal box size for ligand docking and virtual screening against experimental and predicted binding pockets. Journal of Cheminformatics *7*, 18. 10.1186/s13321-015-0067-5.

48. Feng, M., Patel, D., Dervan, J.J., Ceska, T., Suck, D., Haq, I., and Sayers, J.R. (2004). Roles of divalent metal ions in flap endonuclease-substrate interactions. Nature Structural & Molecular Biology *11*, 450-456. 10.1038/nsmb754.

49. Flier, J.S. (2019). Academia and industry: allocating credit for discovery and development of new therapies. Journal of Clinical Investigation *129*, 2172-2174. 10.1172/jci129122.

50. Forli, S., Huey, R., Pique, M.E., Sanner, M.F., Goodsell, D.S., and Olson, A.J. (2016). Computational protein-ligand docking and virtual drug screening with the AutoDock suite. Nature Protocols *11*, 905-919. 10.1038/nprot.2016.051.

51. Frank, G., Qiu, J.H., Zheng, L., and Shen, B.H. (2001). Stimulation of eukaryotic flap endonuclease-1 activities by proliferating cell nuclear antigen (PCNA) is independent of its in vitro interaction via a consensus PCNA binding region. Journal of Biological Chemistry *276*, 36295-36302. 10.1074/jbc.M103397200.

52. Frearson, J., and Wyatt, P. (2010). Drug discovery in academia: the third way? Expert Opinion on Drug Discovery *5*, 909-919. 10.1517/17460441.2010.506508.

53. Friedrich-Heineken, E., and Hubscher, U. (2004). The Fen1 extrahelical 3 '-flap pocket is conserved from archaea to human and regulates DNA substrate specificity. Nucleic Acids Research *32*, 2520-2528. 10.1093/nar/gkh576.

54. Fukushima, S., Itaya, M., Kato, H., Ogasawara, N., and Yoshikawa, H. (2007). Reassessment of the in vivo functions of DNA polymerase I and RNase h in bacterial cell growth. Journal of Bacteriology *189*, 8575-8583. 10.1128/jb.00653-07.

55. Gardiner, P., Cox, R.J., and Grime, K. (2019). Plasma protein binding as an optimisable parameter for acidic drugs. Drug Metabolism and Diposition 47, 865-873. 10.1124/dmd.119.087163.

56. Garforth, S.J., Ceska, T.A., Suck, D., and Sayers, J.R. (1999). Mutagenesis of conserved lysine residues in bacteriophage T5 5 '-3 ' exonuclease suggests separate mechanisms of endo- and exonucleolytic cleavage. Proceedings of the National Academy of Sciences of the United States of America *96*, 38-43. 10.1073/pnas.96.1.38.

57. Garforth, S.J., and Sayers, J.R. (1997). Structure-specific DNA binding by bacteriophage T5 5'->3' exonuclease. Nucleic Acids Research *25*, 3801-3807. 10.1093/nar/25.19.3801.

58. Garg, P., Stith, C.M., Sabouri, N., Johansson, E., and Burgers, P.M. (2004). Idling by DNA polymerase delta maintains a ligatable nick during lagging-strand DNA replication. Genes & Development *18*, 2764-2773. 10.1101/gad.1252304.

59. Garvey, C., McBride, T., Nevin, L., Peiperl, L., Ross, A., Simpson, P., Turner, R., and Editors, P.M. (2016). Antimicrobial Resistance: Is the World UNprepared? Plos Medicine *13*, e1002130. 10.1371/journal.pmed.1002130.

60. Ghosh, S., Goldgur, Y., and Shuman, S. (2020). Mycobacterial DNA polymerase I: activities and crystal structures of the POL domain as apoenzyme and in complex with a DNA primer-template and of the full-length FEN/EXO POL enzyme. Nucleic Acids Research *48*, 3165-3180. 10.1093/nar/gkaa075.

61. Giglione, C., Boularot, A., and Meinnel, T. (2004). Protein N-terminal methionine excision. Cellular and Molecular Life Sciences *61*, 1455-1474. 10.1007/s00018-004-3466-8.

62. Gilbert, D.N., Guidos, R.J., Boucher, H.W., Talbot, G.H., Spellberg, B., Edwards, J.E., Jr., Scheld, W.M., Bradley, J.S., Bartlett, J.G., and Infect Dis Soc, A. (2010). The 10 x '20 Initiative: Pursuing a Global Commitment to Develop 10 New Antibacterial Drugs by 2020. Clinical Infectious Diseases *50*, 1081-1083. 10.1086/652237.

63. Gloor, J.W., Balakrishnan, L., and Bambara, R.A. (2010). Flap Endonuclease 1 Mechanism Analysis Indicates Flap Base Binding Prior to Threading. Journal of Biological Chemistry *285*, 34922-34931. 10.1074/jbc.M110.165902.

64. Gorrec, F. (2009). The MORPHEUS protein crystallization screen. Journal of Applied Crystallography *42*, 1035-1042. 10.1107/s0021889809042022.

65. Grant, S.G.N., Jessee, J., Bloom, F.R., and Hanahan, D. (1990). Differential plasmid rescue from transgenic mouse DNAs into Escherichia coli methylation-restriction mutants. Proceedings of the National Academy of Sciences of the United States of America *87*, 4645-4649. 10.1073/pnas.87.12.4645.

66. Greco, I., Molchanova, N., Holmedal, E., Jenssen, H., Hummel, B.D., Watts, J.L., Hakansson, J., Hansen, P.R., and Svenson, J. (2020). Correlation between hemolytic activity, cytotoxicity and systemic in vivo toxicity of synthetic antimicrobial peptides. Scientific Reports *10*, 13206. 10.1038/s41598-020-69995-9.

67. Greene, A.L., Snipe, J.R., Gordenin, D.A., and Resnick, M.A. (1999). Functional analysis of human FEN1 in Saccharomyces cerevisiae and its role in genome stability. Human Molecular Genetics *8*, 2263-2273. 10.1093/hmg/8.12.2263.

68. Guo, Z.G., Kanjanapangka, J., Liu, N., Liu, S.B., Liu, C.W., Wu, Z.X., Wang, Y.J., Loh, T., Kowolik, C., Jamsen, J., *et al*. (2012). Sequential Posttranslational Modifications Program FEN1 Degradation during Cell-Cycle Progression. Molecular Cell *47*, 444-456. 10.1016/j.molcel.2012.05.042.

69. Guo, Z.G., Zheng, L., Xu, H., Dai, H.F., Zhou, M.A., Pascua, M.R., Chen, Q.M., and Shen, B.H. (2010). Methylation of FEN1 suppresses nearby phosphorylation and facilitates PCNA binding. Nature Chemical Biology *6*, 766-773. 10.1038/nchembio.422.

70. Haas, K.L., and Franz, K.J. (2009). Application of Metal Coordination Chemistry To Explore and Manipulate Cell Biology. Chemical Reviews *109*, 4921-4960. 10.1021/cr900134a.

71. Harrington, J.J., and Lieber, M.R. (1994). The characterization of a mammalian DNA structure-specific endonuclease. Embo Journal *13*, 1235-1246. 10.1002/j.1460-2075.1994.tb06373.x.

72. Harrington, J.J., and Lieber, M.R. (1995). DNA structural elements required for FEN-1 binding. Journal of Biological Chemistry *270*, 4503-4508.

73. Hasan, S., Stucki, M., Hassa, P.O., Imhof, R., Gehrig, P., Hunziker, P., Hubscher, U., and Hottiger, M.O. (2001). Regulation of human flap endonuclease-1 activity by acetylation through the transcriptional coactivator p300. Molecular Cell *7*, 1221-1231. 10.1016/s1097-2765(01)00272-6.

74. He, L.F., Zhang, Y.L., Sun, H.F., Jiang, F., Yang, H., Wu, H., Zhou, T., Hu, S.C., Kathera, C.S., Wang, X.J., *et al*. (2016). Targeting DNA Flap Endonuclease 1 to Impede Breast Cancer Progression. Ebiomedicine *14*, 32-43. 10.1016/j.ebiom.2016.11.012.

75. Helleday, T., Petermann, E., Lundin, C., Hodgson, B., and Sharma, R.A. (2008). DNA repair pathways as targets for cancer therapy. Nature Reviews Cancer *8*, 193-204. 10.1038/nrc2342.

76. Henneke, G., Koundrioukoff, S., and Hubscher, U. (2003). Phosphorylation of human Fen1 by cyclin-dependent kinase modulates its role in replication fork regulation. Oncogene *22*, 4301-4313. 10.1038/sj.onc.1206606.

77. Heyduk, T., Ma, Y.X., Tang, H., and Ebright, R.H. (1996). Fluorescence anisotropy: Rapid, quantitative assay for protein-DNA and protein-protein interaction. Methods in Enzymology *274*, 492-503. 10.1016/s0076-6879(96)74039-9.

78. Hobbs, L.J., and Nossal, N.G. (1996). Either bacteriophage T4 RNase H or Escherichia coli DNA polymerase I is essential for phage replication. Journal of Bacteriology *178*, 6772-6777. 10.1128/jb.178.23.6772-6777.1996.

79. Hosfield, D.J., Mol, C.D., Shen, B.H., and Tainer, J.A. (1998). Structure of the DNA repair and replication endonuclease and exonuclease FEN-1: Coupling DNA and PCNA binding to FEN-1 activity. Cell *95*, 135-146. 10.1016/s0092-8674(00)81789-4.

80. Houston, D.R., and Walkinshaw, M.D. (2013). Consensus Docking: Improving the Reliability of Docking in a Virtual Screening Context. Journal of Chemical Information and Modeling *53*, 384-390. 10.1021/ci300399w.

81. Howard, J.A.K., Hoy, V.J., Ohagan, D., and Smith, G.T. (1996). How good is fluorine as a hydrogen bond acceptor? Tetrahedron *52*, 12613-12622. 10.1016/0040-4020(96)00749-1.

82. Hu, X., Balaz, S., and Shelver, W.H. (2004). A practical approach to docking of zinc metalloproteinase inhibitors. Journal of Molecular Graphics & Modelling *22*, 293-307. 10.1016/j.jmgm.2003.11.002.

83. Hu, X., and Shelver, W.H. (2003). Docking studies of matrix metalloproteinase inhibitors: zinc parameter optimization to improve the binding free energy prediction. Journal of Molecular Graphics & Modelling *22*, 115-126. 10.1016/s1093-3263(03)00153-0.

84. Hutzler, J.M., Obach, R.S., Dalvie, D., and Zientek, M.A. (2013). Strategies for a comprehensive understanding of metabolism by aldehyde oxidase. Expert Opinion on Drug Metabolism & Toxicology *9*, 153-168. 10.1517/17425255.2013.738668.

85. Hwang, K.Y., Baek, K., Kim, H.Y., and Cho, Y. (1998). The crystal structure of flap endonuclease-1 from Methanococcus jannaschii. Nature Structural Biology *5*, 707-713. 10.1038/1406.

86. Iacobuzio-Donahue, C.A., Maitra, A., Olsen, M., Lowe, A.W., Van Heek, N.T., Rosty, C., Walter, K., Sato, N., Parker, A., Ashfaq, R., *et al*. (2003). Exploration of global gene expression patterns in pancreatic adenocarcinoma using cDNA microarrays. American Journal of Pathology *162*, 1151-1162. 10.1016/s0002-9440(10)63911-9.

87. Irwin, J.J., Raushel, F.M., and Shoichet, B.K. (2005). Virtual screening against metalloenzymes for inhibitors and substrates. Biochemistry *44*, 12316-12328. 10.1021/bi050801k.

88. Jin, Y.H., Ayyagari, R., Resnick, M.A., Gordenin, D.A., and Burgers, P.M.J. (2003). Okazaki fragment maturation in yeast - II. Cooperation between the polymerase and 3 '-5 '-exonuclease activities of Pol delta in the creation of a ligatable nick. Journal of Biological Chemistry *278*, 1626-1633. 10.1074/jbc.M209803200.

89. Jin, Y.H., Obert, R., Burgers, P.M.J., Kunkel, T.A., Resnick, M.A., and Gordenin, D.A. (2001). The 3 '-> 5 ' exonuclease of DNA polymerase delta can substitute for the 5' flap endonuclease Rad27/Fen1 in processing Okazaki fragments and preventing genome instability.

Proceedings of the National Academy of Sciences of the United States of America *98*, 5122-5127. 10.1073/pnas.091095198.

90. Judson, R., Kavlock, R., Martin, M., Reif, D., Houck, K., Knudsen, T., Richard, A., Tice, R.R., Whelan, M., Xia, M.H., *et al*. (2013). Perspectives on Validation of High-Throughput Assays Supporting 21st Century Toxicity Testing. Altex-Alternatives to Animal Experimentation *30*, 51-66. 10.14573/altex.2013.1.051.

91. Kabsch, W. (2010a). Integration, scaling, space-group assignment and post-refinement. Acta Crystallographica Section D-Biological Crystallography *66*, 133-144. 10.1107/s0907444909047374.

92. Kantardjieff, K.A., and Rupp, B. (2003). Matthews coefficient probabilities: Improved estimates for unit cell contents of proteins, DNA, and protein-nucleic acid complex crystals. Protein Science *12*, 1865-1871. 10.1110/ps.0350503.

93. Kao, H.-I., Henricksen, L.A., Liu, Y., and Bambara, R.A. (2002). Cleavage specificity of Saccharomyces cerevisiae flap endonuclease 1 suggests a double-flap structure as the cellular substrate. Journal of Biological Chemistry *277*, 14379-14389. 10.1074/jbc.M110662200.

94. Kao, H.I., Veeraraghavan, J., Polaczek, P., Campbell, J.L., and Bambara, R.A. (2004). On the roles of Saccharomyces cerevisiae Dna2p and flap endonuclease 1 in Okazaki fragment processing. Journal of Biological Chemistry *279*, 15014-15024. 10.1074/jbc.M313216200.

95. Karplus, P.A., and Diederichs, K. (2012). Linking Crystallographic Model and Data Quality. Science *336*, 1030-1033. 10.1126/science.1218231.

96. Kim, J.M., Sohn, H.Y., Yoon, S.Y., Oh, J.H., Yang, J.O., Kim, J.H., Song, K.S., Rho, S.M., Yoo, H.S., Kim, Y.S., *et al*. (2005). Identification of gastric cancer-related genes using a cDNA microarray containing novel expressed sequence tags expressed in gastric cancer cells. Clinical Cancer Research *11*, 473-482.

97. Kim, K., Biade, S., and Matsumoto, Y. (1998). Involvement of flap endonuclease 1 in base excision DNA repair. Journal of Biological Chemistry *273*, 8842-8848. 10.1074/jbc.273.15.8842.

98. Kim, Y., Eom, S.H., Wang, J.M., Lee, D.S., Suh, S.W., and Steitz, T.A. (1995). Crystal structure of Thermus aquaticus DNA polymerase. Nature *376*, 612-616. 10.1038/376612a0.

99. Kishimoto, W., Ishiguro, N., Ludwig-Schwellinger, E., Ebner, T., and Schaefer, O. (2014). In Vitro Predictability of Drug-Drug Interaction Likelihood of P-Glycoprotein-Mediated Efflux of Dabigatran Etexilate Based on I (2)/IC50 Threshold. Drug Metabolism and Disposition *42*, 257-263. 10.1124/dmd.113.053769.

100. Krause, A., Combaret, V., Iacono, I., Lacroix, B., Compagnon, C., Bergeron, C., Valsesia-Wittmann, S., Leissner, P., Mougin, B., and Puisieux, A. (2005). Genome-wide analysis of gene expression in neuroblastomas detected by mass screening. Cancer Letters *225*, 111-120. 10.1016/j.canlet.2004.10.035.

101. Kucherlapati, M., Yan, K., Kuraguchi, M., Zhao, J., Lia, M., Heyer, J., Kane, M.F., Fan, K.H., Russell, R., Brown, A.M.C., *et al*. (2002). Haploinsufficiency of Flap endonuclease

(Fen1) leads to rapid tumor progression. Proceedings of the National Academy of Sciences of the United States of America *99*, 9924-9929. 10.1073/pnas.152321699.

102. Kukol, A. (2011). Consensus virtual screening approaches to predict protein ligands. European Journal of Medicinal Chemistry *46*, 4661-4664. 10.1016/j.ejmech.2011.05.026.

103. Lancey, C., Tehseen, M., Raducanu, V.S., Rashid, F., Merino, N., Ragan, T.J., Savva, C.G., Zaher, M.S., Shirbini, A., Blanco, F.J., *et al*. (2020). Structure of the processive human Pol delta holoenzyme. Nature Communications *11*, 1109. 10.1038/s41467-020-14898-6.

104. Larsen, E., Gran, C., Saether, B.E., Seeberg, E., and Klungland, A. (2003). Proliferation failure and gamma radiation sensitivity of Fen1 null mutant mice at the blastocyst stage. Molecular and Cellular Biology *23*, 5346-5353. 10.1128/mcb.23.15.5346-5353.2003.

105. Larsen, E., Kleppa, L., Meza, T.J., Meza-Zepeda, L.A., Rada, C., Castellanos, C.G., Lien, G.F., Nesse, G.J., Neuberger, M.S., Laerdahl, J.K., *et al*. (2008). Early-onset lymphoma and extensive embryonic apoptosis in two domain-specific Fen1 mice mutants. Cancer Research *68*, 4571-4579. 10.1158/0008-5472.can-08-0168.

106. Larson, S.B., Day, J.S., Nguyen, C., Cudney, R., and McPherson, A. (2008). Progress in the development of an alternative approach to macromolecular crystallization. Crystal Growth & Design *8*, 3038-3052. 10.1021/cg800174n.

107. LaTulippe, E., Satagopan, J., Smith, A., Scher, H., Scardino, P., Reuter, V., and Gerald, W.L. (2002). Comprehensive gene expression analysis of prostate cancer reveals distinct transcriptional programs associated with metastatic disease. Cancer Research *62*, 4499-4506.

108. Ledford, H. (2015). End of cancer atlas prompts rethink. Nature *517*, 128-129. 10.1038/517128a.

109. Lee, S.-H., Princz, L.N., Kluegel, M.F., Habermann, B., Pfander, B., and Biertuempfel, C. (2015). Human Holliday junction resolvase GEN1 uses a chromodomain for efficient DNA recognition and cleavage. Elife *4*, e12256. 10.7554/eLife.12256.

110. Leeson, P.D., and Springthorpe, B. (2007). The influence of drug-like concepts on decision-making in medicinal chemistry. Nature Reviews Drug Discovery *6*, 881-890. 10.1038/nrd2445.

111. Lipinski, C., and Hopkins, A. (2004). Navigating chemical space for biology and medicine. Nature *432*, 855-861. 10.1038/nature03193.

112. Lipinski, C.A., Lombardo, F., Dominy, B.W., and Feeney, P.J. (2001). Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Advanced Drug Delivery Reviews *46*, 3-26. 10.1016/s0169-409x(00)00129-0.

113. Lippert, T., and Rarey, M. (2009). Fast automated placement of polar hydrogen atoms in protein-ligand complexes. Journal of Cheminformatics *1*, 13. 10.1186/1758-2946-1-13.

114. Liu, Y., Kao, H.I., and Bambara, R.A. (2004). Flap endonuclease 1: A central component of DNA metabolism. Annual Review of Biochemistry *73*, 589-615. 10.1146/annurev.biochem.73.012803.092453.

115. Ma, L., Cao, X., Wang, H., Lu, K., Wang, Y., Tu, C., Dai, Y., Meng, Y., Li, Y., Yu, P., *et al*. (2019). Discovery of Myricetin as a Potent Inhibitor of Human Flap Endonuclease 1, Which Potentially Can Be Used as Sensitizing Agent against HT-29 Human Colon Cancer Cells. Journal of Agricultural and Food Chemistry *67*, 1656-1665. 10.1021/acs.jafc.8b05447.

116. Macarron, R., Banks, M.N., Bojanic, D., Burns, D.J., Cirovic, D.A., Garyantes, T., Green, D.V.S., Hertzberg, R.P., Janzen, W.P., Paslay, J.W., *et al*. (2011). Impact of high-throughput screening in biomedical research. Nature Reviews Drug Discovery *10*, 188-195. 10.1038/nrd3368.

117. Maga, G., Villani, G., Tillement, V., Stucki, M., Locatelli, G.A., Frouin, I., Spadari, S., and Hubscher, U. (2001). Okazaki fragment processing: Modulation of the strand displacement activity of DNA polymerase delta by the concerted action of replication protein A, proliferating cell nuclear antigen, and flap endonuclease-1. Proceedings of the National Academy of Sciences of the United States of America *98*, 14298-14303. 10.1073/pnas.251193198.

118. Maia, E.H.B., Assis, L.C., de Oliveira, T.A., da Silva, A.M., and Taranto, A.G. (2020). Structure-Based Virtual Screening: From Classical to Artificial Intelligence. Frontiers in Chemistry *8*, 343. 10.3389/fchem.2020.00343.

119. Mannhold, R., Poda, G.I., Ostermann, C., and Tetko, I.V. (2009). Calculation of Molecular Lipophilicity: State-of-the-Art and Comparison of Log P Methods on More Than 96,000 Compounds. Journal of Pharmaceutical Sciences *98*, 861-893. 10.1002/jps.21494.

120. Mase, T., Kubota, K., Miyazono, K.-i., Kawarabayasi, Y., and Tanokura, M. (2011). Structure of flap endonuclease 1 from the hyperthermophilic archaeon Desulfurococcus amylolyticus. Acta Crystallographica F Structural Biology Communications *67*, 209-213. 10.1107/s1744309110053030.

121. Matsui, E., Musti, K.V., Abe, J., Yamasaki, K., Matsui, I., and Harata, K. (2002). Molecular structure and novel DNA binding sites located in loops of flap endonuclease-1 from Pyrococcus horikoshii. Journal of Biological Chemistry *277*, 37840-37847. 10.1074/jbc.M205235200.

122. Matsuzaki, K. (2009). Control of cell selectivity of antimicrobial peptides. Biochimica Et Biophysica Acta-Biomembranes *1788*, 1687-1692. 10.1016/j.bbamem.2008.09.013.

123. Matthews, B.W. (1968). Solvent content of protein crystals. Journal of Molecular Biology *33*, 491-497. 10.1016/0022-2836(68)90205-2.

124. McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., and Read, R.J. (2007). Phaser crystallographic software. Journal of Applied Crystallography *40*, 658-674. 10.1107/s0021889807021206.

125. McPherson, A., and Cudney, B. (2006). Searching for silver bullets: An alternative strategy for crystallizing macromolecules. Journal of Structural Biology *156*, 387-406. 10.1016/j.jsb.2006.09.006.

126. Meng, X.Y., Zhang, H.X., Mezei, M., and Cui, M. (2011). Molecular Docking: A Powerful Approach for Structure-Based Drug Discovery. Current Computer Aided Drug Design *7*, 146-157. 10.2174/157340911795677602.

127. Mierendorf, R.C., Morris, B.B., Hammer, B., and Novy, R.E. (1998). Expression and purification of recombinant proteins using the pET system. Molecular Diagnosis of Infectious Diseases *13*, 257-292. 10.1385/0-89603-485-2:257.

128. Morris, G.M., Huey, R., Lindstrom, W., Sanner, M.F., Belew, R.K., Goodsell, D.S., and Olson, A.J. (2009). AutoDock4 and AutoDockTools4: Automated Docking with Selective Receptor Flexibility. Journal of Computational Chemistry *30*, 2785-2791. 10.1002/jcc.21256.

129. Mueser, T.C., Nossal, N.G., and 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*, 1101-1112. 10.1016/s0092-8674(00)81310-0.

130. Mullard, A. (2020). 2019 FDA drug approvals. Nature Reviews Drug Discovery *19*, 79-84.

131. Murante, R.S., Rust, L., and 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*, 30377-30383. 10.1074/jbc.270.51.30377.

132. Murray, C.W., and Rees, D.C. (2009). The rise of fragment-based drug discovery. Nature Chemistry *1*, 187-192. 10.1038/nchem.217.

133. Nagata, Y., Mashimo, K., Kawata, M., and Yamamoto, K. (2002). The roles of Klenow processing and flap processing activities of DNA polymerase I in chromosome instability in Escherichia coli K12 strains. Genetics *160*, 13-23.

134. Nazarkina, Z.K., Lavrik, O.I., and Khodyreva, S.N. (2008). Flap endonuclease 1 and its role in eukaryotic DNA metabolism. Molecular Biology *42*, 357-370. 10.1134/s0026893308030035.

135. Nikolova, T., Christmann, M., and Kaina, B. (2009). FEN1 is Overexpressed in Testis, Lung and Brain Tumors. Anticancer Research *29*, 2453-2459.

136. O'Boyle, N.M., Banck, M., James, C.A., Morley, C., Vandermeersch, T., and Hutchison, G.R. (2011). Open Babel: An open chemical toolbox. Journal of Cheminformatics *3*, 33. 10.1186/1758-2946-3-33.

137. Olsen, J.V., Vermeulen, M., Santamaria, A., Kumar, C., Miller, M.L., Jensen, L.J., Gnad, F., Cox, J., Jensen, T.S., Nigg, E.A., *et al.* (2010). Quantitative Phosphoproteomics Reveals Widespread Full Phosphorylation Site Occupancy During Mitosis. Science Signaling *3*, ra3. 10.1126/scisignal.2000475.

138. Orans, J., McSweeney, E.A., Iyer, R.R., Hast, M.A., Hellinga, H.W., Modrich, P., and Beese, L.S. (2011). Structures of Human Exonuclease 1 DNA Complexes Suggest a Unified Mechanism for Nuclease Family. Cell *145*, 212-223. 10.1016/j.cell.2011.03.005.

139. Osguthorpe, D.J., Sherman, W., and Hagler, A.T. (2012). Exploring Protein Flexibility: Incorporating Structural Ensembles From Crystal Structures and Simulation into Virtual Screening Protocols. Journal of Physical Chemistry B *116*, 6952-6959. 10.1021/jp3003992.

140. Pagano, M., and Faggio, C. (2015). The use of erythrocyte fragility to assess xenobiotic cytotoxicity. Cell Biochemistry and Function *33*, 351-355. 10.1002/cbf.3135.

141. Palacio-Rodriguez, K., Lans, I., Cavasotto, C.N., and Cossio, P. (2019). Exponential consensus ranking improves the outcome in docking and receptor ensemble docking. Scientific Reports *9*, 5142. 10.1038/s41598-019-41594-3.

142. Panda, H., Jaiswal, A.S., Corsino, P.E., Armas, M.L., Law, B.K., and Narayan, S. (2009). Amino Acid Asp181 of 5 '-Flap Endonuclease 1 Is a Useful Target for Chemotherapeutic Development. Biochemistry *48*, 9952-9958. 10.1021/bi9010754.

143. Parenteau, J., and Wellinger, R.J. (1999). Accumulation of single-stranded DNA and destabilization of telomeric repeats in yeast mutant strains carrying a deletion of RAD27. Molecular and Cellular Biology *19*, 4143-4152. 10.1128/mcb.19.6.4143.

144. Parenteau, J., and Wellinger, R.J. (2002). Differential processing of leading- and lagging-strand ends at Saccharomyces cerevisiae telomeres revealed by the absence of Rad27p nuclease. Genetics *162*, 1583-1594.

145. Payne, D.J., Gwynn, M.N., Holmes, D.J., and Pompliano, D.L. (2007). Drugs for bad bugs: confronting the challenges of antibacterial discovery. Nature Reviews Drug Discovery *6*, 29-40. 10.1038/nrd2201.

146. Pflugrath, J.W. (2004). Macromolecular cryocrystallography - methods for cooling and mounting protein crystals at cryogenic temperatures. Methods *34*, 415-423. 10.1016/j.ymeth.2004.03.032.

147. Pottel, J., Therrien, E., Gleason, J.L., and Moitessier, N. (2014). Docking Ligands into Flexible and Solvated Macromolecules. 6. Development and Application to the Docking of HDACs and other Zinc Metalloenzymes Inhibitors. Journal of Chemical Information and Modeling *54*, 254-265. 10.1021/ci400550m.

148. Qiu, J.Z., Bimston, D.N., Partikian, A., and Shen, B.H. (2002). Arginine residues 47 and 70 of human flap endonuclease-1 are involved in DNA substrate interactions and cleavage site determination. Journal of Biological Chemistry *277*, 24659-24666. 10.1074/jbc.M111941200.

149. Ragheb, M.N., Thomason, M.K., Hsu, C., Nugent, P., Gage, J., Samadpour, A.N., Kariisa, A., Merrikh, C.N., Miller, S.I., Sherman, D.R., and Merrikh, H. (2019). Inhibiting the Evolution of Antibiotic Resistance. Molecular Cell *73*, 157-165. 10.1016/j.molcel.2018.10.015.

150. Reagan, M.S., Pittenger, C., Siede, W., and Friedberg, E.C. (1995). Characterizations of a mutant strain of Saccharomyces cerevisiae with a deletion of the RAD27 gene, a structural homolog of the RAD2 nucleotide excision repair gene. Journal of Bacteriology *177*, 364-371. 10.1128/jb.177.2.364-371.1995.

151. Riccardi, L., Genna, V., and De Vivo, M. (2018). Metal-ligand interactions in drug design. Nature Reviews Chemistry *2*, 100-112. 10.1038/s41570-018-0018-6.

152. Robins, P., Pappin, D.J.C., Wood, R.D., and Lindahl, T. (1994). Structural and functional homology between mammalian DNase IV and the 5′-nuclease domain of Escherichia coli DNA polymerase I. Journal of Biological Chemistry *269*, 28535-28538.

153. Rodriguez, A., Villegas, E., Montoya-Rosales, A., and Rivas-Santiago, B. (2014). Characterization of Antibacterial and Hemolytic Activity of Synthetic Pandinin 2 Variants and

Their Inhibition against Mycobacterium tuberculosis. Plos One *9*, e101742. 10.1371/journal.pone.0101742.

154. Rossi, M.L., and Bambara, R.A. (2006). Reconstituted Okazaki fragment processing indicates two pathways of primer removal. Journal of Biological Chemistry *281*, 26051-26061. 10.1074/jbc.M604805200.

155. Sakurai, S., Kitano, K., Yamaguchi, H., Hamada, K., Okada, K., Fukuda, K., Uchida, M., Ohtsuka, E., Morioka, H., and Hakoshima, T. (2005). Structural basis for recruitment of human flap endonuclease 1 to PCNA. Embo Journal *24*, 683-693. 10.1038/sj.emboj.7600519.

156. Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989). Molecular Cloning: a Laboratory Manual, Second Edition, Vols. 1, 2 and 3. Cold Spring Harbor Laboratory Press: Cold Spring Harbor, New York, USA. Illus.

157. Sampathi, S., Bhusari, A., Shen, B.H., and Chai, W.H. (2009). Human Flap Endonuclease I Is in Complex with Telomerase and Is Required for Telomerase-mediated Telomere Maintenance. Journal of Biological Chemistry *284*, 3682-3690. 10.1074/jbc.M805362200.

158. Sander, T., Freyss, J., von Korff, M., and Rufener, C. (2015). Data Warrior: An Open-Source Program For Chemistry Aware Data Visualization And Analysis. Journal of Chemical Information and Modeling *55*, 460-473. 10.1021/ci500588j.

159. Sanoh, S., Tayama, Y., Sugihara, K., Kitamura, S., and Ohta, S. (2015). Significance of aldehyde oxidase during drug development: Effects on drug metabolism, pharmacokinetics, toxicity, and efficacy. Drug Metabolism and Pharmacokinetics *30*, 52-63. 10.1016/j.dmpk.2014.10.009.

160. Sato, M., Girard, L., Sekine, I., Sunaga, N., Ramirez, R.D., Kamibayashi, C., and Minna, J.D. (2003). Increased expression and no mutation of the Flap endonuclease (FEN1) gene in human lung cancer. Oncogene *22*, 7243-7246. 10.1038/sj.onc.1206977.

161. Savage, N. (2020). Overcoming resistance. Nature *586*, S55-S56. 10.1038/d41586-020-02886-1.

162. Sayers, J.R., and Eckstein, F. (1991). A single-strand specific endonuclease activity copurifies with overexpressed T5 D15 exonuclease. Nucleic Acids Research *19*, 4127-4132. 10.1093/nar/19.15.4127.

163. Setlow, P., and Kornberg, A. (1972). Deoxyribonucleic acid polymerase: two distinct enzymes in one polypeptide. II. A proteolytic fragment containing the 5′ leads to 3′ exonuclease function. Restoration of intact enzyme functions from the two proteolytic fragments. Journal of Biological Chemistry *247*, 232-240.

164. Shah, S., Dunten, P., Stiteler, A., Park, C.K., and Horton, N.C. (2015). Structure and specificity of FEN-1 from Methanopyrus kandleri. Proteins-Structure Function and Bioinformatics *83*, 188-194. 10.1002/prot.24704.

165. Sharma, S., Sommers, J.A., and Brosh, R.M. (2004). In vivo function of the conserved non-catalytic domain of Werner syndrome helicase in DNA replication. Human Molecular Genetics *13*, 2247-2261. 10.1093/hmg/ddh234.

166. Sharma, S., Sommers, J.A., Gary, R.K., Friedrich-Heineken, E., Hubscher, U., and Brosh, R.M. (2005). The interaction site of Flap Endonuclease-1 with WRN helicase suggests a coordination of WRN and PCNA. Nucleic Acids Research *33*, 6769-6781. 10.1093/nar/gki1002.

167. Shen, B.H., Nolan, J.P., Sklar, L.A., and Park, M.S. (1996). Essential amino acids for substrate binding and catalysis of human flap endonuclease 1. Journal of Biological Chemistry *271*, 9173-9176. 10.1074/jbc.271.16.9173.

168. Shen, B., Singh, P., Liu, R., Qiu, J., Zheng, L., Finger, D., and Alas, S. (2005). Multiple but dissectible functions od FEN-1 nucleases in nucleic acid processing, genome stability and diseases. BioEssays *27*, 717-729. 10.1002/bies.20255.

169. Shi, R., Wang, Y., Gao, Y., Xu, X., Mao, S., Xiao, Y., Song, S., Wang, L., Tian, B., Zhao, Y., *et al*. (2020). Succinylation at a key residue of FEN1 is involved in the DNA damage response to maintain genome stability. American Journal of Cell Physiology *4*, 657-666. 10.1152/ajpcell.00137.2020.

170. Shi, Y.Q., Hellinga, H.W., and Beese, L.S. (2017). Interplay of catalysis, fidelity, threading, and processivity in the exo- and endonucleolytic reactions of human exonuclease I. Proceedings of the National Academy of Sciences of the United States of America *114*, 6010-6015. 10.1073/pnas.1704845114.

171. Singh, P., Yang, M., Dai, H., Yu, D., Huang, Q., Tan, W., Kernstine, K.H., Lin, D., and Shen, B. (2008). Overexpression and Hypomethylation of Flap Endonuclease 1 Gene in Breast and Other Cancers. Molecular Cancer Research *6*, 1710-1717. 10.1158/1541-7786.mcr-08-0269.

172. Smietana, K., Siatkowski, M., and Moller, M. (2016). Trends in clinical success rates. Nature Reviews Drug Discovery *15*, 379-380. 10.1038/nrd.2016.85.

173. Sommers, C.H., Miller, E.J., Dujon, B., Prakash, S., and Prakash, L. (1995). Conditional lethality of null mutations in RTH1 that encodes the yeast counterpart of a mammalian 5′-exonuclease to 3′-exonuclease required for lagging-strand DNA-synthesis in reconstituted systems. Journal of Biological Chemistry *270*, 4193-4196. 10.1074/jbc.270.9.4193.

174. Sooriyaarachchi, M., George, G.N., Pickering, I.J., Narendran, A., and Gailer, J. (2016). Tuning the metabolism of the anticancer drug cisplatin with chemoprotective agents to improve its safety and efficacy. Metallomics *8*, 1170-1176. 10.1039/c6mt00183a.

175. Spyrakis, F., and Cavasotto, C.N. (2015). Open challenges in structure-based virtual screening: Receptor modeling, target flexibility consideration and active site water molecules description. Archives of Biochemistry and Biophysics *583*, 105-119. 10.1016/j.abb.2015.08.002.

176. Stepanenko, A.A., and Dmitrenko, V.V. (2015). HEK293 in cell biology and cancer research: phenotype, karyotype, tumorigenicity, and stress-induced genome-phenotype evolution. Gene *569*, 182-190. 10.1016/j.gene.2015.05.065.

177. Stucki, M., Jonsson, Z.O., and Hubscher, U. (2001). In eukaryotic flap endonuclease 1, the C terminus is essential for substrate binding. Journal of Biological Chemistry *276*, 7843-7849. 10.1074/jbc.M008829200.

178. Studier, F.W. (2005). Protein production by auto-induction in high-density shaking cultures. Protein Expression and Purification *41*, 207-234. 10.1016/j.pep.2005.01.016.

179. Studier, F.W. (2014). Stable Expression Clones and Auto-Induction for Protein Production in E-coli. Structural Genomics: General Applications *1091*, 17-32. 10.1007/978-1-62703-691-7_2.

180. Studier, F.W., and Moffatt, B.A. (1986). Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. Journal of Molecular Biology *189*, 113-130. 10.1016/0022-2836(86)90385-2.

181. Sun, X.M., Wu, P., Zheng, L., Thrower, D., Partikian, A., Qiu, J.Z., and Shen, B.H. (2002). Suppression of Saccharomyces cerevisiae rad27 null mutant phenotypes by the 5 ' nuclease domain of Escherichia coli DNA polymerase I. Current Genetics *41*, 379-388. 10.1007/s00294-002-0323-x.

182. Svenson, J., Brandsdal, B.O., Stensen, W., and Svendsen, J.S. (2007). Albumin binding of short cationic antimicrobial micropeptides and its influence on the in vitro bactericidal effect. Journal of Medicinal Chemistry *50*, 3334-3339. 10.1021/jm0703542.

183. Syson, K., Tomlinson, C., Chapados, B.R., Sayers, J.R., Tainer, J.A., Williams, N.H., and Grasby, J.A. (2008). Three metal ions participate in the reaction catalyzed by T5 flap endonuclease. Journal of Biological Chemistry *283*, 28741-28746. 10.1074/jbc.M801264200.

184. Takebe, T., Imai, R., and Ono, S. (2018). The Current Status of Drug Discovery and Development as Originated in United States Academia: The Influence of Industrial and Academic Collaboration on Drug Discovery and Development. Cts-Clinical and Translational Science *11*, 597-606. 10.1111/cts.12577.

185. Teasley, D.C., Parajuli, S., Nguyen, M., Moore, H.R., Alspach, E., Lock, Y.J., Honaker, Y., Saharia, A., Piwnica-Worms, H., and Stewart, S.A. (2015). Flap Endonuclease 1 Limits Telomere Fragility on the Leading Strand. Journal of Biological Chemistry *290*, 15133-15145. 10.1074/jbc.M115.647388.

186. Thandapani, P., Couturier, A.M., Yu, Z.B., Li, X., Couture, J.F., Li, S., Masson, J.Y., and Richard, S. (2017). Lysine methylation of FEN1 by SET7 is essential for its cellular response to replicative stress. Oncotarget *8*, 64918-64931. 10.18632/oncotarget.18070.

187. Theuretzbacher, U., Outterson, K., Engel, A., and Karlen, A. (2020). The global preclinical antibacterial pipeline. Nature Reviews Microbiology *18*, 275-285. 10.1038/s41579-019-0288-0.

188. Tian, S., Sun, H.Y., Pan, P.C., Li, D., Zhen, X.C., Li, Y.Y., and Hou, T.J. (2014). Assessing an Ensemble Docking-Based Virtual Screening Strategy for Kinase Targets by Considering Protein Flexibility. Journal of Chemical Information and Modeling *54*, 2664-2679. 10.1021/ci500414b.

189. Tishkoff, D.X., Boerger, A.L., Bertrand, P., Filosi, N., Gaida, G.M., Kane, M.F., and Kolodner, R.D. (1997). Identification and characterization of Saccharomyces cerevisiae

EXO1, a gene encoding an exonuclease that interacts with MSH2. Proceedings of the National Academy of Sciences of the United States of America *94*, 7487-7492. 10.1073/pnas.94.14.7487.

190. Topatana, W., Juengpanich, S., Li, S., Cao, J., Hu, J., Lee, J., Suliyanto, K., Ma, D., Zhang, B., Chen, M., and Cai, X. (2020). Advances in synthetic lethality for cancer therapy: cellular mechanism and clinical translation. Journal of Hematology & Oncology *13*, 118. 10.1186/s13045-020-00956-5.

191. Trott, O., and Olson, A.J. (2010). Software News and Update AutoDock Vina: Improving the Speed and Accuracy of Docking with a New Scoring Function, Efficient Optimization, and Multithreading. Journal of Computational Chemistry *31*, 455-461. 10.1002/jcc.21334.

192. 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.H., *et al*. (2011). Human Flap Endonuclease Structures, DNA Double-Base Flipping, and a Unified Understanding of the FEN1 Superfamily. Cell *145*, 198-211. 10.1016/j.cell.2011.03.004.

193. Tsutakawa, S.E., Sarker, A.H., Ng, C., Arvai, A.S., Shin, D.S., Shih, B., Jiang, S., Thwin, A.C., Tsai, M.-S., Willcox, A., *et al*. (2020). Human XPG nuclease structure, assembly, and activities with insights for neurodegeneration and cancer from pathogenic mutations. Proceedings of the National Academy of Sciences of the United States of America *117*, 14127-14138. 10.1073/pnas.1921311117.

194. Tsutakawa, S.E., Thompson, M.J., Arvai, A.S., Neil, A.J., Shaw, S.J., Algasaier, S.I., Kim, J.C., Finger, D., Jardine, E., Gotham, V.J.B., *et al*. (2017). Phosphate steering by Flap Endonuclease 1 promotes 5 '-flap specificity and incision to prevent genome instability. Nature Communications *8*, 16145. 10.1038/ncomms16145.

195. Tumey, L.N., Bom, D., Huck, B., Gleason, E., Wang, J., Silver, D., Brunden, K., Boozer, S., Rundlett, S., Sherf, B., *et al*. (2005). The identification and optimization of a N-hydroxy urea series of flap endonuclease 1 inhibitors. Bioorganic & Medicinal Chemistry Letters *15*, 277-281. http://dx.doi.org/10.1016/j.bmcl.2004.10.086.

196. Tumey, L.N., Huck, B., Gleason, E., Wang, J., Silver, D., Brunden, K., Boozer, S., Rundlett, S., Sherf, B., Murphy, S., *et al*. (2004). The identification and optimization of 2,4-diketobutyric acids as flap endonuclease 1 inhibitors. Bioorganic & Medicinal Chemistry Letters *14*, 4915-4918. http://dx.doi.org/10.1016/j.bmcl.2004.07.028.

197. Vallur, A.C., and Maizels, N. (2010). Distinct Activities of Exonuclease 1 and Flap Endonuclease 1 at Telomeric G4 DNA. Plos One *5*, e8908. 10.1371/journal.pone.0008908.

198. van Pel, D.M., Barrett, I.J., Shimizu, Y., Sajesh, B.V., Guppy, B.J., Pfeifer, T., McManus, K.J., and Hieter, P. (2013). An Evolutionarily Conserved Synthetic Lethal Interaction Network Identifies FEN1 as a Broad-Spectrum Target for Anticancer Therapeutic Development. Plos Genetics *9*, e1003254. 10.1371/journal.pgen.1003254.

199. Vandooren, J., Geurts, N., Martens, E., Van den Steen, P.E., and Opdenakker, G. (2013). Zymography methods for visualizing hydrolytic enzymes. Nature Methods *10*, 211-220. 10.1038/nmeth.2371.

200. Veber, D.F., Johnson, S.R., Cheng, H.Y., Smith, B.R., Ward, K.W., and Kopple, K.D. (2002). Molecular properties that influence the oral bioavailability of drug candidates. Journal of Medicinal Chemistry *45*, 2615-2623. 10.1021/jm020017n.

201. Vieira, T.F., and Sousa, S.F. (2019). Comparing AutoDock and Vina in Ligand/Decoy Discrimination for Virtual Screening. Applied Sciences-Basel *9*, 4538. 10.3390/app9214538.

202. Vipond, I.B., and Halford, S.E. (1995). Specific DNA recognition by *Eco*RV restriction endonuclease induced by calcium ions. Biochemistry *34*, 1113-1119. 10.1021/bi00004a002.

203. Virtanen, J.A., Cheng, K.H., and Somerharju, P. (1998). Phospholipid composition of the mammalian red cell membrane can be rationalized by a superlattice model. Proceedings of the National Academy of Sciences of the United States of America *95*, 4964-4969. 10.1073/pnas.95.9.4964.

204. Wadhwa, P., Jain, P., and Jadhav, H.R. (2015). QSAR and Docking Studies of N-Hydroxy Urea Derivatives as Flap Endonuclease-1 Inhibitors. Current Computer-Aided Drug Design *11*, 346-352. 10.2174/1573409912666151124233628.

205. Wang, G.M., and Zhu, W.L. (2016). Molecular docking for drug discovery and development: a widely used approach but far from perfect. Future Medicinal Chemistry *8*, 1707-1710. 10.4155/fmc-2016-0143.

206. Wang, Z., Sun, H.Y., Yao, X.J., Li, D., Xu, L., Li, Y.Y., Tian, S., and Hou, T.J. (2016). Comprehensive evaluation of ten docking programs on a diverse set of protein-ligand complexes: the prediction accuracy of sampling power and scoring power. Physical Chemistry Chemical Physics *18*, 12964-12975. 10.1039/c6cp01555g.

207. Wani, T.A., Bakheit, A.H., Abounassif, M.A., and Zargar, S. (2018). Study of Interactions of an Anticancer Drug Neratinib With Bovine Serum Albumin: Spectroscopic and Molecular Docking Approach. Frontiers in Chemistry *6*, 47. 10.3389/fchem.2018.00047.

208. Ward, T.A., McHugh, P.J., and Durant, S.T. (2017). Small molecule inhibitors uncover synthetic genetic interactions of human flap endonuclease 1 (FEN1) with DNA damage response genes. Plos One *12*, e0179278. 10.1371/journal.pone.0179278.

209. Waring, M.J., Arrowsmith, J., Leach, A.R., Leeson, P.D., Mandrell, S., Owen, R.M., Pairaudeau, G., Pennie, W.D., Pickett, S.D., Wang, J.B., *et al*. (2015). An analysis of the attrition of drug candidates from four major pharmaceutical companies. Nature Reviews Drug Discovery *14*, 475-486. 10.1038/nrd4609.

210. Waterman, D.G., Winter, G., Gildea, R.J., Parkhurst, J.M., Brewster, A.S., Sauter, N.K., and Evans, G. (2016). Diffraction-geometry refinement in the DIALS framework. Acta Crystallographica Section D-Structural Biology *72*, 558-575. 10.1107/s2059798316002187.

211. Wei, Q., Shen, J., Wang, D., Han, X., Shi, J., Zhao, L., and Teng, Y. (2021). A bibliometric analysis of researches on flap endonuclease 1 from 2005 to 2019. Bmc Cancer *21*, 374. 10.1186/s12885-021-08101-2.

212. Whitaker, A.M. and Freudenthal B.D. (2018). APE1: A skilled nucleic acid surgeon. DNA repair *71*, 93-100. 10.1016/j.dnarep.2018.08.012.

213. 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., *et al*. (2018). MolProbity: More and better reference data for improved all-atom structure validation. Protein Science *27*, 293-315. 10.1002/pro.3330.

214. Winter, G. (2010). xia2: an expert system for macromolecular crystallography data reduction. Journal of Applied Crystallography *43*, 186-190. 10.1107/s0021889809045701.

215. Wlodawer, A., Minor, W., Dauter, Z., and Jaskolski, M. (2008). Protein crystallography for non-crystallographers, or how to get the best (but not more) from published macromolecular structures. Febs Journal *275*, 1-21. 10.1111/j.1742-4658.2007.06178.x.

216. Xie, C., Wang, K., and Chen, D. (2016). Flap endonuclease 1 silencing is associated with increasing the cisplatin sensitivity of SGC-7901 gastric cancer cells. Molecular Medicine Reports *13*, 386-392. 10.3892/mmr.2015.4567.

217. Xu, H., Chen, X.Y., Xu, X.L., Shi, R.Y., Suo, S.S., Cheng, K.Y., Zheng, Z.G., Wang, M.X., Wang, L.Y., Zhao, Y., *et al*. (2016). Lysine Acetylation and Succinylation in HeLa Cells and their Essential Roles in Response to UV-induced Stress. Scientific Reports *6*, 30212. 10.1038/srep30212.

218. Xu, H., Shi, R.Y., Han, W.C., Cheng, J.H., Xu, X.L., Cheng, K.Y., Wang, L.Y., Tian, B., Zheng, L., Shen, B.H., *et al*. (2018). Structural basis of 5 ' flap recognition and protein-protein interactions of human flap endonuclease 1. Nucleic Acids Research *46*, 11315-11325. 10.1093/nar/gky911.

219. Xu, Y., Potapova, O., Leschziner, A.E., Grindley, N.D.F., and Joyce, C.M. (2001). Contacts between the 5 ' nuclease of DNA polymerase I and its DNA substrate. Journal of Biological Chemistry *276*, 30167-30177. 10.1074/jbc.M100985200.

220. Yang, W., Lee, J.Y., and Nowotny, M. (2006). Making and breaking nucleic acids: Two-Mg2+-ion catalysis and substrate specificity. Molecular Cell *22*, 5-13. 10.1016/j.molcel.2006.03.013.

221. Yap, T.A., Plummer, R., Azad, N.S., and Helleday, T. (2019). The DNA Damaging Revolution: PARP Inhibitors and Beyond. American Society of Clinical Oncology educational book. American Society of Clinical Oncology. Annual Meeting *39*, 185-195. 10.1200/edbk_238473.

222. Yuriev, E., Holien, J., and Ramsland, P.A. (2015). Improvements, trends, and new ideas in molecular docking: 2012-2013 in review. Journal of Molecular Recognition *28*, 581-604. 10.1002/jmr.2471.

223. Zeng, X., Che, X., Liu, Y.-P., Qu, X.-J., Xu, L., Zhao, C.-Y., Zheng, C.-L., Hou, K.-Z., and Teng, Y. (2017). FEN1 knockdown improves trastuzumab sensitivity in human epidermal growth factor 2-positive breast cancer cells. Experimental and Therapeutic Medicine *14*, 3265-3272. 10.3892/etm.2017.4873.

224. Zhang, F., Xue, J., Shao, J., and Jia, L. (2012). Compilation of 222 drugs' plasma protein binding data and guidance for study designs. Drug Discovery Today *17*, 475-485. 10.1016/j.drudis.2011.12.018.

225. Zhang, J.H., Chung, T.D.Y., and Oldenburg, K.R. (1999). A simple statistical parameter for use in evaluation and validation of high throughput screening assays. Journal of Biomolecular Screening *4*, 67-73. 10.1177/108705719900400206.

226. Zheng, L., Dai, H.F., Zhou, M., Li, M., Singh, P., Qiu, J.Z., Tsark, W., Huang, Q., Kernstine, K., Zhang, X.M., *et al*. (2007). Fen1 mutations result in autoimmunity, chronic inflammation and cancers. Nature Medicine *13*, 812-819. 10.1038/nm1599.

227. Zheng, L., Zhou, M., Chai, Q., Parrish, J., Xue, D., Patrick, S.M., Turchi, J.J., Yannone, S.M., Chen, D., and Shen, B.H. (2005). Novel function of the flap endonuclease 1 complex in processing stalled DNA replication forks. Embo Reports *6*, 83-89. 10.1038/sj.embor.7400313.

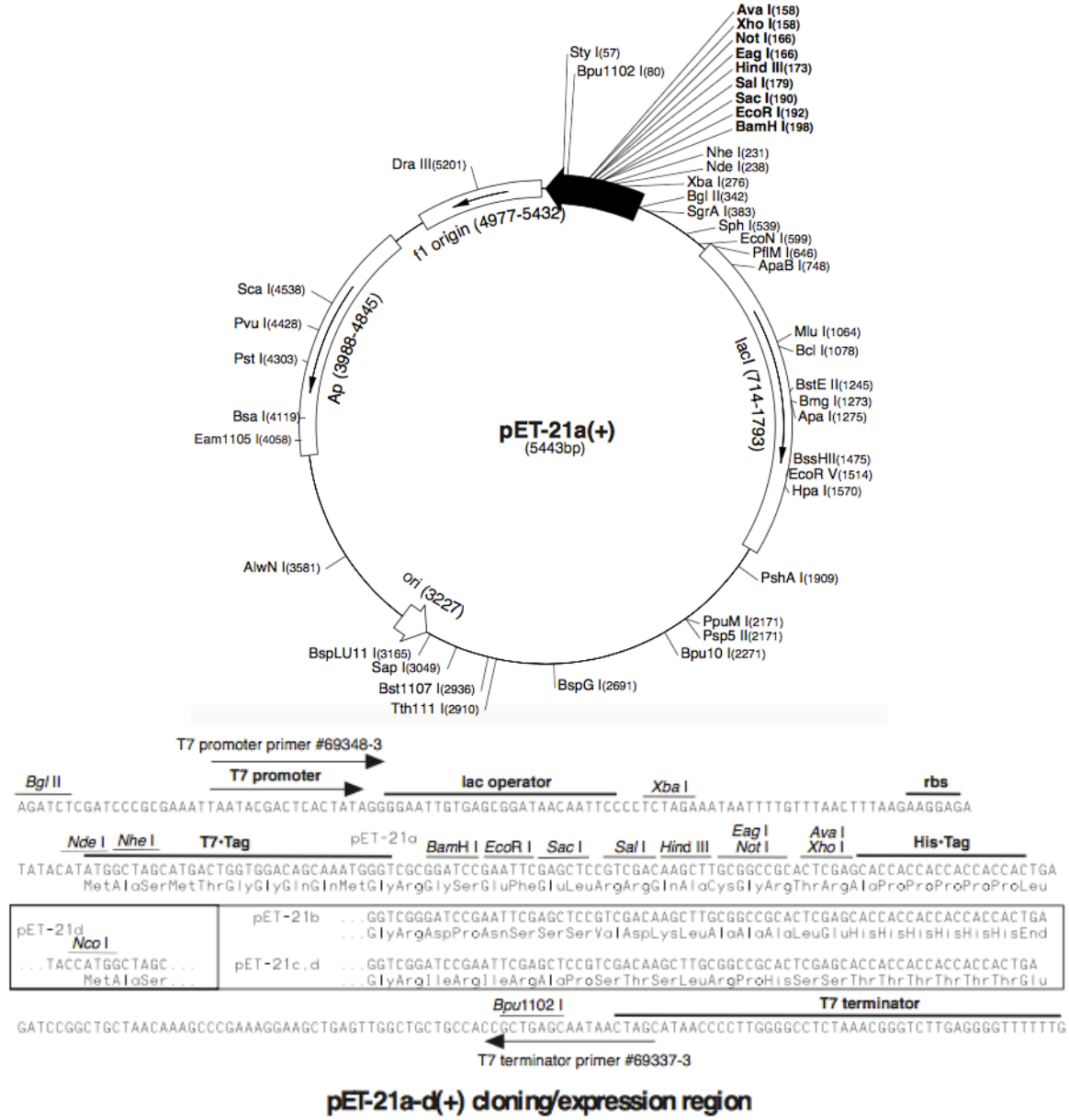
228. Zheng, L., Zhou, M.A., Guo, Z.G., Lu, H.M., Qian, L.M., Dai, H.F., Qiu, J.Z., Yakubovskaya, E., Bogenhagen, D.F., Demple, B., and Shen, B.H. (2008). Human DNA2 Is a Mitochondrial Nuclease/Helicase for Efficient Processing of DNA Replication and Repair Intermediates. Molecular Cell *32*, 325-336. 10.1016/j.molcel.2008.09.024.

229. Zhou, H.J., Di Palma, S., Preisinger, C., Peng, M., Polat, A.N., Heck, A.J.R., and Mohammed, S. (2013). Toward a Comprehensive Characterization of a Human Cancer Cell Phosphoproteome. Journal of Proteome Research *12*, 260-271. 10.1021/pr300630k.

230. Zwaal, R.F.A., and Vandeene.Ll (1968). Protein patterns of red cell membranes from different mammalian species. Biochimica Et Biophysica Acta *163*, 44-49. 10.1016/0005-2736(68)90031-x.

Supplementary Information

Supplementary Information 1 – pET-21a(+) vector map



pET-21a-d(+) cloning/expression region

Feature	Location
T7 terminator	26-72
His® Tag coding sequence	
Multiple cloning sites (BamHI - XhoI)	158-203
T7® Tag coding sequence	207-239
T7 transcription start site	310
T7 promoter	311-327
lacI coding sequence	714-1793
pBR322 origin of replication	3227
ampicillin resistance (also referred as Bla or Ap)	3988-4845
f1 origin of ssDNA replication	4977-5432

Figure SI.1. pET-21a(+) circular vector map and its polylinker sequence. Vector details, map and DNA sequences stated above were adapted from the online available Novagen pET System Manual.

<u>Supplementary Information 2</u> – Protein and inserted gene sequences

```
atgggtattcagggtctggcaaaactgattgcagatgttgcaccgagcgccattcgcgaa
 M  G  I  Q  G  L  A  K  L  I  A  D  V  A  P  S  A  I  R  E
aatgatattaaaagctattttggtcgcaaagttgcaattgatgccagcatgagcatttat
 N  D  I  K  S  Y  F  G  R  K  V  A  I  D  A  S  M  S  I  Y
cagtttctgattgcagttcgtcagggtggtgatgttctgcagaatgaagaaggtgaaacc
 Q  F  L  I  A  V  R  Q  G  G  D  V  L  Q  N  E  E  G  E  T
acctctcatctgatgggtatgttttatcgcaccattcgcatgatggaaaatggcattaaa
 T  S  H  L  M  G  M  F  Y  R  T  I  R  M  M  E  N  G  I  K
ccggtgtatgtttttgatggtaaacctccgcagctgaaaagcggtgaactggcaaaacgt
 P  V  Y  V  F  D  G  K  P  P  Q  L  K  S  G  E  L  A  K  R
agcgaacgtcgtgcagaagcagaaaaacagctgcagcaggcacaggcagccggtgcagaa
 S  E  R  R  A  E  A  E  K  Q  L  Q  Q  A  Q  A  A  G  A  E
caggaagttgaaaaattcaccaaacgcctggtgaaagtgaccaaacagcataatgatgaa
 Q  E  V  E  K  F  T  K  R  L  V  K  V  T  K  Q  H  N  D  E
tgcaaacatctgctgagcctgatgggtattccgtatctggatgcaccgagcgaagccgaa
 C  K  H  L  L  S  L  M  G  I  P  Y  L  D  A  P  S  E  A  E
gcaagctgtgcagcactggttaaagcaggtaaagtttatgcagcagccaccgaagatatg
 A  S  C  A  A  L  V  K  A  G  K  V  Y  A  A  A  T  E  D  M
gattgtctgacctttggttctccggttctgatgcgtcatctgaccgcaagcgaagcaaaa
 D  C  L  T  F  G  S  P  V  L  M  R  H  L  T  A  S  E  A  K
aaactgccgattcaggaatttcatctgagccgtattctgcaggaactgggtctgaatcag
 K  L  P  I  Q  E  F  H  L  S  R  I  L  Q  E  L  G  L  N  Q
gaacagtttgttgatctgtgcattctgctgggtagcgattattgtgaaagcattcgtggt
 E  Q  F  V  D  L  C  I  L  L  G  S  D  Y  C  E  S  I  R  G
attggtccgaaacgtgccgttgatctgattcagaaacataaaagcattgaagaaattgtg
 I  G  P  K  R  A  V  D  L  I  Q  K  H  K  S  I  E  E  I  V
cgtcgtctggatccgaataaatatccggtgccggaaaattggctgcataaagaagcacat
 R  R  L  D  P  N  K  Y  P  V  P  E  N  W  L  H  K  E  A  H
cagctgtttctggaaccggaagttctggacccggaaagcgttgaactgaaatggtctgaa
 Q  L  F  L  E  P  E  V  L  D  P  E  S  V  E  L  K  W  S  E
ccgaatgaagaagaactgattaaatttatgtgcggtgaaaaacagtttagcgaagaacgt
 P  N  E  E  E  L  I  K  F  M  C  G  E  K  Q  F  S  E  E  R
attcgtagcggtgttaaacgtctgagcaaaagccgtcagggtagcacccagggccgcctg
 I  R  S  G  V  K  R  L  S  K  S  R  Q  G  S  T  Q  G  R  L
gatgatttcttcaaggtgaccggctcactgtcttcagctaagcgcaaggagccggaaccg
 D  D  F  F  K  V  T  G  S  L  S  S  A  K  R  K  E  P  E  P
aagggctccactaagaagaaggcaaagactggggcagcagggaagtttaaacgcggaaaa
 K  G  S  T  K  K  K  A  K  T  G  A  A  G  K  F  K  R  G  K
taatgaagctt
 -  -  S
```

Figure SI.2. Protein and inserted gene sequences of wild-type full-length HsFEN1. Full-length wild-type HsFEN1 was initially cloned in Jon Sayer's laboratory by a previous PhD student, Sarah Louise Oates. N-terminal methionine (Met, M) is shown in red, whereas the 44 amino acid long C-terminal is shown in yellow. Both protein and gene sequences confirmation was carried out in Mass Spectrometry Facility and Core Genomic Facility at the University of Sheffield. DNA sequence was translated using the ExPASy translate tool.

```
atgggtattcagggtctggcaaaactgattgcagatgttgcaccgagcgccattcgcgaa
 M  G  I  Q  G  L  A  K  L  I  A  D  V  A  P  S  A  I  R  E
aatgatattaaaagctattttggtcgcaaagttgcaattgatgccagcatgagcatttat
 N  D  I  K  S  Y  F  G  R  K  V  A  I  D  A  S  M  S  I  Y
cagtttctgattgcagttcgtcagggtggtgatgttctgcagaatgaagaaggtgaaacc
 Q  F  L  I  A  V  R  Q  G  G  D  V  L  Q  N  E  E  G  E  T
acctctcatctgatgggtatgttttatcgcaccattcgcatgatggaaaatggcattaaa
 T  S  H  L  M  G  M  F  Y  R  T  I  R  M  M  E  N  G  I  K
ccggtgtatgtttttgatggtaaacctccgcagctgaaaagcggtgaactggcaaaacgt
 P  V  Y  V  F  D  G  K  P  P  Q  L  K  S  G  E  L  A  K  R
agcgaacgtcgtgcagaagcagaaaaacagctgcagcaggcacaggcagccggtgcagaa
 S  E  R  R  A  E  A  E  K  Q  L  Q  Q  A  Q  A  A  G  A  E
caggaagttgaaaaattcaccaaacgcctggtgaaagtgaccaaacagcataatgatgaa
 Q  E  V  E  K  F  T  K  R  L  V  K  V  T  K  Q  H  N  D  E
tgcaaacatctgctgagcctgatgggtattccgtatctggatgcaccgagcgaagccgaa
 C  K  H  L  L  S  L  M  G  I  P  Y  L  D  A  P  S  E  A  E
gcaagctgtgcagcactggttaaagcaggtaaagtttatgcagcagccaccgaaaaaatg
 A  S  C  A  A  L  V  K  A  G  K  V  Y  A  A  A  T  E  K  M
gattgtctgacctttggttctccggttctgatgcgtcatctgaccgcaagcgaagcaaaa
 D  C  L  T  F  G  S  P  V  L  M  R  H  L  T  A  S  E  A  K
aaactgccgattcaggaatttcatctgagccgtattctgcaggaactgggtctgaatcag
 K  L  P  I  Q  E  F  H  L  S  R  I  L  Q  E  L  G  L  N  Q
gaacagtttgttgatctgtgcattctgctgggtagcgattattgtgaaagcattcgtggt
 E  Q  F  V  D  L  C  I  L  L  G  S  D  Y  C  E  S  I  R  G
attggtccgaaacgtgccgttgatctgattcagaaacataaaagcattgaagaaattgtg
 I  G  P  K  R  A  V  D  L  I  Q  K  H  K  S  I  E  E  I  V
cgtcgtctggatccgaataaatatccggtgccggaaaattggctgcataaagaagcacat
 R  R  L  D  P  N  K  Y  P  V  P  E  N  W  L  H  K  E  A  H
cagctgtttctggaaccggaagttctggatccggaaagcgttgaactgaaatggtctgaa
 Q  L  F  L  E  P  E  V  L  D  P  E  S  V  E  L  K  W  S  E
ccgaatgaagaagaactgattaaatttatgtgcggtgaaaaacagtttagcgaagaacgt
 P  N  E  E  E  L  I  K  F  M  C  G  E  K  Q  F  S  E  E  R
attcgtagcggtgttaaacgtctgagcaaaagccgtcagggtagcacctaa
 I  R  S  G  V  K  R  L  S  K  S  R  Q  G  S  T  -
```

Figure SI.2.1. Protein and inserted gene sequences of truncated D179K HsFEN1. Truncated D179K HsFEN1 was initially cloned in Jon Sayer's laboratory by a previous PhD student, Jing Zhang. This protein is truncated at residue Gln337 (Glutamine, Q) and lacks the flexible C-terminal (44 amino acids long) domain, also known as PCNA binding domain. N-terminal methionine is shown in red, whereas the D179K mutation is shown in magenta. Both protein and gene sequences confirmation was carried out in Mass Spectrometry Facility and Core Genomic Facility at the University of Sheffield. DNA sequence was translated using the ExPASy translate tool.

<u>Supplementary Information 3</u> – Physicochemical properties of full-length WT and truncated D179K HsFEN1 proteins

Table SI.3. Physical and chemical parameters[1] of wild-type full-length HsFEN1 and truncated D179K HsFEN1.

Proteins	Length	Theoretical pI	MW (Da)	Ext. coefficient ($m^{-1}cm^{-1}$)	Abs 0.1% (=1g/L)[2]	A_{280} absorbers	
						Tyrosines (Y)	Tryptophans (W)
Wild-type full-length HsFEN1	380	8.80	42592.98	22920	0.538	8	2
Short-form D179K HsFEN1	336	6.49	37871.58	22920	0.605	8	2

[1]All parameters were analysed using the ExPASy protein parameters (ProtParam) tool, based on the amino acid sequences (Gasteiger *et al*., 2005).

[2]Abs 0.1% was calculated assuming that formation of disulphide bonds between the thiol groups of cysteine residues formed are reduced.

<u>Supplementary Information 4</u> – Assays layout for MaybridgeRO3 fragment library screening

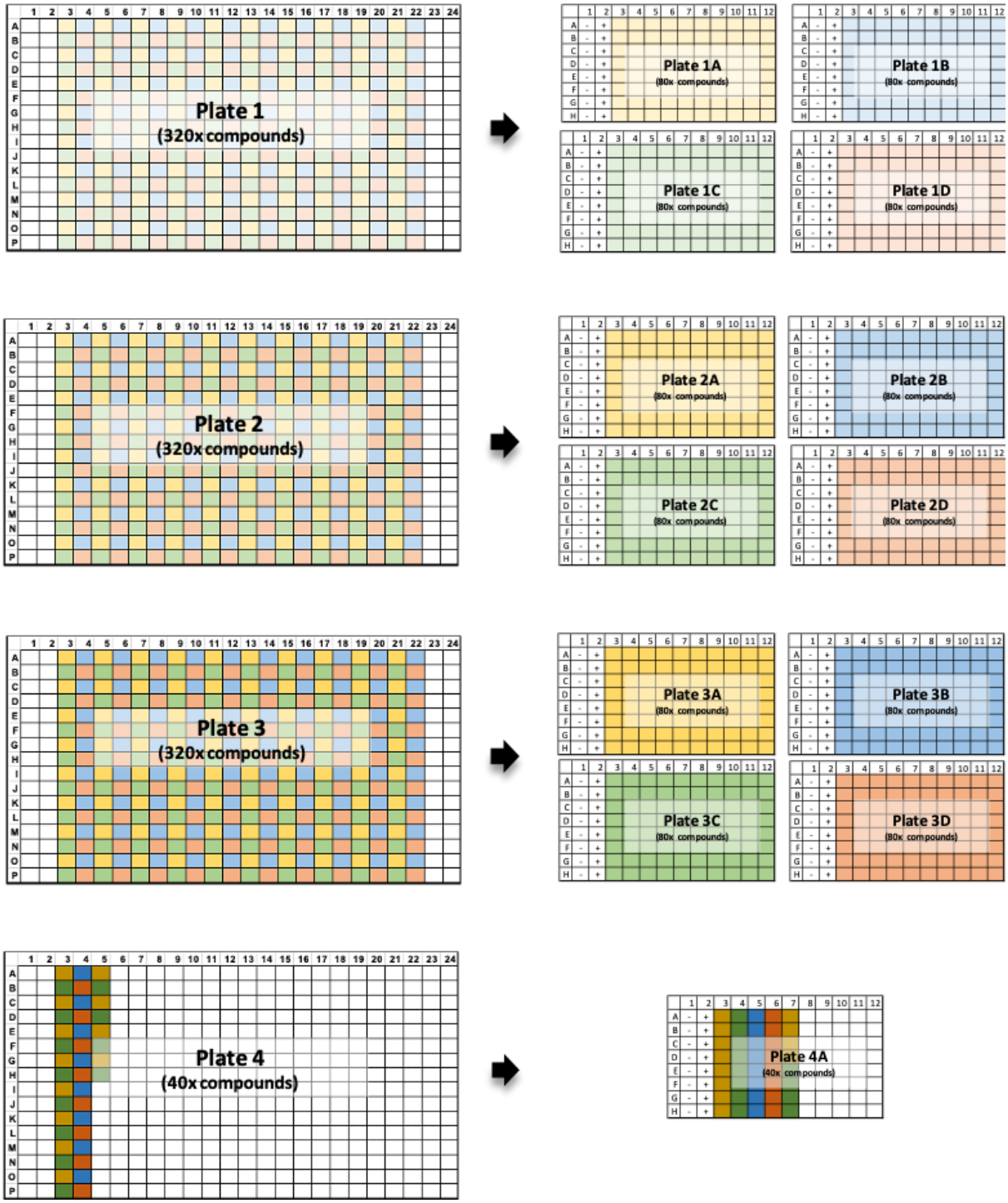


Figure S1.4. Assays layout for HTS of 1,000 ligands from MaybridgeRO3 fragment library. The proposed layout is representative of one complete experiment, providing a single FRET measurement for each of the 1,000 compounds. The 1,000 fragments, stored in four 384-well plates, were screened in a total of 13 x 96-well plate assays over one complete run. Minus (-) and plus (+) on the 13 assay plates are indicative of the negative and positive controls, respectively.

<u>Supplementary Information 5</u> – Collected diffraction data of full-length WT HsFEN1:5ov4ex poor resolution complex

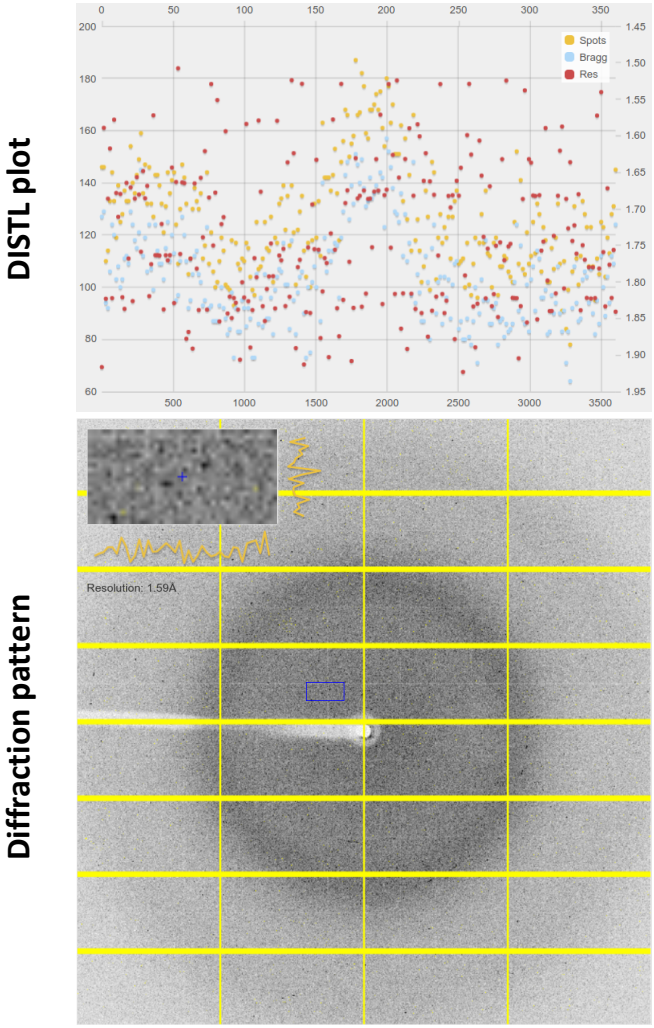


Figure SI.5. DISTL plots and diffraction pattern typology of full-length WT HsFEN1:5ov4ex crystals diffracted at 5.3 Å resolution. Full dataset was collected using iØ4 beamline at Diamond light source, UK (Resolution: 2 Å; Exposure: 0.015 s; Beamsize: 80x20 µm; Ω Osc: 0.10˚; No. Images: 3600; Wavelength: 0.9763 Å; Transmission: 50.16%; Type: SAD). DISTL plots, indicative of the number of images collected and their respective resolutions throughout data collection, and diffraction pattern typologies of each HsFEN1:5ov4ex complex.

Table SI.5. Diffraction data statistics of full-length WT HsFEN1:5ov4ex crystals diffracted at 5.3 Å resolution.[1]

Full-length WT HsFEN1:5ov4ex complex	
Total observations	9855 (385)
Unique observations	1668 (93)
Resolution (Å)	5.28-37.26 (5.28-5.37)
$CC_{1/2}$[2]	0.2 (0.3)
Completeness (%)[3]	99.6 (98.6)
Multiplicity[4]	5.9 (5.4)
Space group	P 1 2 1
Mean I/sig(I)[5]	2.6 (0.8)
R_{meas}[6]	2.16 (2.093)
Unit cell a, b, c (Å), α, β, γ (˚)	61.05, 101.56, 105.79, 90, 106.71, 90
Cryoprotectant % (v/v)[7]	20% Glycerol

[1]Diffraction data statistics collected at 0.9119 Å (iØ3 beamline) and autoprocessed using XDS in xia2 DIALS pipeline (Kabsch, 2010b; Winter, 2010). AIMLESS was used for data scaling and merging (Evans and Murshudov, 2013; Waterman *et al*., 2016). Values in parenthesis are representative of the outer/highest resolution shell.

[2]$CC_{1/2}$ is the calculated Pearson correlation coefficient between two half-split sets of reflections.

[3]Completeness is indicative of the number of all theoretically possible unique observations measured in the collected dataset.

[4]Multiplicity is representative of the average number of observations per individual unique reflection.

[5]Mean I/sig(I) is representative of the average signal-to-noise ratio of the measured intensities in the collected dataset.

[6]R_{meas} is the multiplicity-corrected R value version (Karplus and Diederichs, 2012; Wlodawer *et al*., 2008).

[7]Cryoprotectant, refers to the compound used to prevent ice formation during flash-cooling and data collection.

[8]DISTL plots, diffraction pattern typologies and crystal morphologies are shown in Figure SI.5.

<u>Supplementary Information 6</u> – Mass spectrometry spectra for truncated D179K and full-length WT HsFEN1 proteins.

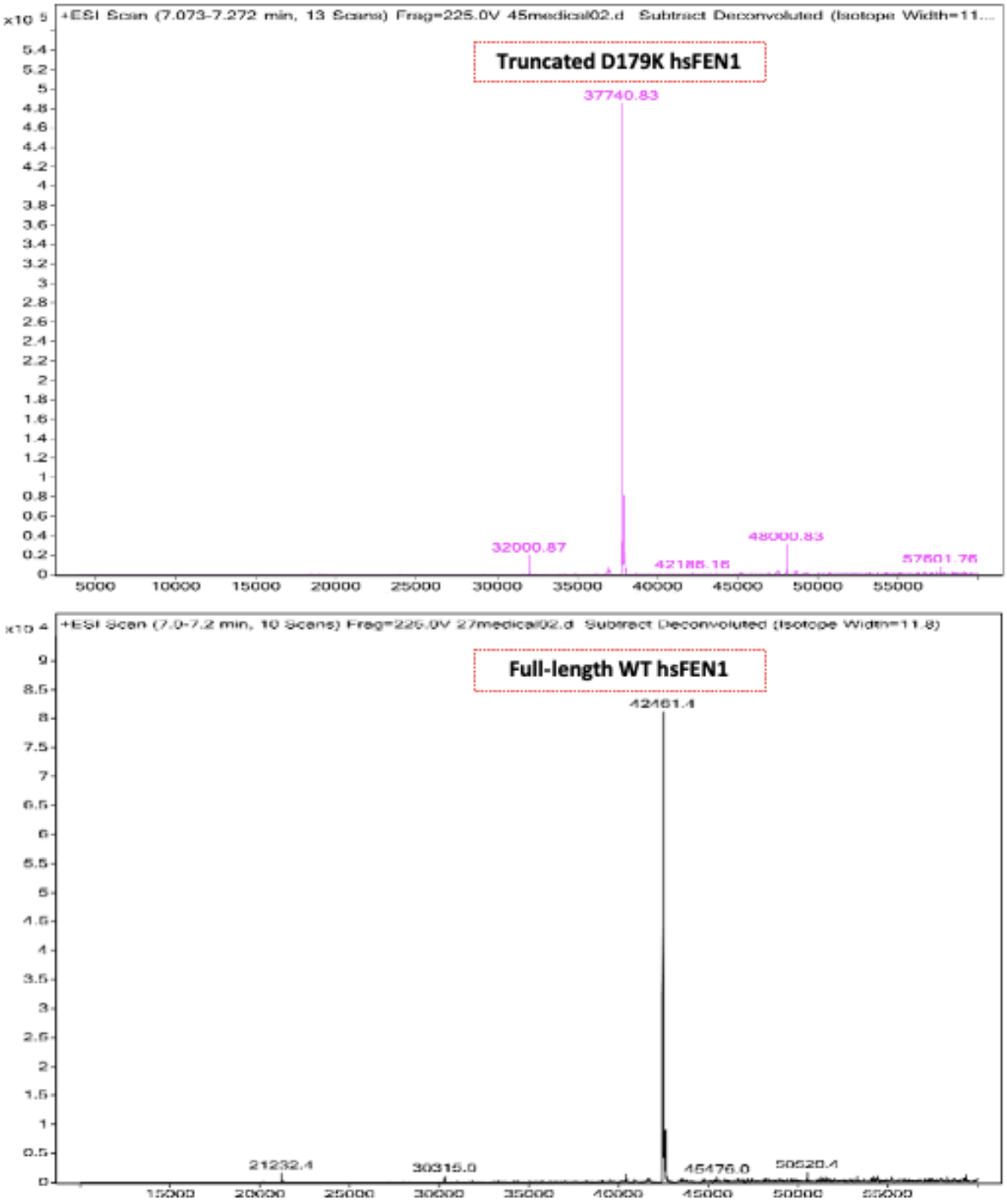


Figure SI.6. Full scan mass spectra of truncated D179K and full-length WT HsFEN1. Base peaks correspond to (A) truncated D179K HsFEN1 and (B) full-length WT HsFEN1. Each spectra shows the masses of detected in the sample molecules (X-axis) against the number of counts (Y-axis), analogous to the relative intensity and concentration. Mass spectrometry results were provided by the Facility of Mass Spectrometry (Department of Chemistry, University of Sheffield).

<u>Supplementary Information 7</u> – Haemolytic activity of ampicillin on 2-5% (v/v) RBCs suspensions and whole blood erythrocytes.

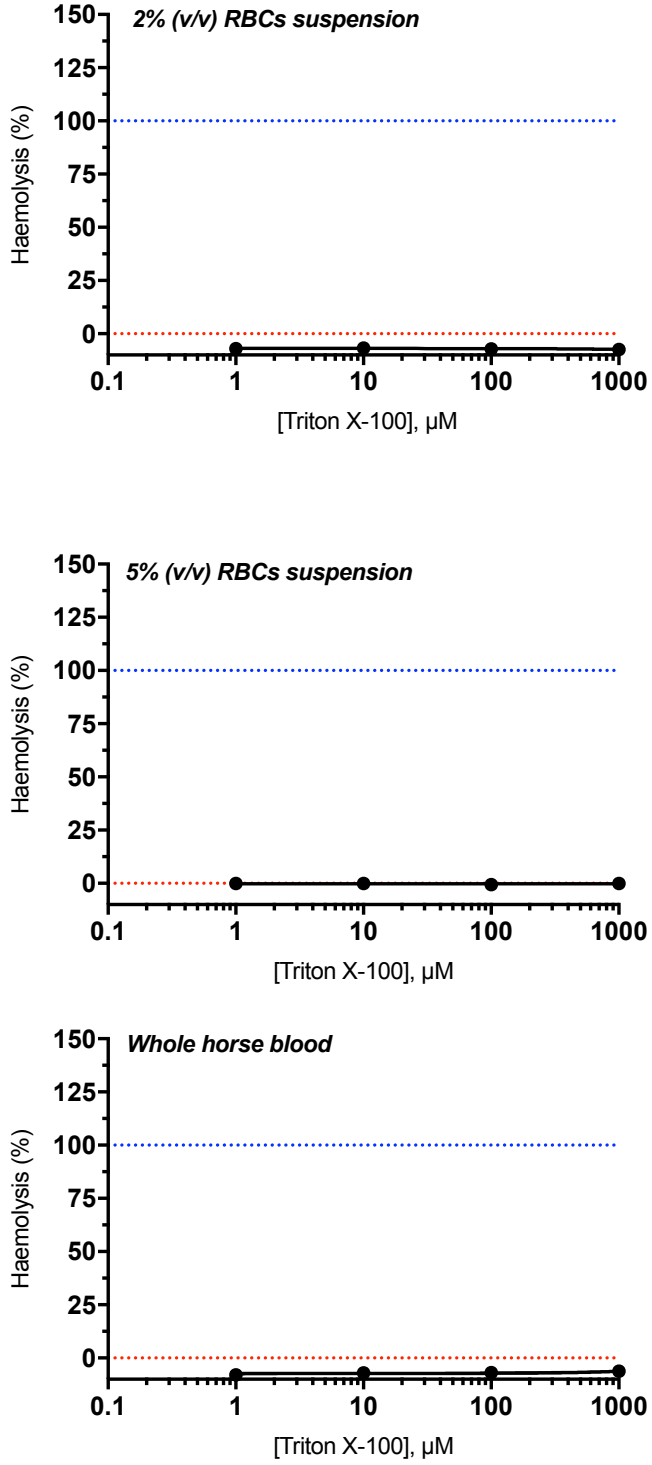


Figure SI.7. Haemolytic activity of ampicillin on 2-5% (v/v) RBCs suspensions and whole blood erythrocytes. Ampicillin haemolytic effect over whole horse blood erythrocytes, 2% and 5% (v/v) RBCs suspension, at 1-1000 µM concentrations. Each value represents the normalised mean of three technical repeats (n=8), based on negative/vehicle and positive/10% (v/v) Triton X-100 controls (n=8). Data were plotted using nonlinear regression [Inhibitor] vs. response -Variable slope (four parameters) model in GraphPad Prism version 8.2. $Z' \geq 0.5$ for all assays performed.

<u>Supplementary Information 8</u> – Electron density maps of WT HsFEN1 active site in complex with processed JT2+5 DNA pre-refinement and rebuilding.

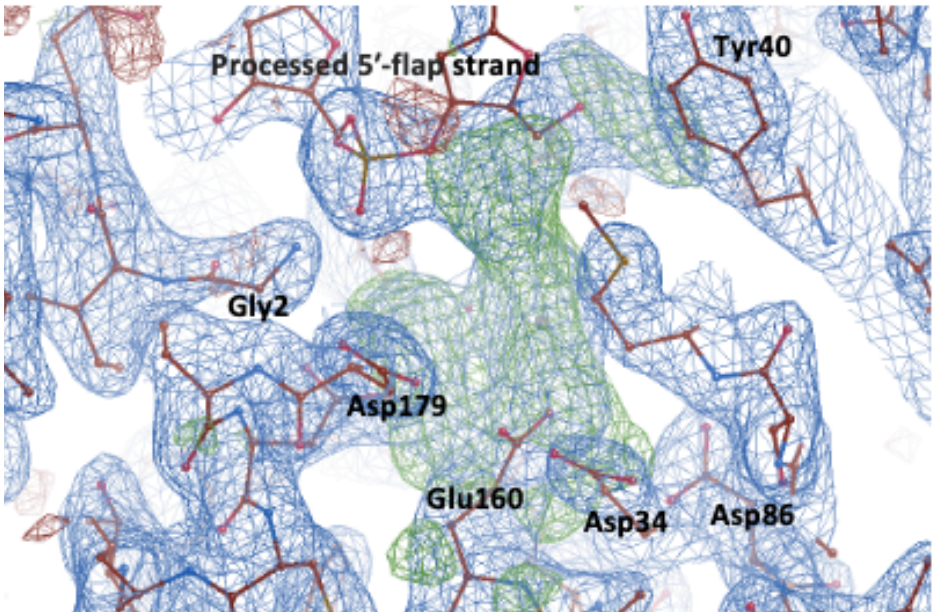


Figure SI.8. Electron density maps of WT HsFEN1 active site in complex with processed JT2+5 DNA pre-refinement and rebuilding. The model is representative of the overall fitting in the electron density map of WT HsFEN1 active site complexed with product JT2+5 substrate. Unmodelled bulb-like electron density (green, positive) corresponds to the active site metals and water molecules pre-refinement and rebuilding (1σ with about 3σ positive map). Tyr40, Gly2 and some of the surrounding carboxylates, composing the active site are also labelled.