Thesis submitted for the degree of Doctor of Philosophy to

the Department of Chemistry, University of Sheffield

Evidence of dynamics and disorder using NMR-spectroscopic techniques applied to human Flap-Endonuclease-1

Supervisors: Prof. Jane Grasby and Prof. Jonathan Waltho

Ian Alexander Bennet

September 2017

Declaration

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

Ian Alexander Bennet

Acknowledgements

It is said that something is always impossible until it is done. Well certainly none of this would have been possible without the great support and encouragement of a number of people. Primarily, I would not even be here if my supervisor Professor Jane Grasby had not provided me with all the help and guidance I received over these years. I am truly grateful for the opportunity to work with her. On top of that I was lucky to have a second supervisor, Professor Jon Waltho who always managed to come up with solutions to my insoluble NMR problems. Of course there is Dr. Dave Finger as well, the major force behind the NMR project, and a constant fount of knowledge, whether you wanted to know it or not! In all seriousness though, I truly enjoyed the endless conversations we had about science be it in the lab or at lunch.

I am very grateful to be part of the BBSRC White Rose doctoral training program, and the huge support the whole system provided.

A large thank you must go to Andrea Hounslow for helping me decipher the complexities of NMR machines and their software, and also for all those hours comparing chemical shifts on a screen. Thank you also to Nicky Baxter for her invaluable help with everything to do with relaxation measurements.

The office was never a dull place, with plenty of fantastic characters. It took some time, but I warmed up eventually to you all. At the start there was of course Jack and Emma, who helped me out of my shell and were the first to make me truly feel welcome. Then there was the Williams trio: Claire, Hannah and Ester, who were a constant source of entertainment. A true change of weather came in when Sana arrived, and we would talk about our love of food, and I would receive Arabic lessons (I didn't get very far). Then there was that soft-spoken sitcom superhero Stevo, who always told the worst jokes. Not to mention the overly dramatic antics of Nazihah and her partner in crime Becca, those two are a force to be reckoned with. These are just a few of the many great people all around the office, and indeed the whole university (you know who you are); I will miss the fun times we spent faffing about!

Finally, for all their support throughout the years, and especially during these 4 years, I would like to thank my parents, John and Debi, to whom I dedicate this thesis.

Abstract

Flap endonuclease 1 (FEN1) is a member of a 5' nuclease superfamily involved in DNA replication and repair. FEN1 hydrolyses the phosphodiester bond one nucleotide into the duplex region of bifurcated double-flapped structures as found in lagging-strand DNA synthesis. These flap structures need to be cut in a very specific manner on the order of around 10^6 times per cell cycle. Therefore FEN1 is seen as an essential enzyme that maintains genomic integrity across all life forms. How FEN1 achieves its molecular recognition for a chemically very similar but structurally different DNA substrate and how it achieves catalysis on a biochemically relevant timescale are key questions to understand the protein system. This thesis describes some of the mechanistic studies used to understand how the structure and dynamics of hFEN1 relate to its function.

It was proposed that T5 bacteriophage FEN was a catalytically perfect, or diffusion-limited enzyme, yet its main rate-limiting step after substrate binding was non-chemistry related. To ascertain whether this was true for hFEN1, the effect of leaving group pKa using 2' modified double flapped substrates on rates of catalysis was measured. It was found that both apparent second order rates and first order single turnover rates of catalysis were insensitive to leaving group pKa. Furthermore by supplementing the reaction with glycerol, an unexpectedly high viscosity dependence was observed. The explanation for this is likely the presence of another physical step in the catalytic cycle affected by viscosity.

Previous structural and biophysical studies of hFEN1 identified a helical arch, which was thought to be disordered as it could accommodate bulky 5' flaps through it. Furthermore, the arch is key for positioning the 5' flap into the active site. Using NMR spectroscopic techniques the solution state conformation of apo-hFEN1 was analysed. The arch was found to be disordered, but the C-terminal portion of it was transiently sampling α-helical φ,ψ space, while the other half was in an extended conformation. Another DNA recognition region, the α2-α3 loop was also found to be disordered.

Various ligands and substrates were found to alter the structure and the dynamics of hFEN1. Addition of substrate DNA slowed the motion of the arch and α2-α3 loop to a millisecond timescale. Equally addition of a single monophosphate nucleotide had an effect on the dynamics of the top of the arch, despite binding in the active site. Furthermore, titration of calcium ions into the active site when DNA was present on the enzyme resulted in large perturbations to substrate recognition sites distant from the active site. This potentially links the specificity of these regions to activity within the active site.

Publications arising from work covered in this thesis

Algasaier, Sana I., Exell, Jack C., Bennet, Ian A., Thompson, Mark J., Gotham, Victoria J.B., Shaw, Steven J., Craggs, Timothy D., Finger, L. David, Grasby, Jane 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

Contents

Abbreviations

AIC	Akaike information criteria
Arg / R	Arginine
Asp / D	Aspartic acid
B_0	Uniform external magnetic field strength
B_{eff}	Effective magnetic field strength
βME	Beta-mercaptoethanol
C(t)	Correlation function
CIDER	Classification of intrinsically disordered ensemble regions
CPMG	Carr-Purcell Meiboom-Gill pulse sequence
CSA	Chemical shift anisotropy
Cys / C	Cysteine
dAMP	Deoxyadenosine monophosphate
DC	Dipolar coupling
DHAP	Dihydroxyacetone phosphate
dHPLC	Denaturing high performance liquid chromatography
DNA	Deoxyribonucleic acid
DSBR	Double-strand break repair
dsDNA	Double Stranded DNA
DTT	Dithiothreitol
E	Enzyme
ECCD	Exciton-coupled circular dichroism
EDTA	Ethylenediaminetetraacetic acid
EQ	Enzyme-product complex
ES	Enzyme-substrate complex
EXO1	Exonuclease-1
FAM	Fluorescein
FEN(1)	Flap Endonuclease
FRET	Förster resonance energy transfer
GAP	D-glyceraldehyde 3-phosphate
GEN1	Gap Endonuclease-1
Glu / E	Glutamic acid
H2tH	Helix two-turn-helix motif

hFEN1	Human Flap Endonuclease-1
His / H	Histidine
HJ	Holliday Junction
hNOE	Heteronuclear nuclear Overahauser effect
HSQC	Heteronuclear single quantum coherence
Ile / I	Isoleucine
IPTG	Isopropyl-β-D-thiogalatopyranoside
J(ω)	Spectral density function
k_{cat}	Catalytic rate constant
k_{cat}/K_M	Apparent second-order rate constant
$k_{CC/RCC}$	Rate of conformational change/reverse conformational change
k_{chem}	Rate of chemistry
K_D	Dissociation constant
K_M	Michaelis constant
$k_{on/off}$	Bimolecular diffusion association/dissociation rate
$k_{release}$	Rate of product release
k_{ST}	Single turnover rate constant
Leu / L	Leucine
LFER	Linear free energy relationship
Lys / K	Lysine
mRNA	Messenger RNA
NSR	Nuclear spin relaxation
P	Product
PCNA	Proliferating Cell Nuclear Antigen
PIP box	PCNA interacting peptide
Pol-α Primase	Polymerase-α Primase hetero-tetramer complex
Pol-δ	Polymerase-δ
Pol-ε	Polymerase- ε
Pro / P	Proline
R_1	Spin-lattice (longitudinal) relaxation rate
$R_{1\rho}$	Rotating frame relaxation rate
R_2	Spin-spin (transverse) relaxation rate
RD	Relaxation dispersion
R_{ex}	Chemical exchange rate

RNA	Ribonucleic acid
RNase A	Pancreatic ribonuclease
S:N	Signal to noise ratio
S^2	Model free order parameter
S_N2	Associative Nucleophilic Substitution
ssDNA	Single Stranded DNA
T5 FEN	T5 bacteriophage Flap Endonuclease
TCEP	Tris(2-carboxyethyl)phosphine
Thr / T	Threonine
TIM	Triosephosphate isomerase
TROSY	Transverse relaxation optimised spectroscopy
Tyr / Y	Tyrosine
Val / V	Valine
W_0	Probability of zero-quantum transition
W_2	Probability of double-quantum transition
WT	Wild type
*X*nt	*X* Nucleotide(s), where *X* is an integer
XPG	Xeroderma Pigmentosum-G
XRN1	Exoribonuclease1
βlg	Brønsted proportionality constant
Δδ	Chemical shift difference
η	Viscosity
τ_C	Correlation time
τ_e	Effective correlation time
Ω-loop	Omega loop

Chapter 1: Introduction

1.1 Nucleases

1.1a Background

It has been shown for many years now that deoxyribonucleic acid (DNA) is the molecule that codes for genetic information.[1,2] However, the DNA molecule is by no means a static structure. DNA is like a book, which is fragile in places and needs constant copying and repair. Hydrolysis, oxidation and methylation of DNA occur endogenously, at high frequency, *in vivo*, which require constant repair by enzymatic pathways.[3] Maintenance of the genome includes nucleobase oxidation removal, nucleobase methylation removal and resolving mismatches of base pairs. Furthermore, DNA needs to be replicated in order to propagate genetic information, requiring an unwinding of the double-helix.[4] Therefore, the importance of maintaining high fidelity during DNA replication and repair is paramount.

Nucleic acids achieve these chemical processes with the aid of a plethora of highly efficient and specific enzymes. There are many pathways that are both monitored and catalysed by nucleic acid-acting enzymes. A few examples of enzymes that work on nucleic acids include polymerases, glycosylases and nucleases. The latter, which are the focus of this thesis, catalyse the hydrolysis of a phosphodiester bond that links two nucleotides together. The diversity of nucleases is striking. They are involved in a wide variety of very important cellular DNA and RNA tasks, such as replication, repair, structural changes, mRNA splicing, RNA degradation, cell defense and apoptosis to name a few.

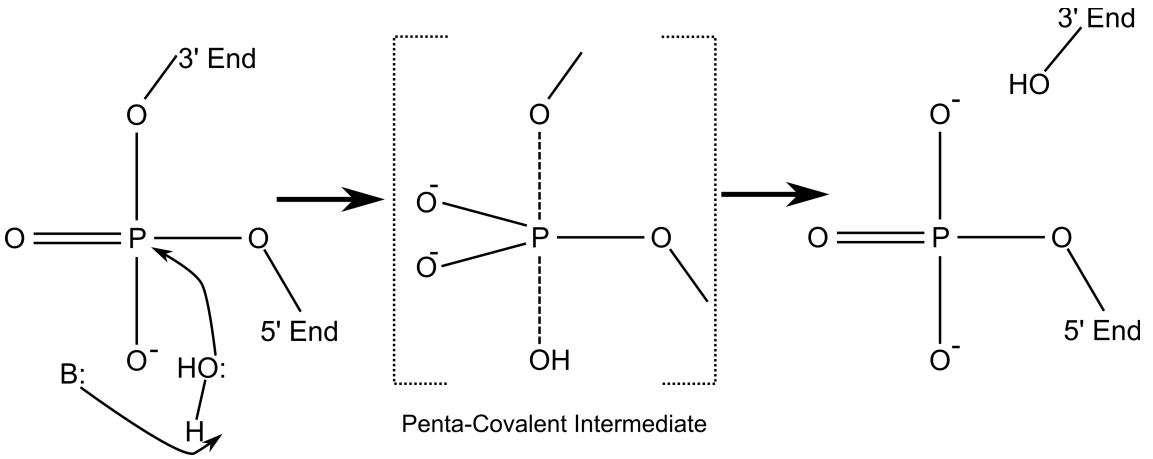

Figure 1.i. **Phosphodiester hydrolysis.** The base (**B:**) activates a nucleophile (in this case H_2O) that attacks the phosphodiester. This results in a penta-covalent intermediate, with the leaving group on the axial position (usually the 3' bridging oxygen). Finally the leaving group is ejected and the product is a 5' phosphate monoester. It should be noted that metal ions could help with (1) creation of the base, (2) stabilisation of the penta-covalent intermediate and (3) reduction in energy of the product.

1.1b Chemistry of nucleases

Chemically, nucleases catalyse a nucleophilic substitution reaction. This was shown to be mainly an associative nucleophilic substitution (S_N2), due to its characteristic inversion of stereochemistry of the product.[5] A phosphodiester bond by itself in water pH 7 at 37 °C has a rate of hydrolysis of 10^{-22} s^{-1} (or one bond broken per $3x10^8$ million years).[6] Nucleases bring the half-life of a phosphodiester bond down to hundreds of seconds, a rate increase of 19 orders of magnitude. Nucleases achieve this by bringing all the required components in close proximity to each other to make the reaction favourable. The phosphodiester bond is positioned for attack by an activated nucleophile (**Fig. 1.i**). Nucleophiles include water, tyrosine, serine or histidine residues, and the 2'-OH group in RNA. Furthermore the penta-covalent intermediate or transition state is usually stabilised by some external positive charge, like that from divalent metal ions. It might seem that such a simple methodology would produce a very conserved pattern of active sites and enzyme structures. The chemistry is deceiving however, as the range of nuclease enzyme structures is vast.

1.1c Substrate specificity

In order for a nuclease to function properly it needs to have a substrate to act on. This is especially important for a nuclease, as the substrate is the genetic material that codes for life, errors or inactivity could result in cancer or other serious genetic illnesses. This regulatory problem could be solved by two distinct molecular mechanisms. Firstly the nuclease could interact with a protein partner, receive post-translational modifications or bind with a co-factor, such as a metal ion. These processes either inhibit or stimulate the nuclease's activity or ability to bind to a substrate, thereby allowing the cell to keep its nuclease activity in check. Secondly, by having a nuclease exhibit a high degree of specificity for their DNA substrates, thereby only hydrolysing DNA or RNA that needs to be cut. This thesis will predominately look at the second mechanism, focusing on how nucleases exhibit an inherent specificity for their substrates.

Nucleases have traditionally been categorized by looking at how they achieve their nucleophilic substitution and what substrates they act upon.[7] Metal ions (as mentioned above) are predominately used in a wide range of nucleases to a) deprotonate water (activate a nucleophile), b) stabilise the negatively charged penta-covalent intermediate or transition state and c) lower the free energy of the product. Magnesium ions (Mg^{2+}) are the most frequently used ions due to their unique coordination geometry and ubiquity within the cell. Nucleases are categorised as one-, two- or three-metal-ion dependent, or may be metal-independent.

In terms of substrate specificity, broadly speaking there are two categories of nucleases: sequence-specific (restriction enzymes), or structure-specific. The former recognises a very select sequence of DNA or RNA and hydrolyses endo-nucleolytically, either double strand or single strand. The mechanism by which selection happens is in part by recognition of the chemistry of the DNA base pairs. The latter type of nuclease, which this thesis will focus on, is the structure-specific type. These enzymes do not primarily discriminate catalytic activity by base-sequences, but rather discriminate based on the secondary structure of DNA.

1.2 The Flap Endonuclease Superfamily

1.2a Summary of the superfamily

One such superfamily of structure-specific nucleases involved in DNA metabolism, replication and repair is the flap endonuclease superfamily. This includes the enzymes exonuclease (EXO1), gap endonuclease (GEN1), exoribonuclease (XRN1), xeroderma pigmentosum-G (XPG) and flap endonuclease (FEN1). The family members are all linked by their similar protein fold and two metal ion-dependency. These enzymes highlight the diversity in activity and cellular function of nucleases, even within one family. Each protein still performs a simple 5’ phosphodiester hydrolysis with remarkably similar active sites, but each one has its preferred substrate.

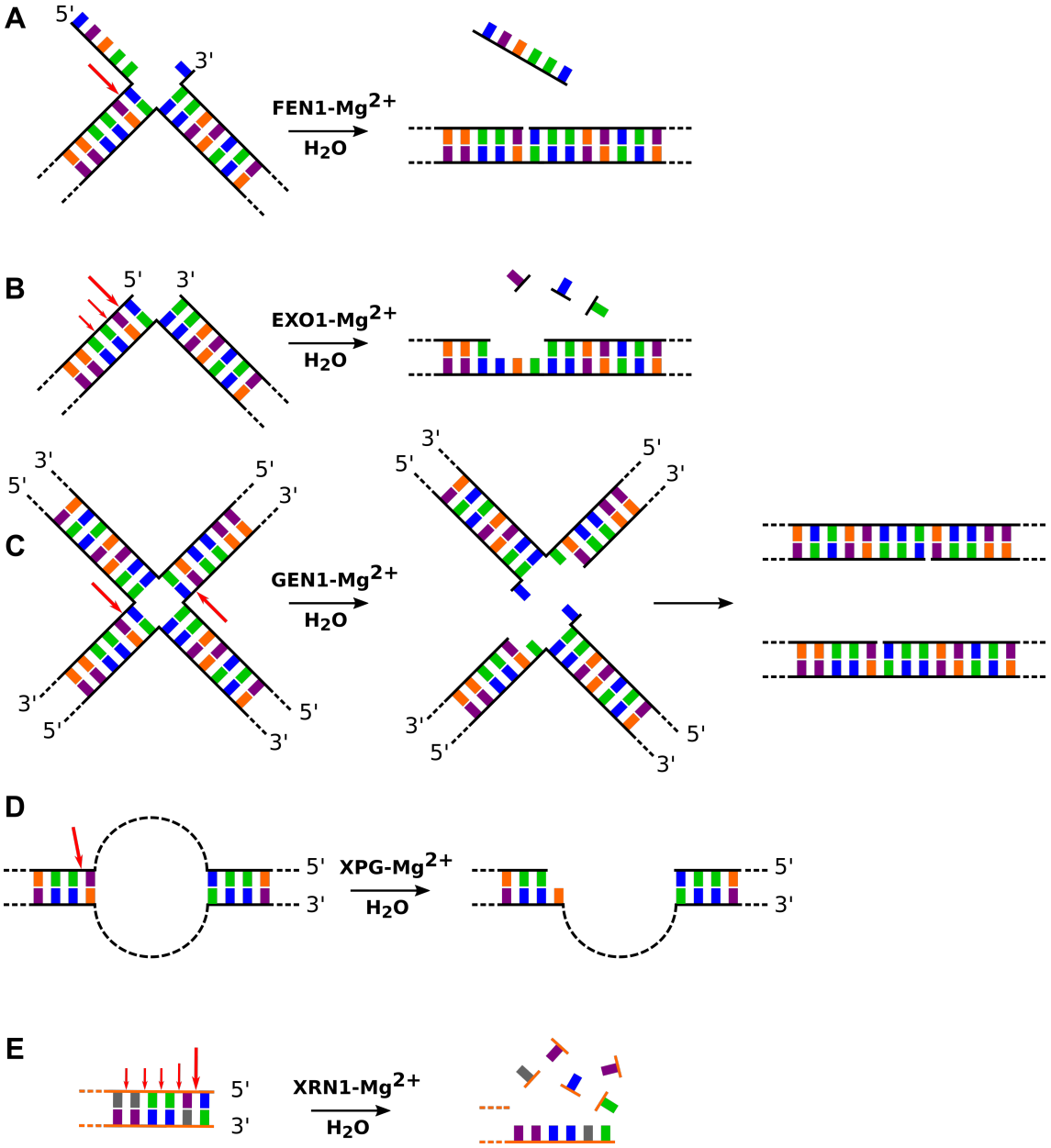

Figure 1.ii. **Superfamily substrates and their sites of activity.** Sites of phosphodiester hydrolysis are denoted in red arrows, DNA has a black backbone and RNA has an orange backbone. (A) FEN1 hydrolyses double-flapped DNA, with an endonucleolytic reaction occurring one nucleotide into the dupliex. (B) EXO1 hydrolyses a nicked DNA exonucleolytically with slight processivity. (C) GEN1 homodimer hydrolyses two mirrored sites on a Holliday junction to resolve it. (D) XPG hydrolyses large bubble like structures (dotted line) on one side of the bubble. (E) XRN1 hydrolyses RNA bases in a 5' to 3' direction highly processively.

FEN1 hydrolyses, very specifically, one nucleotide into the duplex of flap-like structures found in replication and repair pathways (**Fig.1.ii.a**).[8] EXO1 is involved in double-strand break repair and mismatch repair.[9,10] Furthermore it is only found in eukaryotes. Its primary substrate is that of a 5' blunt end, or a nicked piece of DNA (**Fig. 1.ii.b)**. There is some evidence of limited processivity, as it can cleave roughly 6-7 nucleotides exonucleolytically before stopping. GEN1 is the main holliday junction (HJ) 'resolvase', allowing for these potentially dangerous DNA

structures to be removed from the cell. It is the only superfamily member that is confirmed as a dimer, and uses the mirrored phosphodiester hydrolysis activity to cut each side of the HJ to achieve two double-stranded products (**Fig. 1.ii.c**). XPG cuts open large 'bubble' structures in DNA found during nucleotide excision repair (**Fig. 1.ii.d**), and works as part of the XP pathway.[11] Finally, XRN1 cuts very precessively ribonucleotides, and is involved in mRNA degredation (**Fig. 1.ii.e**).[11]

1.2b The role of FEN1 in DNA replication

FEN1 is involved in many processes, including long-patch base excision repair and ribonucleotide excision repair, but it is most vitally involved in replication.[8,12] The process of DNA replication uses a large host of proteins to regulate the entire process. Most are outside the scope of this thesis. To understand why FEN1 is so vital it is necessary to look at eukaryotic replication. It starts at specific sites on eukaryotic chromosomes called origin sites.[13] In brief, origins are regulated by various replication initiation factors, which have to form a pre-replicative complex on the DNA before the origin can 'fire' and hence replication can proceed. The regulation of these pre-replicative complexes is known as 'licensing' of replication and is essential to ensure only one set of replication is carried out in one cell growth cycle.[14]

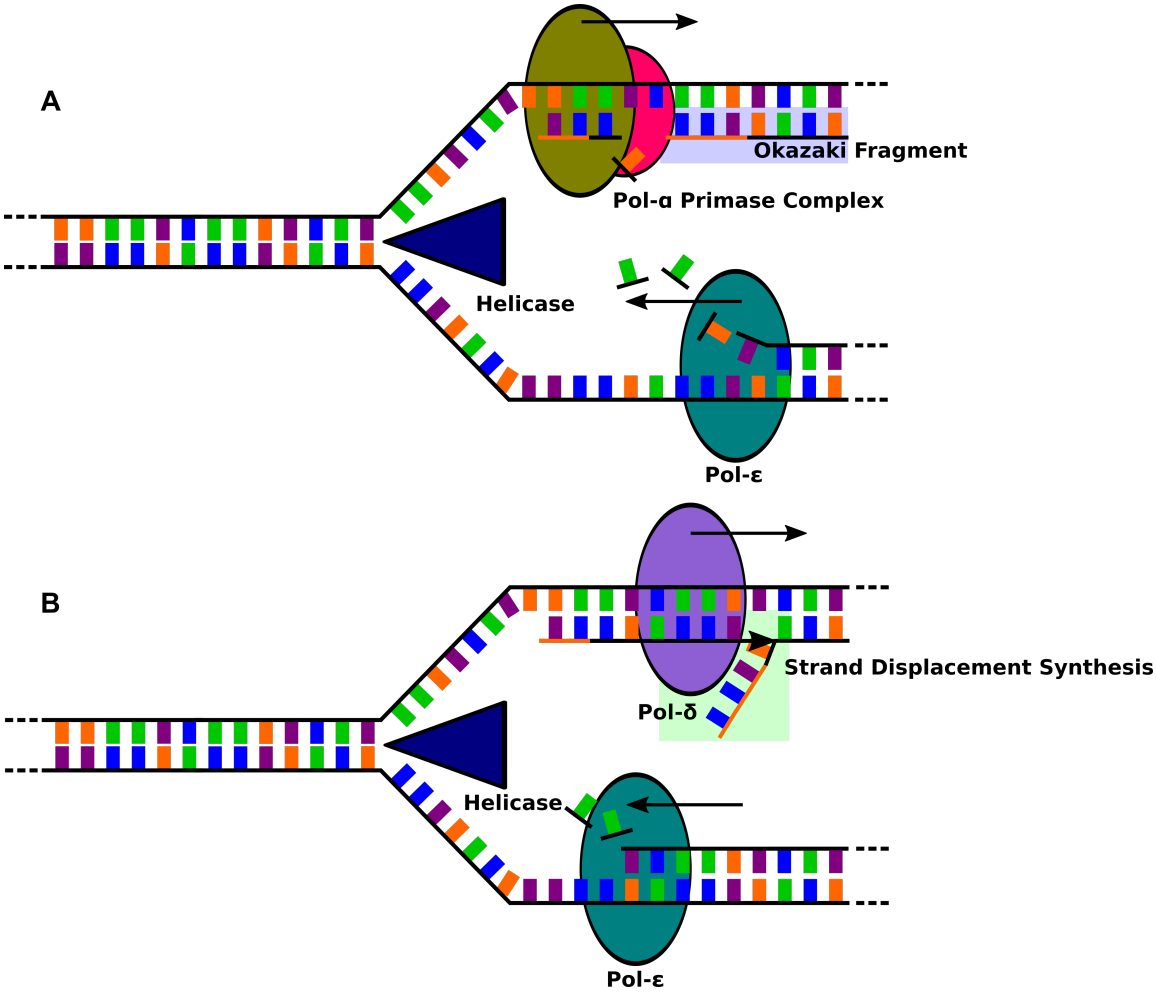

Figure 1.iii. Cartoon representation of a replication fork. (A) DNA is unpaired by helicase and splits into a leading strand and a lagging strand. On the leading strand Pol-ε polymerises new DNA continuously towards the replication fork. On the lagging strand Primase binds a few RNA primers before Pol-α extends them a bit more. (B) Pol-δ takes over from Pol-α and polymerises until it runs into the previously made Okazaki fragment, whereby strand displacement synthesis occurs, creating a 5' flap like structure, which is the main substrate for FEN1.

Once the origin fires, it is aided by a trimeric protein scaffold, the proliferating cell nuclear antigen (PCNA), which guides other proteins on and off the replication site.[15] The DNA helix is unwound and unzipped by a helicase, which forms a replication fork (**Fig. 1.iii.a**). Synthesis starts with a Polymerase-α Primase hetero-tetramer complex (Pol-α Primase).[16] The Primase segment can synthesise *de novo* RNA primers (about 5-7) from the ribonucleotides in the nucleus on the replication site. The Pol-α segment of the complex is proficient at extending already synthesised strands, so it uses the RNA primer as a scaffold to extend a small number of DNA nucleotides (about 7-14). The Pol-α Primase complex is then mediated off the site and Polymerase-δ (Pol-δ) takes over long strand DNA synthesis (**Fig. 1.iii.b**).

Polymerases are another metal-dependent enzyme involved with phosphodiester formation and hydrolysis. However, Pol-δ can only synthesise DNA precessively in a 5' to 3' direction,

which presents a problem. On the template strand the DNA is synthesised continuously *towards* the replication fork in a 5' to 3' direction. The lagging strand however is going in the opposite direction, which means that the polymerase can only synthesise in a 5' to 3' direction *away* from the fork, resulting in discontinuous polymerisation. Pol-δ latches on to a small Pol-α synthesised DNA strand near the fork and synthesises as far as it can forming small strands of DNA called Okazaki fragments. Crucially, when a Pol-δ from the preceding fragment runs in to the 5' end of the former fragment strand displacement synthesis occurs (**Fig. 1.iii.b**). This is where the polymerase displaces the preceding RNA/DNA strand resulting in a RNA/DNA flap structure.[17] These flap structures need to be removed and the double strand product ligated in order for the lagging strands to form an identical replicate of the original DNA. The DNA needs to be cleaved in a structure-specific manner, without regard to the sequence of the DNA and it needs to occur roughly 50 million times per cell cycle.[17] This is why FEN1 is a crucial enzyme for most life on earth.

In fact, FENs are found in all organisms from early phage through to mammals. Older FENs such as bacteriophage T5 FEN have a preference for a different 5' exonucleolytic activity that trims flaps rather than cutting them.[18] The sheer ubiquity of FENs would suggest that the protein is of significant importance to maintaining the integrity of the genetic code across all life forms. This has been demonstrated in mammals where a FEN1 knock out in the mouse genome is embryonically lethal.[19] Studies have also been carried on the yeast *Saccharomyces cerevisiae* where a null mutation of Rad27 (a FEN homologue in yeast) display increased genomic instability, with a strong mutator phenotype.[20] On the other hand if mammalian FEN1 is overexpressed in the genome it leads to an increase in tumor aggressiveness, perhaps increasing the stability or turnover of the DNA in cancerous cells.[21]

1.3 FEN1: structure and function

1.3a A review of the function of hFEN1

FEN1 is a DNA-structure specific nuclease, therefore it catalyses the endonucleolytic removal of 5' flaps by hydrolysing one nucleotide (1nt) into the synthesised double strand region of DNA with a high degree of specificity. The preferred substrate structure is a double-flapped DNA molecule with any length of 5' flap on the downstream duplex and a 1nt 3' flap on the upstream double strand (**Fig. 1.ii.a**).[22] This structure is present at the junction of the two Okazaki fragments *in vivo*. Using two Mg^{2+} ions as catalysts, FEN1 hydrolyses the DNA 1nt into the downstream region via a penta-covalent phosphate intermediate or transition state, leaving a

5'-phosphate at the end of the reaction.[23] This generates a nicked DNA because the upstream 3'-flap is complementary to the template DNA. This final structure is a substrate for DNA ligase and can immediately be processed to complete replication, thus minimising extra DNA repair steps.

The main questions are therefore, how does FEN1 achieve this molecular recognition for a chemically very similar but structurally different DNA substrate, and how does it catalyse on a biochemically relevant timescale? To answer these questions there have been numerous structural, kinetic and biophysical analyses done on FEN1. The detailed focus of this thesis will cover mainly human FEN1 or hFEN1, but some insights can be gained from looking across phyla at the wide variety of FENs and how they compare to each other.

1.3b Structures of FEN1 with DNA

The sequence of hFEN1 has been well characterized, there are two domains: a nuclease-core domain present from residues 1-336, and a protein-binding domain from residues 337-380. The protein-binding domain links to many other proteins (including PCNA) that mediate FEN1's activity *in vivo* and are important in its overall cellular function.[24,25] The protein however, retains its catalytic function and DNA-binding capabilities in the absence of these 43 residues, albeit with a slightly higher Michaelis constant (K_M).[26] FENs across phyla share a similar primary amino acid sequence to hFEN1 as well as similar structural motifs.

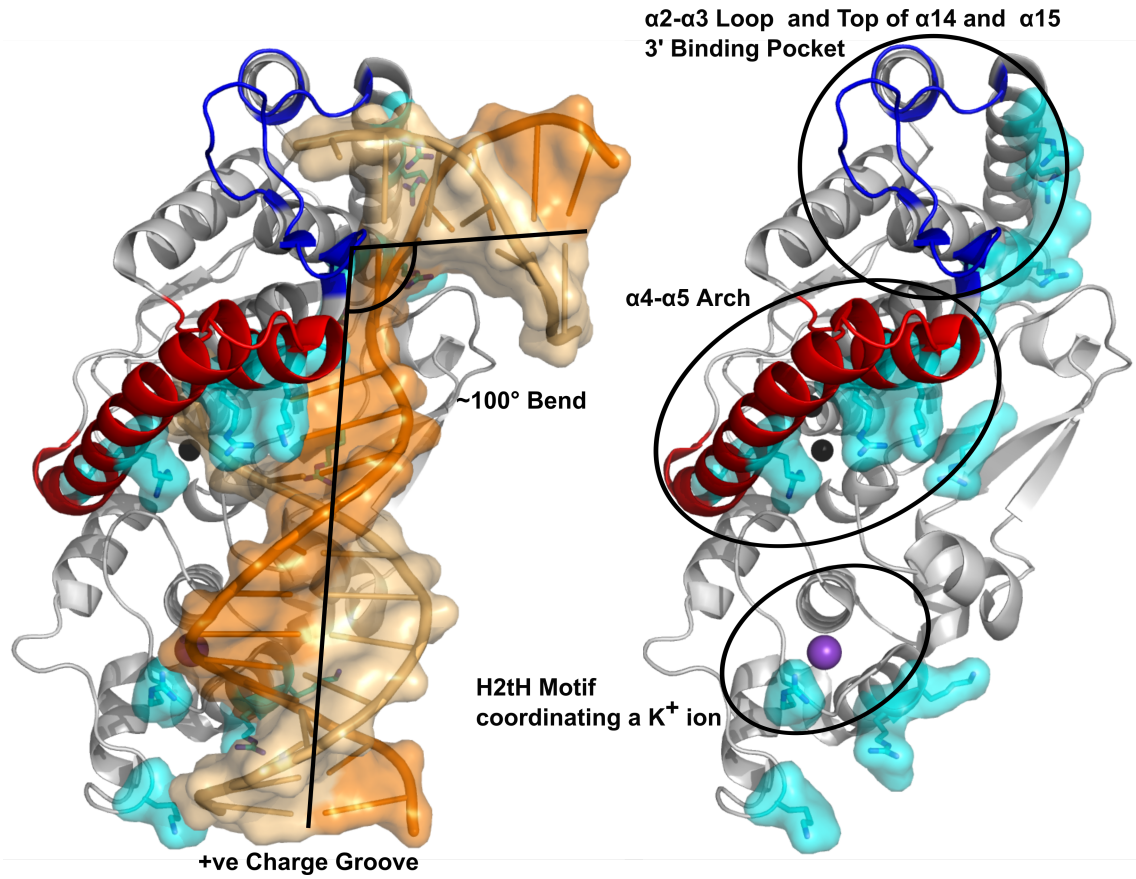

Figure 1.iv. DNA recognition sites on hFEN1. The two figures are from a product-bound hFEN1-DNA crystal structure complex (3Q8K).[27] The right hand side figure has the DNA hidden for clarity. The highlighted cyan residues are basic residues, which lie along a positive charge groove on the protein. There is also a Helix-two-turn-Helix motif (H2tH) with a K^+ ion coordinated in it for additional DNA binding. The DNA when bound forms a 100° bend by having two juxtaposed binding sites, and by wedging the top of α2 into the junction, forcing the nicked DNA into this conformation. The α4-α5 archway houses the active site at its base with two Mg^{2+} metal ions (black spheres), but also houses many basic residues that contact the 5' flap and the template.

Across all FENs the nuclease-core structure consists of a 7-stranded β-sheet platform, which is surrounded by 15 α-helices. There is a long positively charged groove down the center for general dsDNA binding (**Fig. 1.iv**). A potassium ion is coordinated via peptide carbonyls in a helix-two-turn-helix motif (H2tH), a double-stranded DNA binding motif that binds backbone DNA at the downstream region of the template strand. At the junction of the duplex DNA, the nucleic acid is forced into a bent configuration by approximately 100° around one phosphodiester bond by the positioning of the protein helix α2, as a hydrophobic wedge into the junction (**Fig. 1.iv**). The DNA is further held in place by a second site upstream. Contacts in α3 and the α14-α15 loop help anchor the upstream DNA helping adopt the DNA into a bent conformation. This provides specificity towards a dsDNA that can easily bend at that particular phosphodiester bond. A weaker and broader specificity towards B-DNA is seen, as the H2tH motif is roughly one pitch length (34 Å) away from the bent flap structure. In this way most of the initial substrate recognition is done by interactions between the template strand of DNA

and the protein and, as is to be expected for a sequence independent nuclease, its main recognition sites are on the phosphate backbone, not the base pairs.

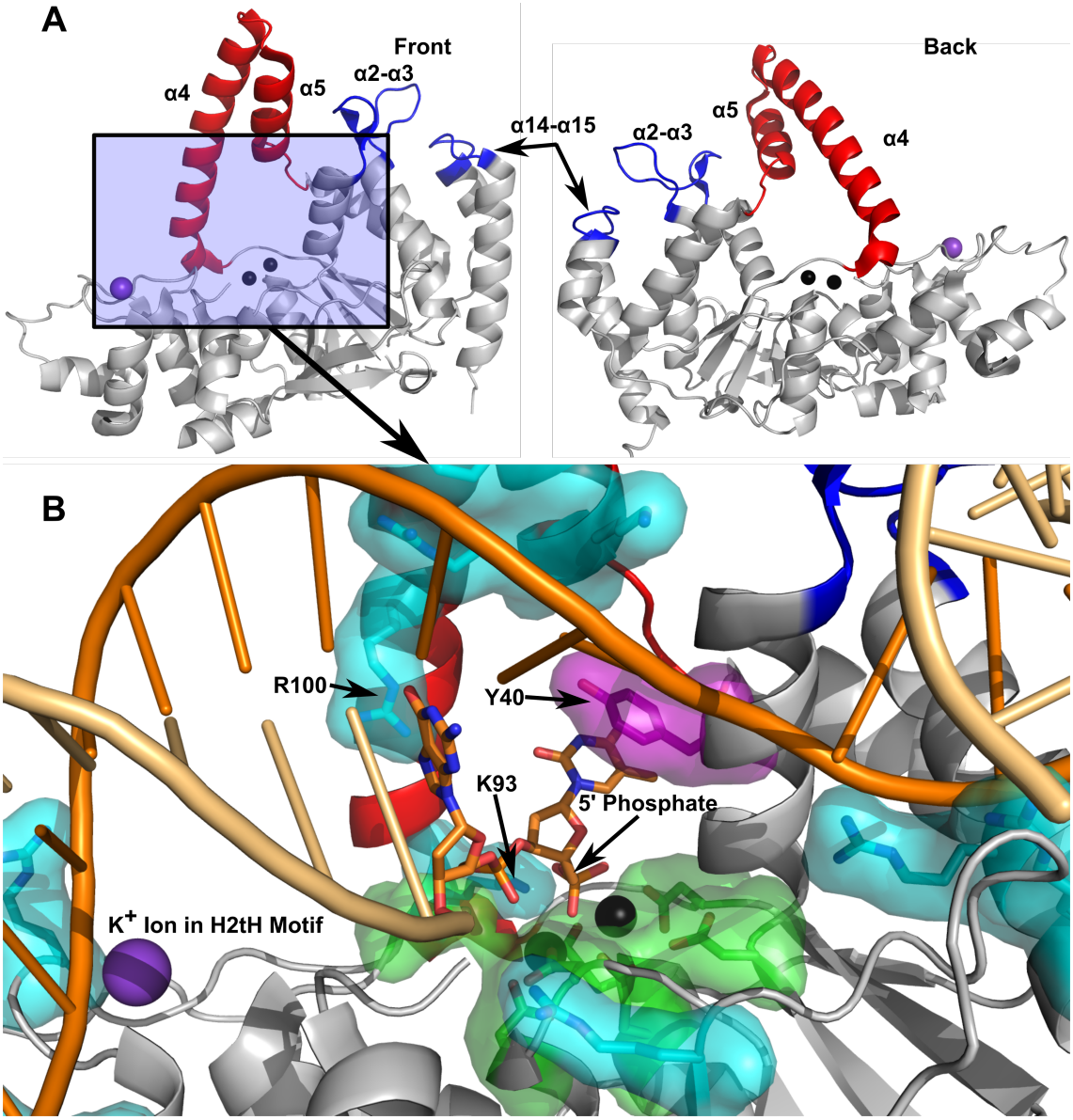

Figure 1.v. A close view of the active site of hFEN1. (A) Overview of the hFEN1 crystal 3Q8K (as seen in **Fig 1.iv**),[27] showing the 3' binding pocket (blue), arch (red) and K^+ ion (purple). The crystal structure uses Sm^{3+} ions (black spheres) to aid crystal packing and inhibit activity but *in vivo* this would be the location of two Mg^{2+} ions. DNA is hidden for clarity. (B) Close-up of the active site showing more of the specific interactions. Y40 (magenta) stacks with the terminal 5' nucleotide base of the cleaved product. Carboxylates (green) coordinate with the Mg^{2+} ions to hold them in the ideal coordination to stabilise the negative 5' monophosphate. Basic residues in the arch (in cyan, R100 and K93) hold the 5' flap down in position. There is a visible deformation of the two bases going into the active site.

A look at the active site of hFEN1 in more detail shows it is located at the base of a 'helical arch' (α4-α5), with two Mg^{2+} ions held in place with four inner-sphere and three outer-sphere carboxylic groups (**Fig. 1.v.b**). The presence of Mg^{2+} ions has been shown to be crucial for catalysis, due to reasons listed above (**Section 1.1c**). It should be noted however that the active

site of hFEN1 can accommodate a wide range of multi-valent metal ions. In the crystal structure above (3Q8K) Sm^{3+} ions are used for crystallisation and allow for DNA binding. Furthermore, Ca^{2+} ions allow for a binding and bending of the DNA.[28] However, Ca^{2+} and Sm^{3+} do not facilitate catalysis. It is not quite understood why, but it is likely to do with the size and coordination geometry of the ions, which do not coordinate correctly with the phosphodiester.

In X-ray crystal structures of both product (3Q8K)[27] and substrate (5UM9)[29] bound hFEN1 there is a high degree of secondary structure ordered around the 5' terminus, with it snugly fitting between the helical arch. The flap is held in place by several basic residues in the arch (K93, R100 and R104) and a conserved nucleobase-stacking residue (Y40) (**Fig. 1.v.b**). In this way the helical arch and other residues deliver the hydrolytic phosphodiester into the correct coordination between the two Mg^{2+} ions allowing chemistry to occur.

A further structural recognition domain of hFEN1 is the 3' flap-binding pocket, which consists of the α2- α3 loop and α14-α15 loop. These protein sequences associate with a 1nt long 3' DNA flap. This not only allows for an enhanced specificity for the preferred dsDNA double-flapped structures, but also allows specific hydrolysis of the phosphodiester 1nt into the downstream dsDNA region. There is structural evidence of connectivity between the α2-α3 loop and the helical arch as several important sidechains of residues from the top of α2 are very close in space to the bottom of α5 (i.e., R47 and I44 and V131). Therefore it has been suggested that allostery between the α2-α3 loop and the helical arch play an important role in substrate recognition.

In summary, substrate recognition and DNA binding have been associated with these three regions:

1) Juxtaposed DNA binding between downstream binding sites and upstream binding sites, which confer specificity for dsDNA able to bend.
2) The helical arch which accommodates and delivers the single stranded 5' flap to the metal ions for chemistry to occur.
3) The recognition of the 3' flap in the α2-α3 loop and a14-a15 pocket causing potential allosteric changes elsewhere in the protein and DNA.

There is however a discrepancy between X-ray crystal structures of a wide variety FENs. So far only structures crystallized with DNA have been examined. There are important structural changes in X-ray crystal structures without DNA.

1.3c Protein Dynamics

In a crystal structure of hFEN1 without DNA (1UL1) we see that portions of the arch region, α2-α3 loop and the hitherto unmentioned region of the β-pin (β6-β7) are unobservable.[24] This could be for many crystallographic reasons such as a high B-Factor, which indicates the amount of static or mobile dynamics of a region or just simply a result of refinement errors. In a field of phyla we actually observe the arch region and α2-α3 loop in a range of frozen conformations, sometimes missing, sometimes in a random coil, sometimes in discernable secondary structure, or a mix between all three (**Fig. 1.vi**). Whether the range of structures arises due to differences in the crystal lattice or crystallization conditions rather than a dynamic property of the FEN protein is unknown.

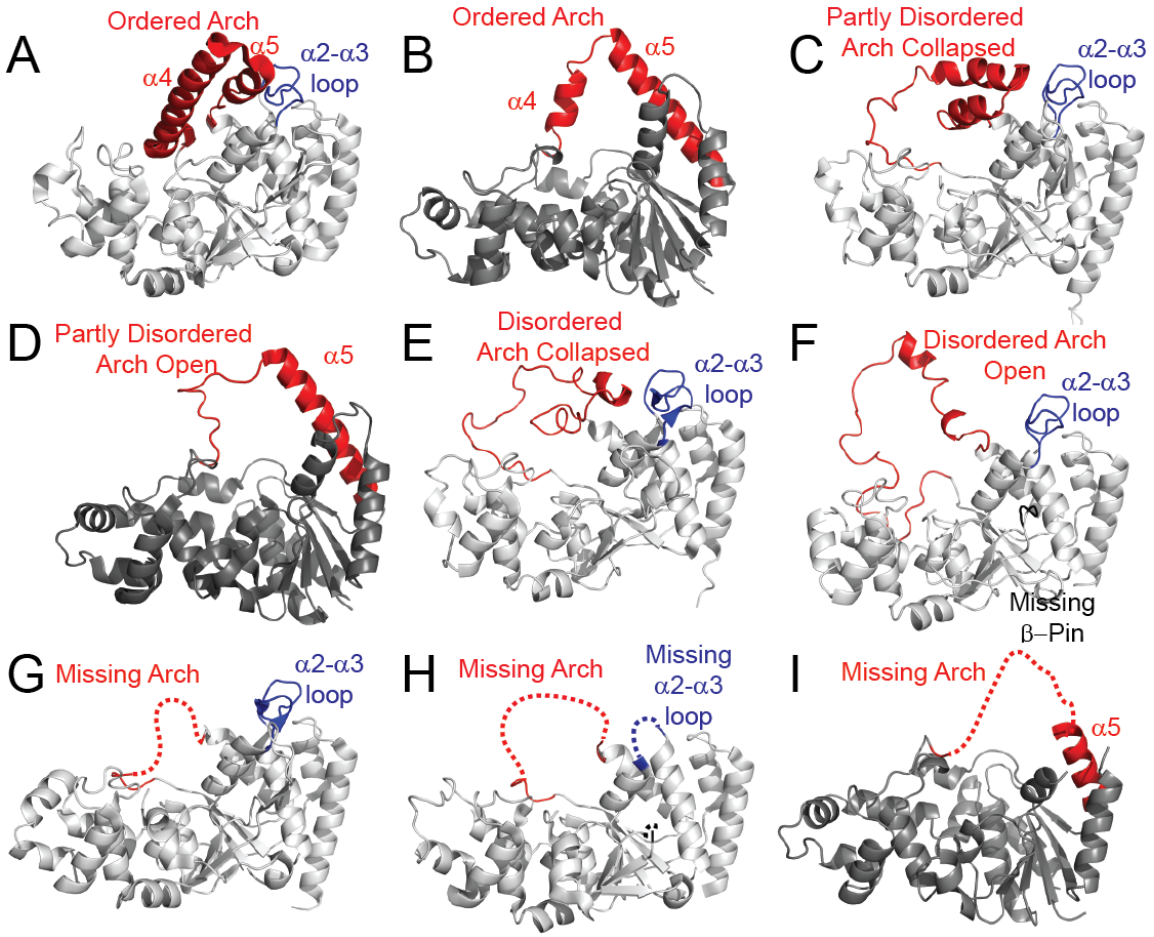

Figure 1.vi. The arch regions and α2-α3 loops of FENs from other organisms. Crystal structures of (A) *Desulfurococcus solfataricus* (3ORY)[30] and (B) T5 bacteriophage FEN (1EXN)[18] show an α-helical arch region. Crystal structures from (C) *Pyrococcus furiosus* (1B43)[31] and (D) T5 FEN (5HML and 5HMM)[32] show an arch region with a degree of secondary structure, and in (C) it is collapsed. Equally in the structure of (E) *Pyrococcus horikoshii* (1MC8)[33] there is an arch region with very limited secondary structure and it is collapsed over the active site. In (F) *Methanococcus jannaschii* (1A76 and 1A77)[34] there is an extended conformation of arch with very limited α-helical nature. Finally for (G) *Methanopyrus kandleri* (4WA8)[35], (H) hFEN1 in complex with an inhibitor (5FV7)[26] and (I) T5 FEN (1XO1)[36], there is a lack of electron density in the arch region. Furthermore in the case of (H), the α2-α3 loop is missing.

The supposed dynamics in these regions has been a focus of study for some time. Kinetic and fluorescence anisotropy studies coupled with site-directed mutagenesis have shed much insight into the link between dynamics and catalysis. In particular the role of the arch and its ability to pass the 5' flap through it, or in other words 'thread' the flap. Proline mutations were made in the arch, whereby flexible leucine residues (L97, L111 and L130) were mutated to equally hydrophobic, but rigid prolines. These mutants have relatively unaffected binding affinities (K_D), whereas their single-turnover rates of reaction (k_{ST}) were affected greatly. L97P showed the greatest reduction (>100,000-fold reduction), possibly because it may have disrupted other basic residue contacts (K93, R100).[37] L111P and L130P showed reductions in rate of 15,000-fold and 3,000-fold respectively. These residues were at the top of the arch, far away from the active site but still affected the single-turnover rate significantly. The question of how these mutations cause such catastrophic reductions in rate in a region far away from the active site is difficult to explain using static structural models.

More experiments done with a bulky 5' streptavidin tetramer linked to substrate DNA show further facts regarding the dynamics of the arch region. If streptavidin is added before any protein or catalytically activating Mg^{2+} ions are added it creates a 'blocked' substrate, which slows catalysis considerably (15,000-fold decrease in k_{ST}) (**Fig. 1.vii.a**).[38] If however, the streptavidin is added after pre-mixing protein in the solution and then activating with Mg^{2+} ions, the catalysis is not changed considerably at all, hence illustrating a 'trapped' complex (**Fig. 1.vii.b**). The results show that a blocked substrate cannot react while a trapped substrate can. Therefore the DNA must thread through the arch during catalysis, which was subsequently confirmed by the substrate structure of hFEN1 (5UM9)[29] and T5 bacteriophage FEN1 (5HNK).[32] Furthermore, it was also shown that a short 5' dsDNA as well as ssDNA (gapped substrate) can also be threaded through the gateway and undergo catalysis.[22,38] In the crystal structure of product or substrate however, the arch presents a gap too small for dsDNA. Therefore the arch was postulated to exist in an extended conformation, which must position the catalytic machinery in a state amenable for hydrolysis. Therefore the current hypothesis is that a disorder-to-order transition of the arch must occur for catalysis to happen.

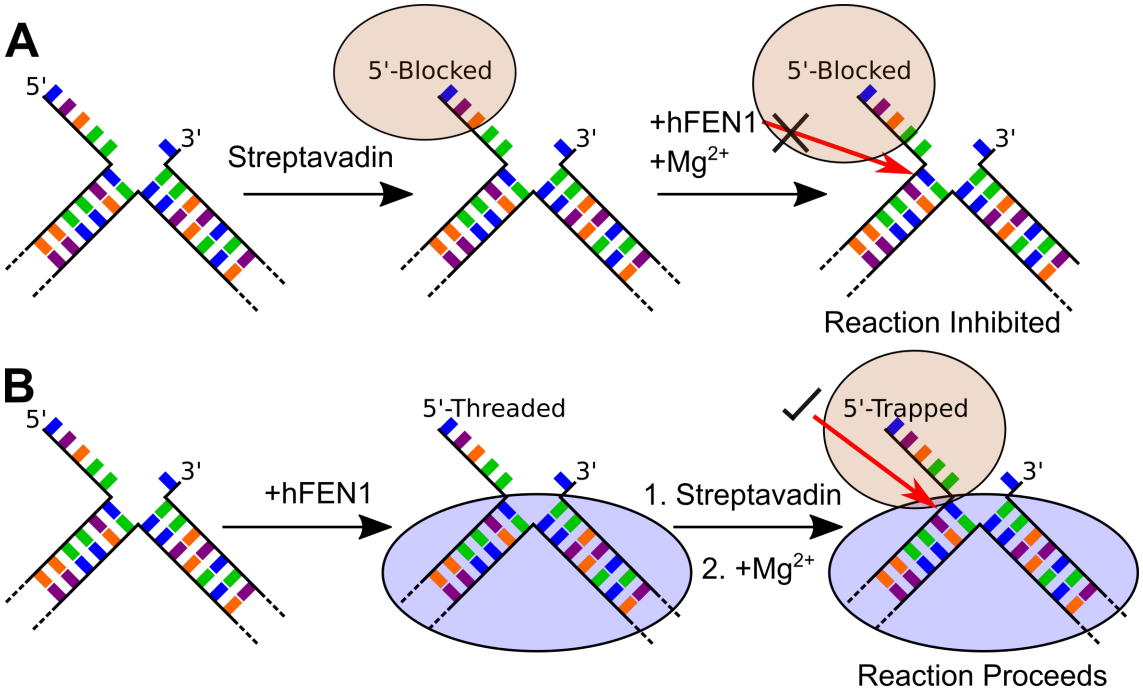

Figure 1.vii. **Illustration of a blocked or trapped reaction schematic.** Both DNA strands contain a 5' biotin attachment, which the streptavidin (orange oval) binds to. (A) Illustration of the sequence of a blocked reaction. First, streptavidin is added to the substrate DNA and allowed to equilibrate. Second, the protein along with catalytically essential Mg^{2+} ions are added, the result is an inhibited reaction, as the protein cannot access the 5' end of the substrate. (B) Illustration of a trapped complex reaction. First protein is added in absence of Mg^{2+} ions, so reaction is halted. Second, addition of the streptavidin allows for the protein to become trapped on the substrate, confirmed by the subsequent addition of Mg^{2+} ions allowing for reaction to occur.

The role of the α2-α3 loop is less known, although it displays similar levels of structural heterogeneity in X-ray crystals. Unpublished results by Dr. Mark Thompson show the loop is important for catalysis however, and reactions without a 1nt 3' flap have single-turnover rates up to 1000-fold lower than the double-flapped substrate. Mutating the residue R47, which is said to mediate contact with α5, to an alanine displays single-turnover rates 300-fold lower than the wild type protein. Furthermore, the structured loop present in 3Q8K fits the categorisation of an omega loop (Ω-loop). Ω-loops are non-regular secondary structures involved in catalysis, folding and stability.[39] Thus it could be very likely that the α2-α3 loop mediates the binding of the 3' flap and hence aids folding of the arch.

1.3d DNA Dynamics

Aside from the protein's conformational change there are also conformational changes associated with the DNA. Förster resonance energy transfer (FRET) analysis with a donor dye on the downstream DNA and an acceptor on the upstream part of the DNA (**Fig. 1.viii.a**) can illustrate how close the two ends of the duplexes are to each other. An assay with increasing concentrations of hFEN1 can illustrate the ability of hFEN1 to bend the substrate. It was shown

that substrate bending happens regardless of many catalytic mutations, a mismatch at the +1 nucleotide (relative to hydrolysis), a lack of 5' flap entirely or even blocking of the 5' flap with streptavidin.[28] This shows that simply having a nicked structure is sufficient for hFEN1 to bend the DNA, despite some modifications or mutants having drastically reduced rates.

Further analysis using single-molecule FRET assays illustrates how hFEN1 protein can induce a 1nt 3' flap. This has lent some weight to an induced fit model of DNA bending.[40] It is likely to be much more complex than this, as the nicked substrate is documented to be able to exhibit extreme flexibility and it could adopt this conformation by itself.[41] For now, whether the protein induces bending by binding or whether the protein selects a transient bent conformation is a point of contention.

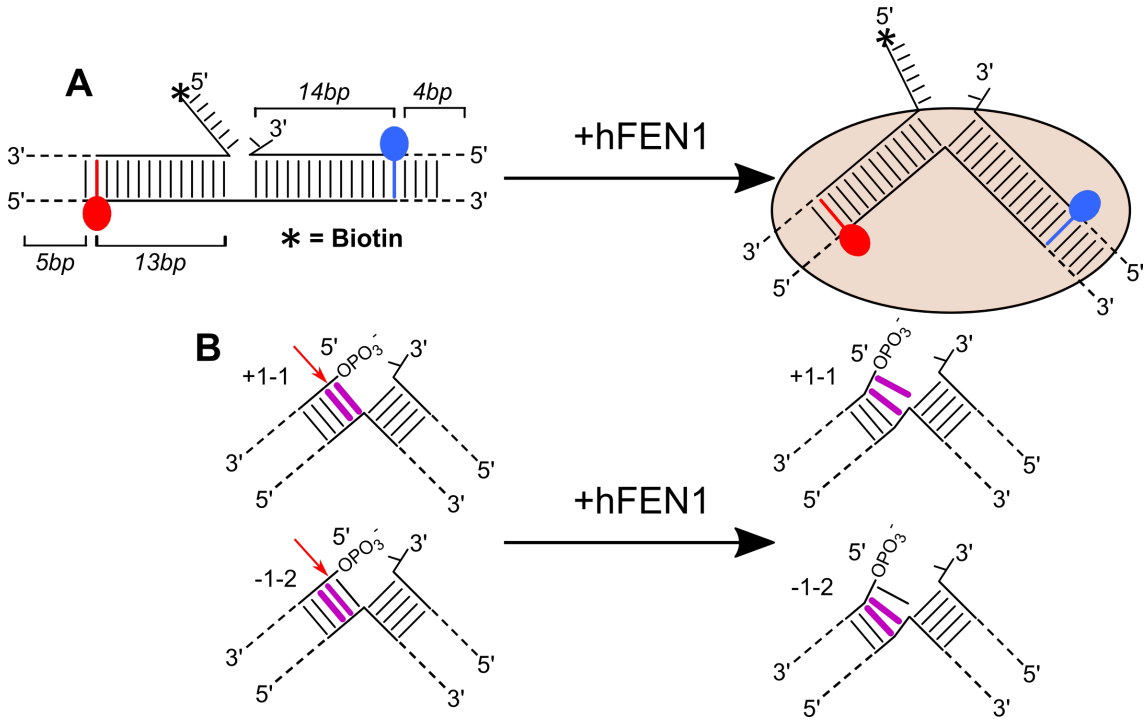

Figure 1.viii. Substrate DNA used to investigate DNA conformational changes. (A) Schematic of the substrate used in the bulk FRET assay with a donor dye, fluorescein (blue) on the upstream section and an acceptor dye, 5-Carboxytetramethylrhodamine (red) on the downstream section. When hFEN1 is added to the DNA it bends around the nick and brings the dyes closer together, increasing the FRET efficiency. (B) Schematics of ECCD substrates used, with tandem amino-purine bases marked in purple. The positioning of the bases (e.g, +1, -1 and -2) are relative to the scissile bond marked with a red arrow. When hFEN1 binds it causes a deformation that is detectable by circular dicrhoism at around 326 nm.

As seen from the DNA-bound crystal structures (3Q8K and 5UM9), the bend in the DNA causes the 5' flap to protrude into the active site region. The active site residues position themselves into the correct orientation to perform cleavage. However, the dsDNA still needs to deform in order to position the phosphate next to the active site. This deformation was visible using exciton-coupled circular dichroism (ECCD). By placing the exciton pairs of 2-aminopurine bases

next to each other (**Fig. 1.viii.b**), it was possible to measure deformations in the structure of the 5' reacting-strand, which showed whether the 5' end of the DNA has twisted into the active site.[28] For the DNA, it must not be blocked by bulky groups, have no mismatch and have a phosphate on the 5' end. For the protein, certain key residues must be present (L130, Y40, D181 and R100) and it must have divalent cations (Ca^{2+} in this case). The physical manifestation of the inferred conformational change is not known but has been postulated to involve: DNA twisting, deformation or double-nucleotide-unpairing, but it is most likely twisting, as observed in the 5UM9 crystal structure.[29] Thus, the process is likely involved in substrate specificity as well as being one of the major conformational changes that influence the rate of catalysis.

1.3e Kinetics

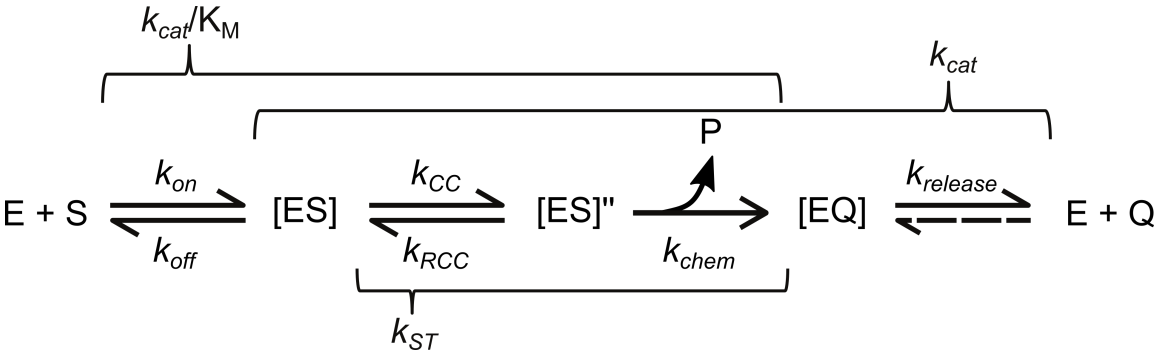

Figure 1.ix. The proposed reaction pathway of hFEN1. Each step has a rate constant associated with it, but most are not directly measurable in bulk, for example k_{chem}. The bracketed values displayed above and below are macroscopic quantities that can be experimentally obtained by kinetic experiments. Therefore k_{cat}/K_M measures all micro-rates up to and including the first irreversible step (k_{chem}). k_{cat} measures everything after substrate binding, and finally k_{st} measures everything after binding but before release. The proposed conformational change rates (k_{CC} and k_{RCC}) could be due to many physical movements and it is thought that the k_{st} reflects the rate of this conformational change. Adapted from (Finger et. al. 2012).[8]

A further look at kinetics can reveal a potential link between single-turnover rates and rates of conformational change. The proposed model for the catalytic cycle of hFEN1 is shown in **figure 1.ix**. Enzyme (E) binds to substrate (S) and forms an enzyme substrate complex (ES). This then has to undergo a conformational change to become chemistry competent (ES)'', at which point reaction occurs and the small 5' flap product (P) is released, forming an enzyme product complex (EQ). The nicked DNA product (Q) then needs to be released in order for the free enzyme to be ready to turnover and bind more (S). There has been a wealth of kinetic studies done to establish which of these processes play an important role.

Firstly the catalytic efficiency (k_{cat}/K_M) of FEN1 is close to the bimolecular diffusion association rates in solution (k_{on}) roughly of 10^7-10^{10} M^{-1} s^{-1}. This suggests that under pseudo first-order

conditions where substrate concentration is lower than K_M, the rate of the hFEN1 reaction is diffusion limited.[42] Under substrate-saturating multiple-turnover conditions, the maximal rate (k_{cat} = 3 s^{-1}) is slower than the respective single-turnover maximal rate (k_{STmax} = 22 s^{-1}), suggesting that product release ($k_{release}$) is the rate-limiting step under saturating substrate conditions.[22] It is interesting to note that product release and diffusion are probably minor factors at the *in vivo* replication fork, as hFEN1 location is mediated by various controls such as the PCNA.[25,43,44] Therefore the kinetic parameters after binding and before product release are likely the most biologically relevant. This means that the rate of conformational change after substrate is bound is likely to be an important factor in the catalytic cycle.

1.3f The Model So far

Conformational changes within the protein and DNA are likely to be key factors in the biologically relevant catalysis for hFEN1. They are also expected to play a key role in substrate recognition and hence aid specificity. There are various proposals as to what some of these conformational changes could be:

1) The arch going from a disordered to an ordered conformation.
2) The DNA bending.
3) Placement of the phosphodiester in the active site (threading and positioning)
4) 3' Flap engagement

The order and relative speed at which any of these processes happen, or even what the physical manifestation of each process is of debate at the moment. From FRET experiments, the DNA is likely to bend regardless of any other process. From the streptavadin trapped/blocked experiments, it is likely that the 5' flap must thread before, or at the same time as, the arch orders or else it gets blocked from entering the active site. Without a 3' flap the rate of reaction is reduced up to 1000-fold, but it is difficult to explain this difference in rate based on its interaction far away from the active site. It is still unknown how fast the arch is moving into its ordered state and how fast 5' DNA placement occurs. It is also a mystery what changes happen to residues in the rest of the protein and how fast they occur. Further information on these conformational changes is not only one of the key means to understanding the mechanism of hFEN1, but it might explain its specificity towards its substrate. In order to ascertain what particular residues are moving and how fast, a probe that can be used to monitor the motions of all the residues in the protein is necessary.

1.4 Measuring dynamics

1.4a Types of dynamic motions in proteins

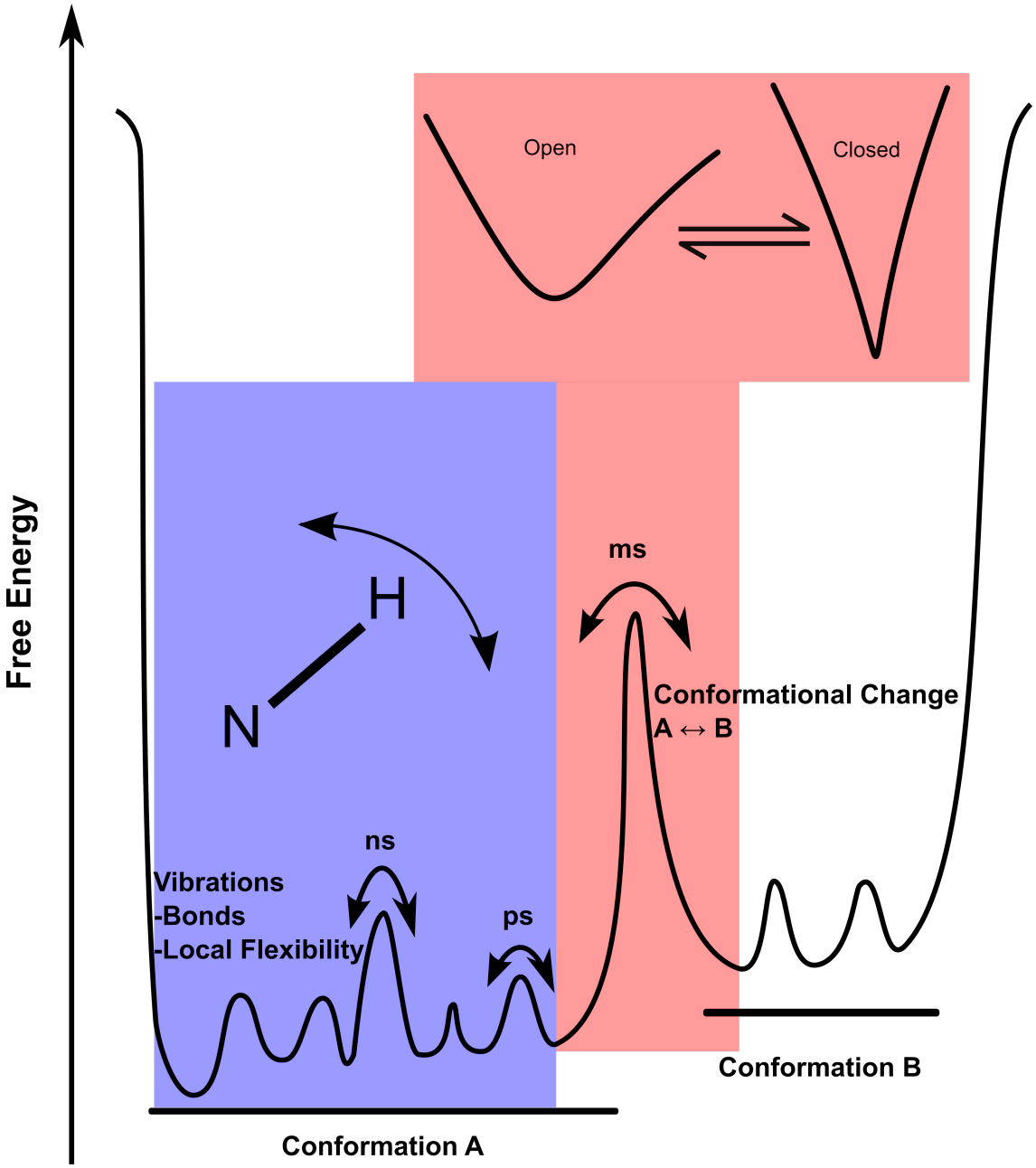

Figure 1.x. A free energy profile of a protein. Bond vibrations and more local flexibility (blue box) will have small, low energy conformational changes, which occur on the ps-ns timescale. A conformational change for regional protein motions (e.g. open to closed transition) (red box) will have a much larger energy barrier that takes place on the μs-ms timescale.

Proteins are in essence folded chains of chiral amino acids. Many 'weak' interactions such as hydrogen bonding, ionic interactions and hydrophobic forces hold them together. Therefore, proteins are not static molecules in solution; they are constantly sampling many conformations at timescales of femtoseconds to hours. A potential energy profile of an enzyme is not a simple curve; it is a complex mixture of transiently formed states, folded to unfolded, or conformation

A to conformation B (**Fig. 1.x**). Slow dynamic changes can be interpreted as loop or domain movements and these can play a key role in catalytic turnover or protein folding events.[45] There is also fast sampling of similar conformational changes on a bond vibrational timescale that allows for fast energy barrier crossing (**Fig. 1.x**).[46] In this way the protein is sampling many states, some of which are non-productive while others drive the catalytic cycle forward. Both fast and slow dynamics of proteins are therefore very important for catalysis as well as signaling,[47] thermostability[48] and folding.[49,50]

There are a few experimental techniques, as well as computational methods that can be utilised to ascertain the rates of protein dynamics. Single-molecule fluorescence, or FRET, involves labeling a donor and acceptor site on a protein with suitable chromophores and measuring the FRET efficiency or a fluorescence for a 'reactive' or 'unreactive' state. This is a powerful technique, particularly if the protein can be easily labeled and can offer very fast measurements in the microsecond timescale. Methods like this are often used in tandem with computational molecular modeling for complex systems. Furthermore, computationally harder quantum modeling is used for situations of electron transfer. However, by far the most important technique in terms of versatility and analytical power is NMR spectroscopy.

1.4b NMR Spectroscopy and protein dynamic measurements

NMR spectroscopy can provide insights into a broad range of dynamic timescales. These methods can show protein structural changes related to side chain rotation, loop motion, molecular tumbling, overall folding and domain reorientation of the protein.[51] There are several ways of measuring dynamics between two states using NMR observables (chemical shift, line width and intensity). These include line-shape analysis, exchange spectroscopy, relaxation dispersion techniques and real-time NMR for slower motions.[51] NMR approaches can also be used to measure exchange between protein states by changing an external property such as temperature or pressure, or by titration of a substrate, product or co-factor. However, a powerful method to initially identify residues of interest as well as to gain a representation of the flexibility of all residues within a protein is nuclear spin relaxation (NSR), and in particular the application of model-free analysis.

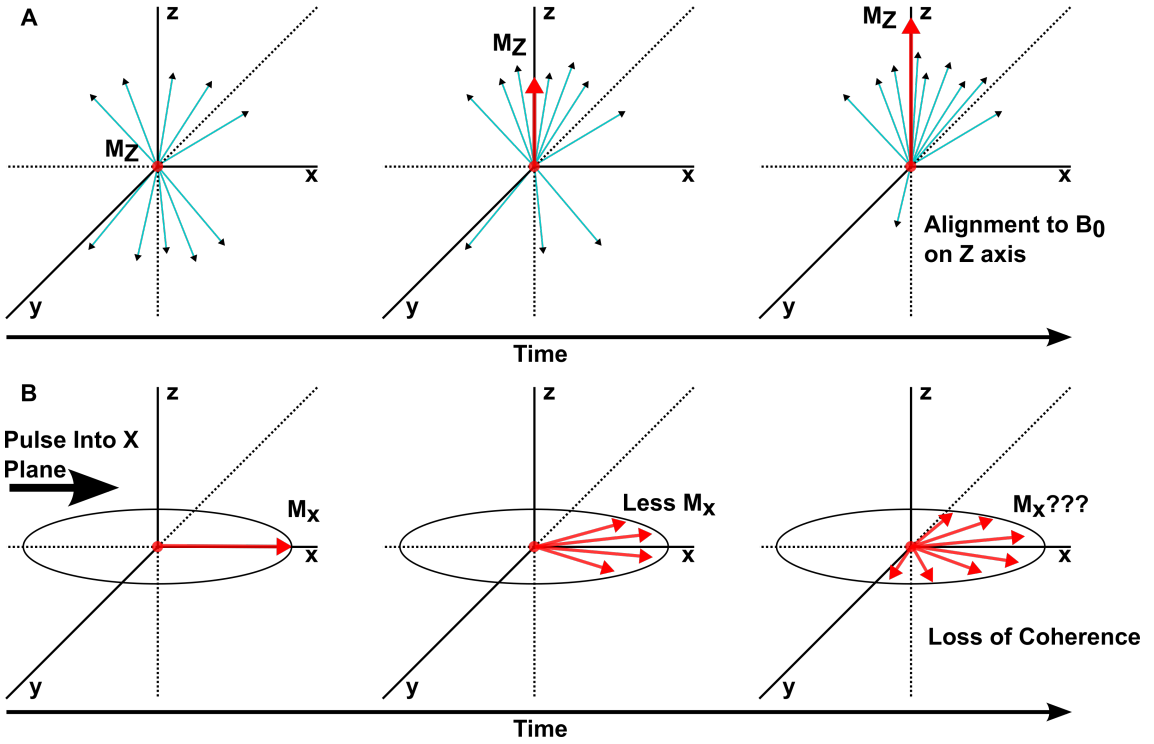

Figure 1.xi. Two mechanisms of spin relaxation. (A) R_1, longitudinal or spin-lattice relaxation is where the bulk spins transfer their energy to neighboring atoms, resulting in a re-equilibration to the Boltzmann distribution. The rate at which this happens is measured as R_1. (B) R_2, transverse or spin-spin relaxation. The figure shows the spins of one proton for all molecules in solution, in the rotating frame of reference. After a 90° pulse into the X-Y plane, initially all spins are coherent. However, minute differences in their rotations compared to the Larmor frequency will cause them to lose coherence, which causes a loss of signal in the X-Y plane, or a relaxation.

NSR uses the site specific NMR spectroscopy observables: Longitudinal Relaxation (R_1), Transverse Relaxation (R_2) and the Heteronuclear Nuclear Overhauser Effect (hNOE). R_1 is the rate at which spins relax to the Boltzmann population of spin-up and spin-down states in an applied Z-axis magnetic field B_0 (**Fig. 1.xi.a**). The value of R_1 can influence the maximum scan rate available for spectrum acquisition. R_2 is the rate of loss of coherence from spins in the rotating X-Y reference frame (**Fig. 1.xi.b**) and results in line-broadening in NMR spectra. For R_1, the spins can only make an energy transition *via* interactions between fluctuating magnetic dipoles at the Larmor precession frequency. R_1 and R_2 relaxation rates are very slow compared to the much faster flexibilities of bonds. Nevertheless, it is possible to link the relatively slow millisecond − second (ms-s) relaxation rates of R_1 and R_2 to picosecond − nanosecond (ps-ns) dynamics of a backbone N-H bond.

The speed of molecular tumbling affects the proportion of the lattice that is available to receive energy from nuclear spins. For large proteins which have slow molecular tumbling, a low proportion of the lattice is available and consequently R_1 is slow. Conversely, small proteins have much faster molecular tumbling and R_1 is fast. R_2 is more complex, as slower molecular

motions such as chemical exchange further enhance R2, so for macromolecules, R_2 is fast. Chemical exchange is the process by which one atom moves to another chemically distinct state, with the resultant loss of coherence. This rate of exchange (R_{ex}) can reveal more about the slower domain movements in proteins. Overall, the tumbling of the molecule therefore affects relaxation process of the whole molecule. However, bond flexibility also changes spin relaxation through two major mechanisms: nuclear dipole-dipole interactions and chemical shift anisotropy.

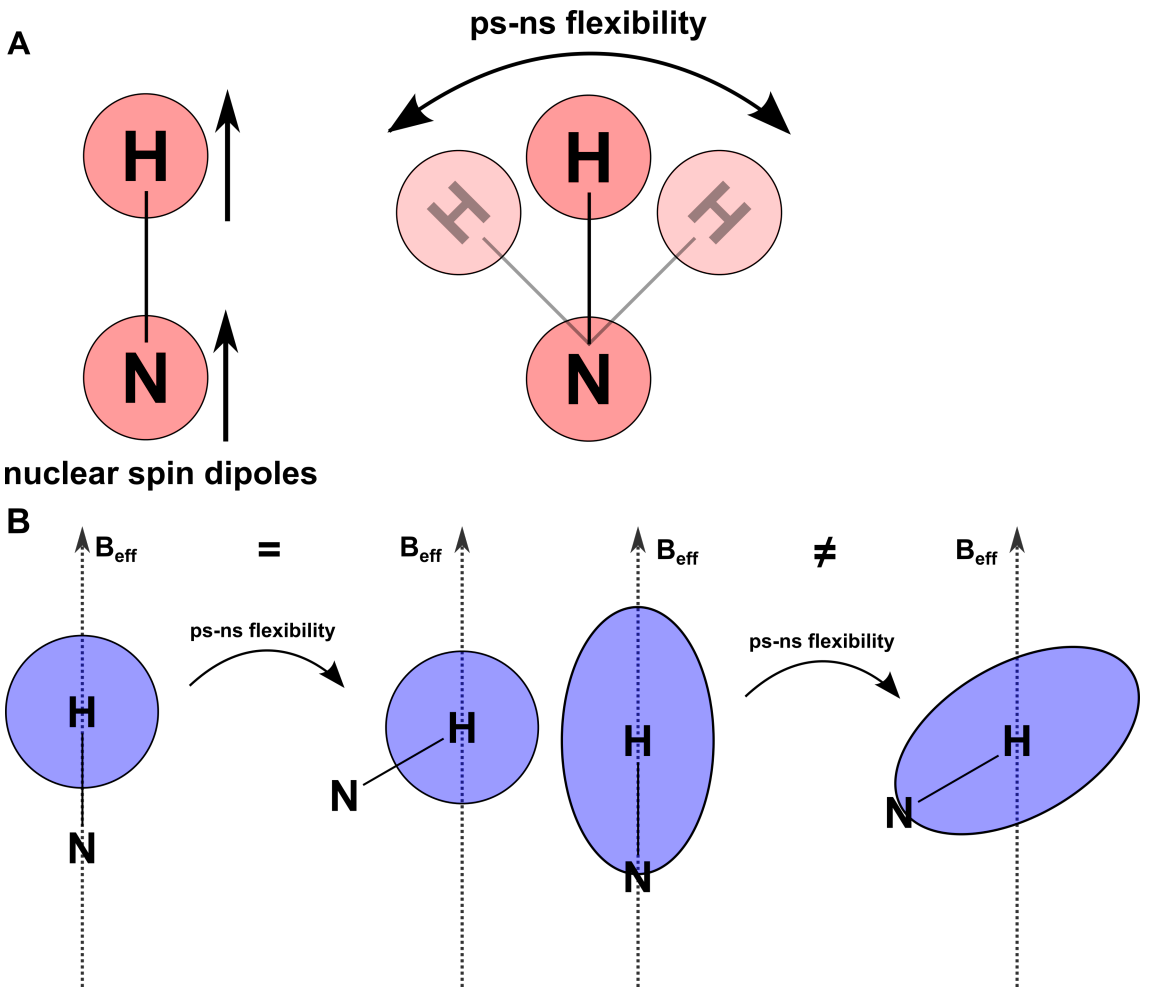

Figure 1.xii. The main two mechanisms that link bond flexibility with R_1 and R_2. (A) Two nuclear spin dipoles are aligned in a magnetic field. When one moves relative to the other, it creates its own magnetic field and will aid the relaxation of its dipole partner. The efficiency of relaxation enhancement is dependent on r^{-6} and the gyromagnetic ratio of the nuclei. (B) Two nuclei are shown with spherical (left) and ellipsoidal (right) electronic orbitals. When both are subjected to an outside magnetic field (B_0) the effective magnetic field observed (B_{eff}) is affected by the shielding of the electrons in the orbitals. Electron density in molecules is almost never symmetrical, hence when nuclei move relative to one another there is a change in the B_{eff}. This gives rise to an anisotropy in the chemical shifts, which fluctuates with fast bond flexibility enhancing relaxation mechanisms.

1. **Dipolar Coupling (DC)** (**Fig. 1.xii.a**) is the through-space interaction between two nuclei and their magnetic dipoles. The distance and orientation of the dipoles in space will

change with the molecular vibrations and rotations and hence create a fluctuating magnetic field in relation to B_0, which will in turn enhance the relaxation rates. The strength of the dipole and the distance between the nuclei will affect this mechanism, with distance being r^{-6} dependent and the strength being depending on the gyromagnetic ratio of the nuclei involved, hence protons one bond away are by far the most efficient relaxation enhancers.

2. **Chemical Shift Anisotropy (CSA)** (**Fig. 1.xii.b**) arises because chemical shielding of the nucleus by its electrons is dependent on its orientation with respect to the applied field B_0. CSA also affects relaxation rates, because as the molecule tumbles slowly or a bond vector rotates at ps-ns times, the electronic orbitals create a fluctuating local magnetic field for every nucleus with respect to B_0, which will change the relaxation rates.

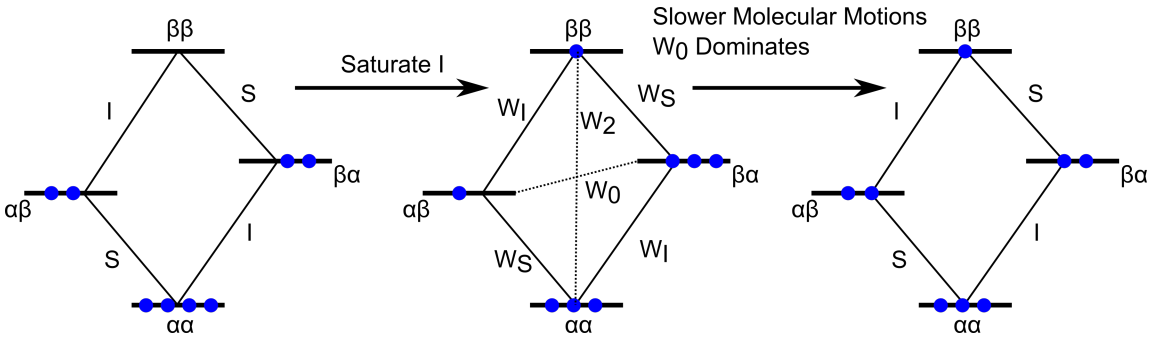

Figure 1.xiii. Quantum mechanism of the heteronuclear NOE (hNOE). Energy levels indicate the four combinations (αα, αβ, βα and ββ) which can be occupied by two nuclear spins. Blue dots indicate relative populations in each level and the **I** and **S** labels indicate which nuclei are undergoing a single quantum transition, either **I** or **S**. The relative 'signal' output is measured as the difference in populations between S single quantum states. Upon saturating **I**, the populations of **I** single quantum transitions are equalised. During this time there is a probability of spins relaxing by either single quantum transition (W_I or W_S), or by double (W_2), or by zero (W_0) quantum transitions. If W_0 dominates, then the **I** transition is less numerous than before and therefore a reduction in signal is observed (in this case a reduction of 4 to 2).

Finally, the hNOE is the transfer of magnetisation via dipolar coupling in free space between different nuclei, typically ^{1}H with either ^{13}C or ^{15}N nuclei. The hNOE is another effect induced by quantum nuclear spin. It too depends on macromolecular molecular motion and the fast motions of bonds. Consider the energy states of two I = 1/2 nuclei **I** and **S** in a magnetic field, there are 4 possibilities of α and β states (αα, αβ, βα and ββ) the transitions of which are related to the gyromagnetic ratio of each nuclei (**Fig. 1.xiii**). If nuclei **I** was saturated, the normal single quantum transitions of **I** would be equalized and no relaxation would take place. However what happens is either a double quantum (flip-flip) or a zero quantum (flip-flop) relaxation of both spins, the probability of each one (W_2 or W_0) determining whether an overall loss (W_0) or gain (W_2) in signal occurs. In large, slow macromolecules the probability W_0 dominates and the degree of W_0 compared to W_2 depends on the speed at which the spin pair are rotating in

solution. Thus, the hNOE provides yet another way of linking an observable measurement to the dynamics of the system.

1.4c Model Free Analysis

Once R_1, R_2 and hNOE measurements have been obtained, they are then linked mathematically to the motions of the molecule by a correlation function or C(t). This is a common way of analysing motions of macromolecules tumbling or even bonds vibrating randomly in a protein. The function gives a representation of the time it takes for the orientation of the molecule or bond vector to become randomised. For an ensemble measurement of one bond for many proteins in solution, the C(t) is assumed to be an exponential, with the time constant, or correlation time (τ_C) being approximately the time it takes a bond to rotate by one radian. A Fourier transformation of C(t) produces a Lorentzian distribution called a spectral density function J(ω), showing the proportion of time a bond is moving at a frequency (ω). This effectively informs what proportion of spins in an ensemble are moving close to the Larmor frequency, and hence how fast they relax.

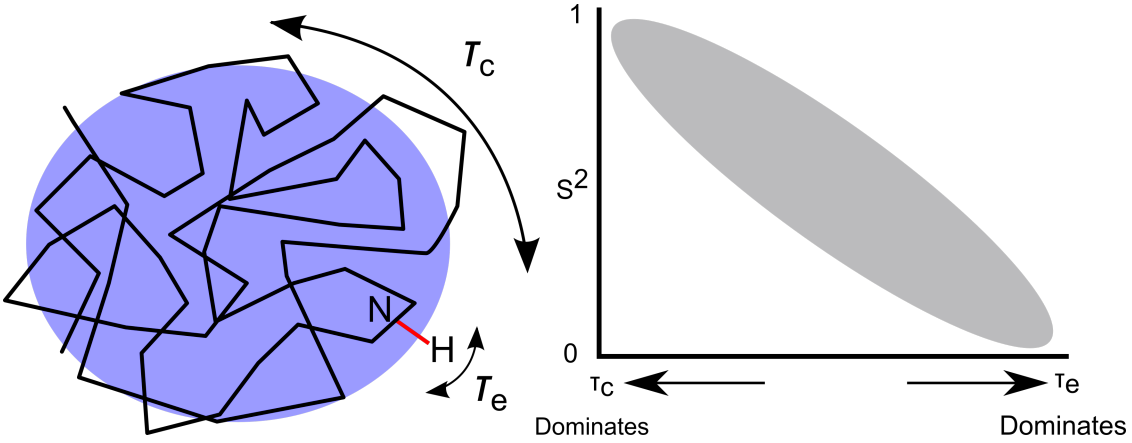

Figure 1.xiv. The principle of Model-Free analysis, and the link between fast bond motions and the S^2 parameter. A cartoon representation of a protein tumbling in solution (left), with a correlation time τ_C illustrating the time for the protein to tumble by roughly one radian and can be interpreted well by R_1 and R_2. According to model-free analysis the very fast motion of a single N-H bond is completely independent to the tumbling of the molecule. These fast motions will enhance the NMR observables R_1 R_2 and hNOE by changing the spectral density function. Thus, when modeled by model-free formalism there will be a difference between an effective correlation time (τ_e) and the rotational correlation time of the molecule (τ_C). The right chart shows S^2 values, and how much the bond is tumbling in line with the overall motions of the molecule (high S^2) or *vice versa*.

The R_1, R_2 and hNOE values can be used to calculate the spectral density function and hence quantify τ_C. Using the model free approach, it is possible to deconvolute the contribution to relaxation from molecular tumbling (τ_C), and from much faster bond motions. The analysis models J(ω) as a combination of τ_C and an effective correlation time, τ_e for fast internal motions. Model free analysis assumes that the motions of individual residues are much faster and

independent of the much slower molecular tumbling of the protein. It also has *(as its name suggests)* no model for the motions; it simply quantifies the motions present. The most useful parameter that is obtained from the model is the S^2 or order parameter, which takes a value between 1 and 0 (**Fig. 1.xiv**). A residue with a perfect match of τ_e with the τ_C shows absolute rigidity, and has an S^2 value of 1. Conversely, a residue with significantly different values for τ_e and τ_C shows absolute flexibility, and has an S^2 value of 0. Order parameters are not a rate, and confusingly are not related to entropy. They have no units; they simply give a general number for the ps-ns timescale motions around each residue. Although S^2 cannot be compared to actual rates, the value is still incredibly useful. It has been shown empirically that for backbone amide ^{15}N sites, rigid structures like α-helices and β-sheets have S^2 values ≥ 0.8, whereas flexible loops, turns and termini have S^2 values between 0.5 and 0.8.[52]

1.4d Exchange and Relaxation Dispersion.

The order parameter provides a good indication whether a residue is more flexible than others or more flexible than its environment would suggest, and can be used to direct further experiments that can quantify rates of exchange (k_{ex}). Consider a nucleus exchanging between two states, where the chemical shift of each state is different. If k_{ex} is much slower than the difference in chemical shift (Δδ in Hz), then two distinct peaks are present with peak intensities reflecting the relative population of each state, this is known as slow exchange. Conversely, if k_{ex} is much faster than Δδ, the spectrum will display an average chemical shift at the weighted average of populations between the two exchanging states; this is known as fast exchange. As the k_{ex} approaches equality with Δδ, the peak broadens out, usually disappearing into the baseline; this is known as intermediate exchange. As mentioned above, the mechanism of R_2 relaxation can be significantly affected by exchange broadening, and several methods have been designed to determine the added effect of exchange on R_2.

One method is to deploy Relaxation Dispersion (RD) experiments that measure the exchange rates (k_{ex}) between two different transiently populated states P_A and P_B, where the minor species P_B has a population of 1-10 %. There are two widely used methods, the first being Carr-Purcell Meiboom-Gill Relaxation Dispersion (CPMG RD). This is used to quantify intermediate/fast exchange processes in the 0.3-10 ms time range.[53] This timescale is relevant to a lot of domain motions and secondary structure changes within proteins; these can influence ligand binding or release and hence in some cases catalytic turnover (See **section 1.5d**: RNase A). Any R_2 relaxation combined with an exchange process (R_{ex}) will decrease the

signal received at acquisition since, $R_{2obv} = R_2^0 + R_{ex}$. In brief, the CPMG pulse sequence refocuses all spins irrespective of chemical shift by using a spin-echo τ-180_{-x}-τ pulse sequence (**Fig. 1.xv**). In a series of 2D experiments with different values of τ, the contribution of the exchange process on R_2 relaxation is decreased by shortening the τ period between pulses (equal to an increase in CPMG pulse frequency or ν_{CPMG}). The increase in ν_{CPMG} means that there is less time for an exchange event to occur during the spin-echo pulse sequence; hence the observed R_2 relaxation is less affected by the exchange process. The observed effective R_2^{eff} is plotted versus ν_{CPMG} and by using fitting models a range of parameters describing the exchange process can be derived from the dispersion profiles (k_{ex}, $|\Delta\delta|$, P_A and R_2^0).

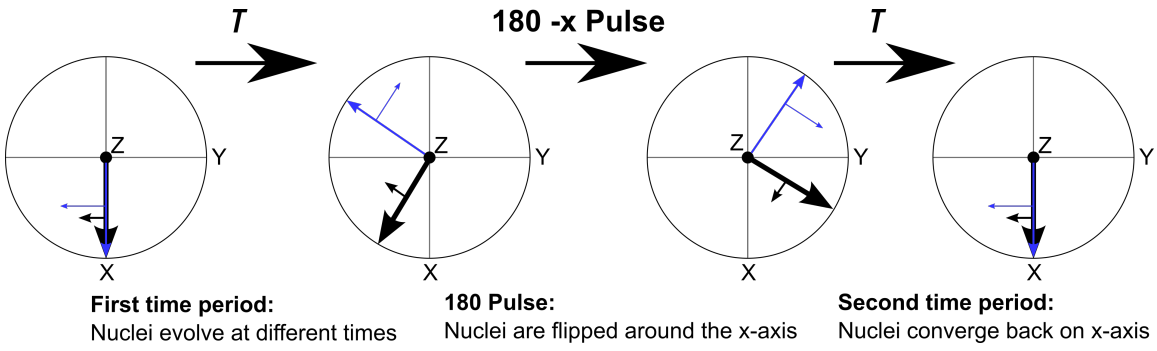

Figure 1.xv. The spin echo pulse sequence for CPMG. The τ-$180_{-x,}$-τ refocuses the magnetisation back to +x irrespective of chemical shift. Even though the chemical shifts of A and B are different and hence they evolve with different times in τ, the fact that they have been 'mirrored' with the 180 pulse should refocus them to the same point after another τ time period. The only reason the signal at x after this pulse sequence would be less than expected is due to an exchange process happening, thus changing the chemical shift in one τ period but not in the other resulting in a loss of signal coherence. Adapted from (Kleckner and Foster 2011).[51]

A second method used to derive information from exchange processes is the rotating frame RD (RT RD), which is applied to faster exchange events with timescales of around 20-100 µs occurring in the intermediate exchange regime. In this method, the magnetisation is 'locked' in place by a radio frequency soft pulse at a tilted angle to the magnetic field B_0; the relaxation is a combination of R_1 and R_2 relaxation rates, designated $R_{1\rho}$.[54] If exchange processes are occurring, the nucleus spends more time in different chemical shift environments during the spin-lock period resulting in a faster relaxation or a larger $R_{1\rho}$. By increasing the effective field strength of the spin-lock pulse it is possible to reduce the frequency that an exchange event can take place. The remaining analysis is very similar to that of CPMG RD, with R_2^{eff} being plotted versus the effective field strength to determine the timescale of the exchange process (k_{ex}) by a subsequent model fitting.

RD is a powerful technique that allows analysis of exchange processes taking place at a wide variety of timescales. NSR and the subsequent model free analysis are equally powerful, allowing for an overview of flexibility in proteins. Both these approaches allow for dynamic measurements at residue resolution. The information gained can shed light on the dynamics of protein systems, which have implications for understanding their catalytic mechanism. Presented below is the NMR spectroscopic work done on hFEN1 prior to my work, and two examples of proteins that have been extensively studied using NMR spectroscopy, which highlight some of the strengths and weaknesses of the NMR techniques. The information from the examples will be applied to novel studies involving hFEN1.

1.5 Studies using NMR spectroscopy on hFEN1 and case studies of NMR techniques applied to dynamic proteins.

1.5a Protein stability and spectral quality optimisation of the hFEN1 protein

The hFEN1 protein without the extended C-terminus is around 38 kDa. Experiments done by Dr. David Finger have shown that trying to keep it soluble above concentrations of 100 µM is difficult at room temperature. The problem with collecting viable NMR spectroscopic data on the protein is therefore is twofold. First, the large molecular weight causes a fast R_2 relaxation rate, meaning less information is gathered per FID, resulting in a decrease in the signal-to-noise ratio (S:N) and an increase in signal broadening. Second, the protein precipitates during pulse sequences causing the uniformity of the sample to deteriorate thus ruining any magnet shimming and causing large linewidths on peaks.

To solve the first problem of large molecular weight, a specific NMR technique can be used: Transverse Relaxation-Optimised SpectroscopY (TROSY). The TROSY pulse sequence was first reported 20 years ago and was shown to decrease the line width of signals dramatically.[55] It works by relying on the destructive interference caused by CSA and dipole-dipole coupling in one of the four components in the signals of a heteronuclear correlation spectrum. Rather than using decoupling methods to merge all four components into one, ¾ of the signal is sacrificed in order to use the very slowly relaxing doubly destructive peak only (**Fig. 1.xvi**). This results in very sharp peaks, which in this case the signal lost is overcome by the reduction of line broadening. With the TROSY pulse program it is possible to achieve meaningful results on larger biomacromolecules (> 28 kDa). For hFEN1 this is ideal and the initial experiments done by Dr. David Finger reported TROSY pulse sequences improving spectral quality significantly.

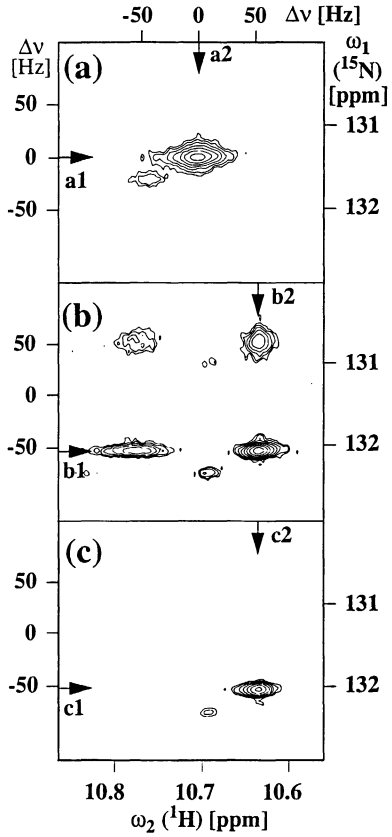

Figure 1.xvi. Transverse Optimised SpectroscopY (TROSY) spectra (A) The decoupled heteronuclear correlation spectrum for one system, there is only one broad peak. (B) The four components associated with the four transitions of the heteronuclear correlation, the bottom right peak being the destructive coherent peak and the top left being the constructive coherent peak. (C) The TROSY spectra. The other three peaks are silenced from the spectra, leaving only the much sharper doubly destructive peak, although much less signal is obtained the sharpness of the peak allows for a much greater S:N. Taken from the patent for TROSY by Pervushin *et. al.*[56]

The second issue of the protein, precipitation, proved far more problematic. A range of sample conditions was experimented by Dr. Finger to improve protein stability. The initial conditions for a ^{15}N-labelled hFEN1 protein sample were: 50 mM sodium phosphate, 1 mM DTT and 0.1 mM EDTA. Temperature, pH, mono- and di-valent ion concentrations were screened for conditions that afforded an optimal ^{1}H-^{15}N TROSY spectrum. Although hFEN1 was initially stable when concentrated to 500 µM in the phosphate buffer, protein precipitation was observed after 12 hours at 4 °C and could not be prevented until the solution was below 100 µM. Therefore the aim was to screen for sample conditions that afforded optimal spectra while minimizing protein precipitation. It was found the temperature that gave the best spectrum with minimal loss of protein due to precipitation was 25 °C; lower temperatures presented broader peaks likely due to slower molecular tumbling, whereas temperatures above 25 °C greatly expedited protein precipitation. The solution pH was adjusted between 6.0 to 8.5. Lowering the pH resulted in an increase in peak broadening, sample turbidity and increased precipitation. This was not surprising as the theoretical pI of the hFEN1 construct is pH 6.0. Although pH 8.5 resulted in the

slowest rate of precipitation, the number of observable peaks was decreased considerably due to the increased rate of amide proton exchange. Thus, pH 7.5 was deemed optimal as it provided the most number of peaks and a moderate rate of precipitation.

1.5b Assignment of hFEN1 protein

The first step towards analysing backbone dynamics is to assign as many amide moieties as possible in a protein. The conditions described above were suitable for an assignment of triply labeled protein with ^{15}N, ^{13}C and ^{2}H. Dr. David Finger and Dr. Andrea Hounslow employed six TROSY triple-resonance experiments: HNCO, HNCA, HNCACB, HN(CA)CO, HN(CO)CA and HN(CO)CACB. Each experiment transfers magnetisation from the amide proton to the nitrogen, and then to a specific carbon on the peptide chain, and then transfers it back to the proton for detection (**Fig. 1.xvii**). Therefore chemical shifts of Cα, Cβ or CO are obtained and they are all correlated to a specific amide. The assignment data was processed on the analysis software *felix* and was confirmed using the *asstools* assignment program.[57] The hFEN1 protein assignment was able to map chemical shifts of backbone amides for 70% of all theoretically observable backbone N-H groups (uploaded to the BioMagResBank under the BMRB accession code 27160). The assignment was lacking in regions of interest, for example at the bottom of α4 and residues known to be in the active site (**Fig. 1.xviii**).

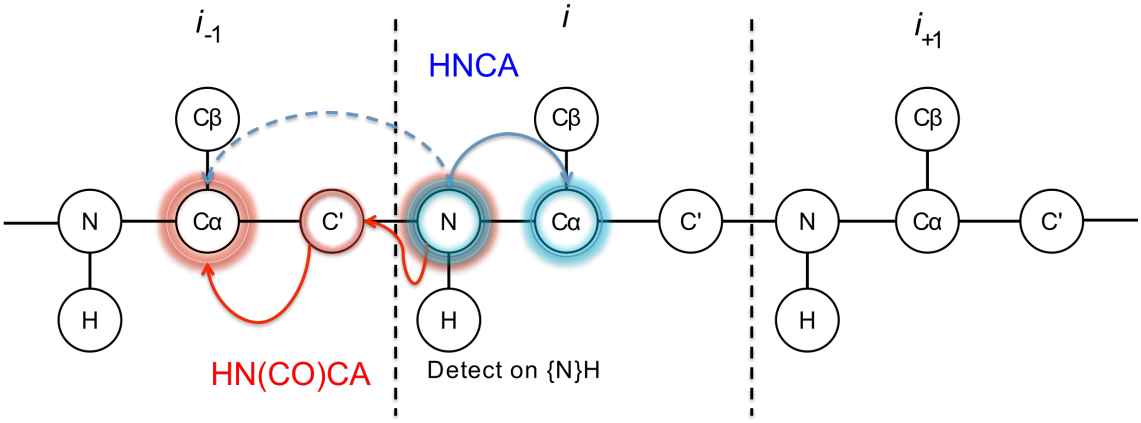

Figure 1.xvii. A resonance transfer diagram for Cα atoms. Arrows represent resonance transfer; the reverse resonance transfer for detection goes in the opposite direction. The directed experiment (Red, HN(CO)CA) first directs the magnetisation to *i-1*'s carbonyl carbon (C') and then subsequently to the *i-1*'s Cα, hence triple resonance only happens for the previous carbon. The undirected (Blue, HNCA) directs the magnetisation to Cα straight away, there is a major peak present at *i*'s Cα, and a minor peak sometimes present at *i-1*'s peak.

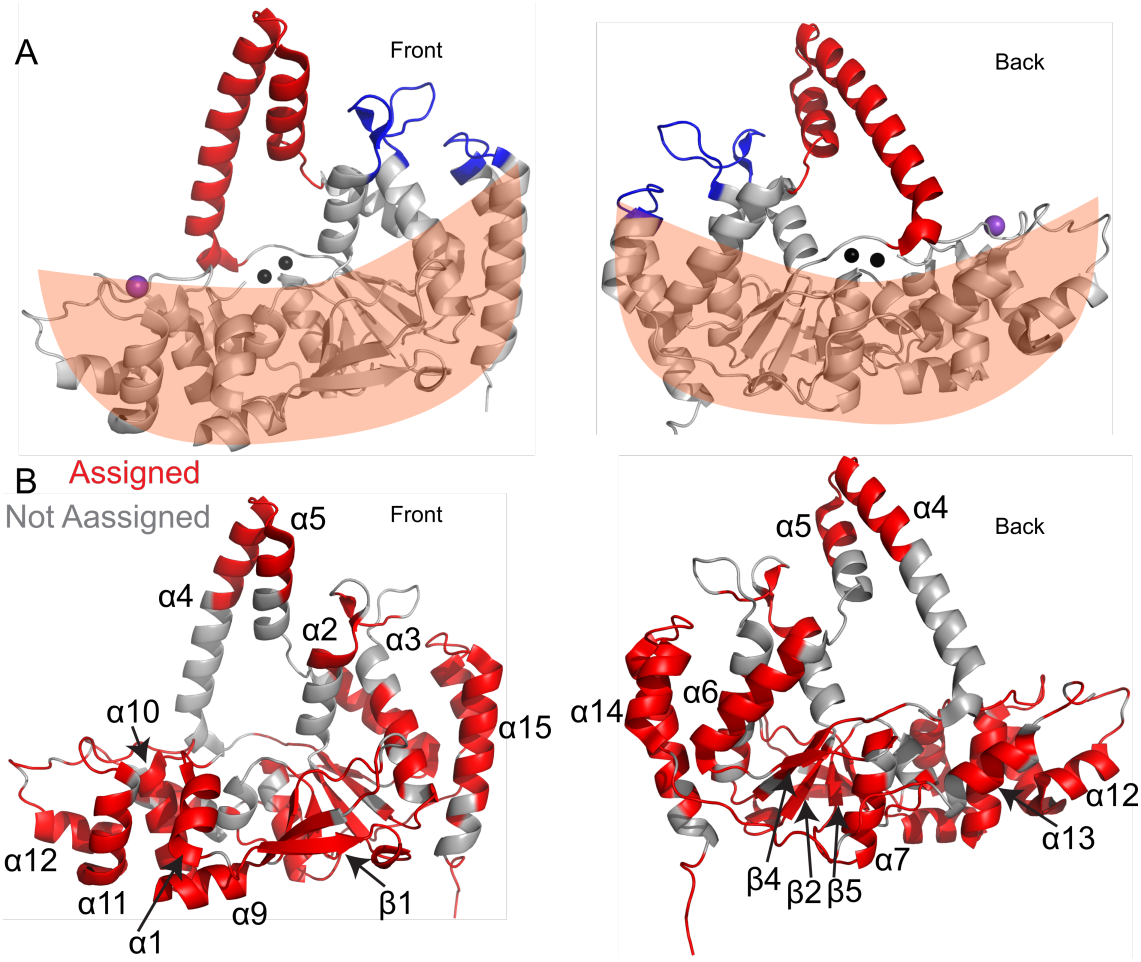

Figure 1.xviii. Cartoon representation of the hFEN1 molecule to illustrate location of assigned backbone amides. (A) A cartoon showing the 3Q8K structure with important regions highlighted, the arch (red), the α2-α3 loop (blue), the saddle region (orange saddle), the positions of the two catalytic metal ions, i.e. the active site (black spheres) and the potassium ion in the H2tH motif (purple sphere). (B) The assigned (red) and unassigned (grey) residues are coloured onto the product crystal structure 3Q8K. Only residues with chemical shifts assigned to a backbone amide are considered assigned. Secondary structure elements are labeled for clarity.

1.5c Protein stability optimisation of buffer, monovalent and divalent salt conditions

The protein in its current state was not soluble for long timescales. Therefore, differing conditions of buffer, monovalent salts and divalent salts were screened for. Dr. Finger experimented with different buffer molecules (Tris, HEPES, potassium phosphate), but they did not make a noticeable difference to the NMR spectra. Therefore HEPES was used for further experiments, as its pKa was closest to the optimal pH of the protein (pH 7.5). Because hFEN1 requires divalent metal ions (Mg^{2+}) for catalysis, Mg^{2+} was added to the protein sample to investigate changes in the spectrum. Small changes were observed in the spectra at concentrations of 8 mM $MgCl_2$ (**Fig. 1.xix**). Most notably changes were observed in the N-terminus of the protein, which is consistent with the N-terminus of hFEN1 pointing towards the active site. Other chemical shift differences were noted in residues next to active site

carboxylates (for example S157 and L190). Furthermore, other diamagnetic multi-valent cations (Ca^{2+}, Zn^{2+} and Lu^{3+}) were also tried. Ca^{2+} afforded spectra similar to the Mg^{2+} (**Fig. 1.xix**), whereas Zn^{2+} resulted in precipitation and Lu^{3+} resulted in poor linewidths. The crystal structure of hFEN1 in complex with DNA (3Q8K) shows that in addition to divalent cation ion binding sites, the protein also has a potassium ion binding site. Titration of KCl into the solution resulted in minor improvements in the spectra with 76 mM resulting in the optimal spectrum. The addition of mono- and di-valent cations did not prevent the slow precipitation.

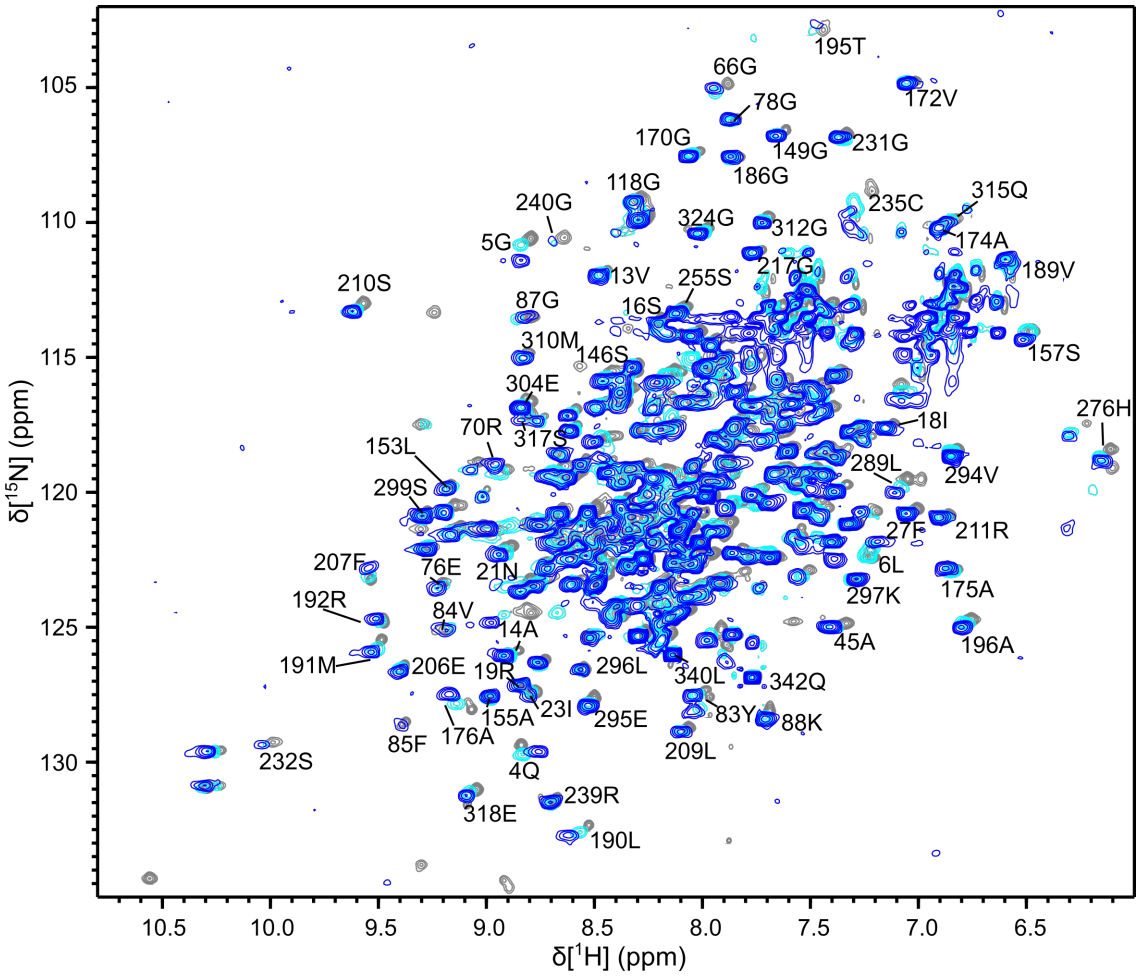

Figure 1.xix. ^{1}H-^{15}N TROSY spectrum of hFEN1 with different divalent cations added. hFEN1 protein in 0.1 mM EDTA (grey), 8 mM $CaCl_2$ (blue) and 8 mM $MgCl_2$ (cyan). Some dispersed residues are labeled, of particular note are 4Q and 5G, which change chemical shifts upon addition of Mg^{2+} ions, and show chemical shift differences between Ca^{2+} ions or Mg^{2+} ions. Some peaks are missing due to poor linewidths in non-perdeuterated samples. Only the EDTA spectrum is from perdeuterated sample.

The hFEN1 construct and its sample conditions set up by Dr. Finger are suitable for an assignment of hFEN1. However the longer timescales needed for NMR relaxation measurements means that sample stability has to be solved before moving onto these experiments. However, the two examples below highlight what sort of analyses could be performed upon protein systems that are soluble for extended periods.

1.5d Example 1. Ribonuclease A: The model dynamic protein system.

Pancreatic ribonuclease, or RNase A has around 300 X-ray crystallographic, neutron diffraction and NMR resolved structures present in the Protein Data Bank due to its high stability and ease of purification. RNase A activity and catalytic cycle have also been studied in great detail, making it an excellent case study of the use of NMR dynamics.[58] In terms of activity, RNase A catalyses transesterification and then hydrolysis of RNA without any metal ions or cofactors. It has been shown that the enzyme has a shallow 'U' shape with the active site of the enzyme lying in the middle of the cleft (**Fig. 1.xx**). Key residues in the active site include H12 and H119, which act as general bases and K41, which acts as a general acid. All these play an important role in the catalytic step by using general acid-base catalysis to hydrolyse the phosphodiester bond.

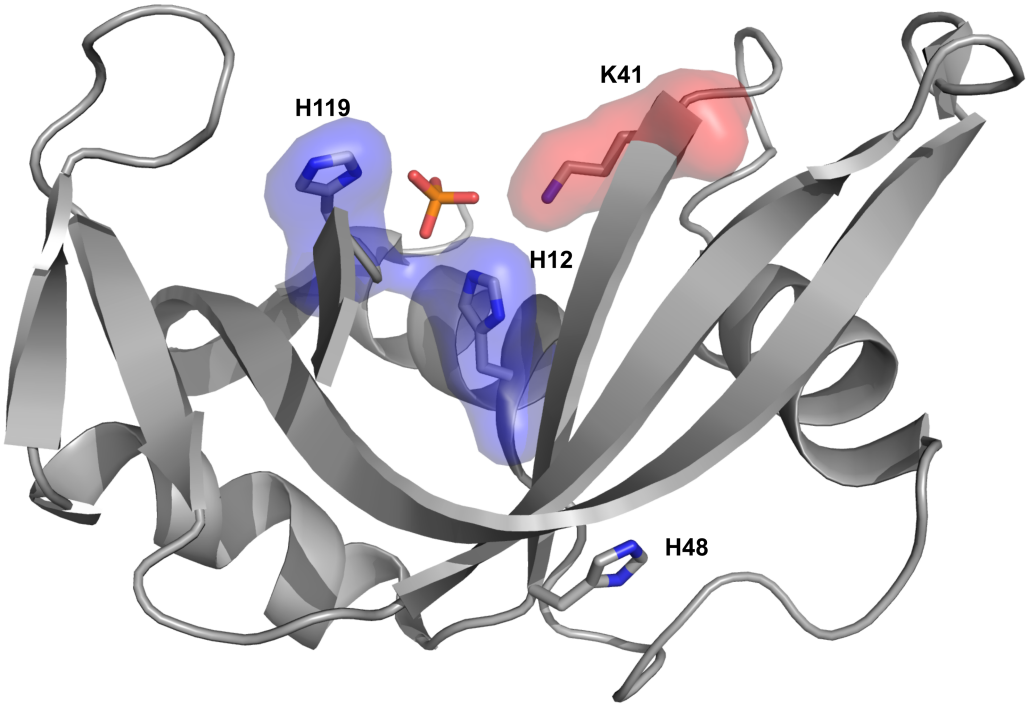

Figure 1.xx. X-ray crystal structure of RNase A (5RSA).[59] The overall structure of the protein is seen as a 'U' shape with a wedge in the middle, where the substrate would pass through. This crystal has a phosphate present in the active site. Active site bases H12 and H119 are present in blue, while the active site acid K41 is seen in red. H48 is far away from the active site (around 18 Å).

Kinetic experiments showed that product release is the rate-limiting step in catalysis, with a k_{cat} of around 1700 s^{-1}. Furthermore, the second-order rate constants approach the value expected for diffusion. So the motion of the protein and the conformational change could have something to do with catalytic function.[60] Various NMR and X-ray structures showed that

residues in the active site sampled different conformers, in particular H119, and it was thought that when a ligand was bound it initiated this conformational change. However, it was subsequently reported that the apo-enzyme was apparently in a constant shift between two distinct states - A and B - confirmed by other NOE-based structural studies.[61,62] Examination by model free analysis showed that certain key active site residues were in fact moving at a ms-to-μs timescale, hence showing inherent flexibility in these residues.[63] To quantify the rate of exchange, CPMG RD NMR experiments showed that the 16 quantifiable residues were moving on average R_{ex} 1640 s^{-1}. Interestingly this rate is almost identical to the catalytic turnover rate (k_{cat}) and that of product release (k_{off}) as measured by line shape analysis.[64] It was implied that these dynamics played a role in product release and therefore the catalytic cycle, but further studies with unreactive ligand substitutes revealed more information.

The result of looking at enzyme (E), unreactive enzyme substrate (ES) and unreactive enzyme product (EP) (**Fig. 1.xxi**) complexes is that, while binding with the substrate or product the structure of the enzyme changes subtly. The main structural difference is seen for H119, which changes conformation from being hydrogen bonded to D121 in the ES complex to solvent exposed in the EP complex. These structures were time-averaged structures by solution state NMR and X-ray crystallography, which would suggest that there is some kind of conformational change occurring. In order to ascertain the dynamics between these lowly populated transient states additional RD NMR was undertaken on these complexes.[64] It was found that there is no change between the dynamics of the protein in E, ES or EP. This means that the dynamics are independent of ligand. Furthermore the states are being sampled at a rate equivalent to k_{cat}, which implies strongly that the enzyme motion is linked to the catalytic turnover. This information, coupled with the X-ray structures indicates that RNase A will open up bind a substrate and release a product all on the timescale of this exchange process.

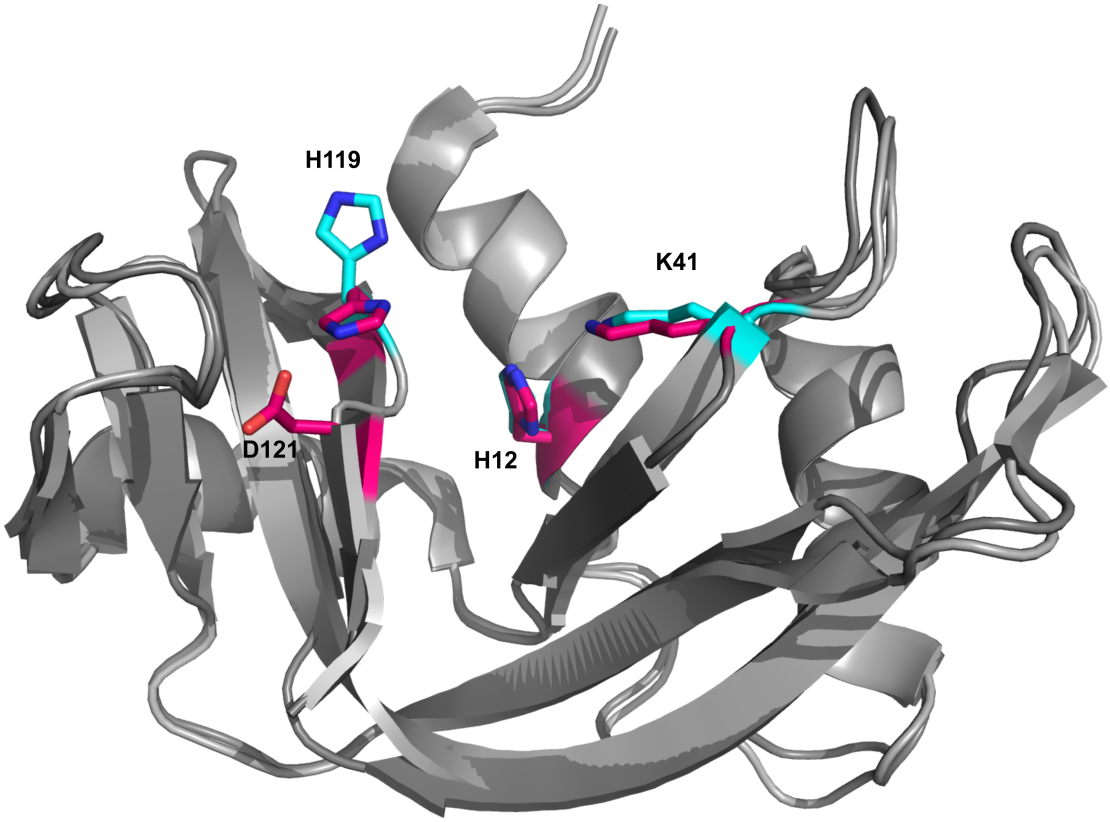

Figure 1.xxi. RNase A structures with product or substrate bound. Highlighted residues from each structure show the main differences between the substrate bound (ES) (1U1B, magenta)[64] and product bound (EP) (1RPF, cyan)[65] structures. The product analogue used was a 3' cytidine monophosphate, while the substrate analogue was an adenine pyrophosphate thiamine dinucleotide (pTp*p*Ap). H119 shows the greatest difference when bound to the substrate, forming a hydrogen bond with D121 indicating a subtle change in structure.

In addition to the many analyses described above, mutagenesis experiments were carried out on the protein to link catalytic rates with k_{off} values as determined by NMR. Residues mutated in the active site were found to increase the k_{off} values of the product considerably which was again linked with a decrease in catalytic turnover.[66] H48, which is a highly conserved residue, was mutated to an alanine and a subsequent decrease in motion and k_{cat} values was noted despite it being 18 Å away from the active site (**Fig. 1.xx**).[67] H48 clearly has some kind of link to the speed of opening and closing of the enzyme suggesting an allosteric effect present.

Overall in this example the NMR dynamic measurements proved not to be a definitive tool to infer catalytic activity when used on their own. Rather NMR was used in conjunction with a plethora of biochemical and biophysical methods in order to provide new insights into the function of the protein. Not only were conformational exchange rates acquired for the opening and shutting of RNase A but also exchange rates for individual residues were obtained. This

allowed a great deal of insight to be gained in the catalytic cycle of the protein, including if any allosteric effects were present.

1.5e Example 2. Triosephosphate Isomerase (TIM): A large molecular weight example.

In some regards RNase A can be seen as an ideal case study for NMR and biochemical techniques coming together to ascertain catalytic activity. As mentioned before, the protein is small, very stable and easy to produce and purify, which is ideal for solution NMR experiments. NMR experiments designed to probe dynamics are typically run for extended time periods at room temperature with relatively high concentrations of protein (0.1 to 1 mM) and with a total of about 500 µL of sample required. This is unsuitable for the main protein of interest hFEN1, which in its current state is not very soluble over extended timescales. The second example, the 53 KDa protein Triosephosphate isomerase (TIM) will highlight how larger proteins can be optimised.

Because of its size, TIM has a very large rotational correlation time of 30 ns at 293 K.[45] The protein is therefore a good example of how TROSY pulse sequences can be used to measure relaxation properties. TROSY also allows unique relaxation experiments, which can be applied to the TIM enzyme, in particular the off-resonance $R_{1\rho}$ experiment, which can be used to observe very fast motions up to rates of 10^5 s^{-1}.[68,69] TIM is the enzyme responsible for isomerising the two products of aldolase in glycolysis, D-glyceraldehyde 3-phosphate (GAP) and dihydroxyacetone phosphate (DHAP).[70] Since only GAP can continue along the glycolytic pathway it is essential for DHAP to be isomerised to GAP in order for 100% of the energy from one glucose molecule to be obtained. The activity of TIM is linked heavily to a Ω-loop structure (loop 6) present near the active site (**Fig. 1.xxii.a**).[39] The loop can move between an open and closed conformation (**Fig. 1.xxii.b,c**). Looking at various structures and computational models, when in the open state, the substrate is free to bind and the product is free to dissociate.[70] When the loop is in the closed state, it conveniently borders the active site allowing for catalysis to occur.[71,72] Furthermore a key residue in catalysis (E165) has been found to be aligned perfectly for catalysis in the closed conformation.[73] The motion between these two states is what is thought to provide the catalytic activity to the protein.[74] So studying the dynamics between the two states should provide a good insight into the catalytic rate.

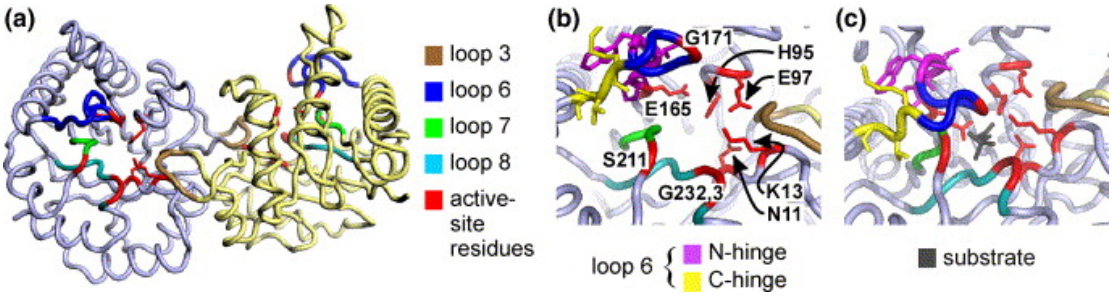

Figure 1.xxii. Illustration of the dynamics of an Ω-loop in and out of the active site. A) The full structure of the TIM dimer, with loops and active-site residues colour labeled. The important loop is loop 6, which can be seen to move in and out of the active site in the free B) enzyme and the substrate bound enzyme C), indicating conformational changes in this region of the protein. These changes have clear implications on catalysis as the labeled key residues G171, E165 are placed into the active site upon substrate binding. Diagram taken from (Kempf *et. al.* 2007).[75]

The closed state is stabilised by many hydrogen bonds between residues in loop 6 and neighboring loop 7. One such interaction thought to be important is the hydrogen bond from the amide of A176 and the O^{η} of Y208. When Y208 is mutated to a hydrophobic phenylalanine it results in a 2400-times decrease in activity (k_{cat}/K_M), suggesting that having a H-bond stabilised closed state impacts on catalytic efficiency.[76] NMR $R_{1\rho}$ RD experiments were done on WT protein residues in loop 6 and an exchange time of roughly 100 µs (or a R_{ex} of 8900 s^{-1}) was observed.[77] Motions in the C and N-termini were found to be very similar, suggesting a global motion. In the mutant Y208F, the same $R_{1\rho}$ experiments were carried out, and site-specific information was gathered on the dynamics in the loop. Interestingly the exchange rate of the N-terminus of the loop in Y208F is very similar to that of the WT protein (around 10 000 s^{-1}). The main difference is in the C-terminus which has flat relaxation curves, suggesting that the population of the open state is stabilised so much that no exchange, and therefore no dynamics are occuring. The NMR data shows that one side of the loop is disrupted more than the other by the loss of the one hydrogen bond.

1.6 Measurement of potential dynamics of hFEN1

1.6a Application of NMR relaxation experiments to the hFEN1 protein system

As is shown in the examples, NMR methods can be applied to larger proteins, and to motions which are on the timescale of 1 ms to 100 µs. Crucially with NMR experiments, they are amino acid residue specific, allowing a great deal of information to be gathered on one protein system. Using NMR measurements to assess dynamics with hFEN1, what questions can be addressed? In hFEN1 the proposed model shows that a slow conformational process plays an important part in catalysis and potentially allows for specificity for the ideal substrate. This conformational

change could be attributed to a variety of structural changes both for the protein and the DNA molecule. In the protein, the arch becomes ordered or structured around the 5' flap, and the 3' flap-binding pocket engages with a 1nt flap. In the DNA, there is bending at the nick and positioning of the 5' flap into the active site. There is speculation to how all these processes act to perform catalysis and at what rate they all happen.

NMR relaxation measurements and subsequent model-free analysis can ascertain S^2 parameters. This will reveal unusually fast-moving residues within hFEN1 and can direct further dynamic studies. This would be applicable to the arch region, as X-ray crystallography snapshots have shown that the arch exhibits a wide variety of conformations in the absence of DNA. Therefore ascertaining the true dynamic properties of the arch in solution, as well as other regions such as the 3' flap-binding pocket would confirm how flexible these regions are. The model-free analysis should also highlight any hitherto unknown regions of flexibility. Finally, the analysis can also be applied to the DNA, allowing for very fast (ns-ps) flexibilities in a nicked substrate to be ascertained, to see if the DNA exhibits fast sampling of a bent conformation.

Model-free analysis can also yield R_{ex} terms, which indicate slow millisecond timescales of motion. A more thorough analysis using RD through CPMG or $R_{1\rho}$ would allow for a classification of these slow motions, and equally would be applicable to the arch. If the arch has to undergo a conformational change when bound to substrate, it would be interesting to know if the arch is already sampling this conformation in the absence of substrate, like in the example of RNase A. Or, equally likely, whether the arch simply adopts this closed conformation when bound to DNA. RD analysis could also reveal other residues moving at similar timescales highlighting slow flexibilities in other parts of the protein, with potential allosteric effects, like in the 3' flap-binding pocket or further afield in the H2tH motif.

The dynamics of residues can also be analysed for substrate and product complexes. As with RNase A, the presence of a substrate/product might not change the dynamics of the underlying conformational change associated with catalysis. Or perhaps there will be a change in dynamics, as in the example with TIM, with implications for catalysis. Furthermore other essential co-factors for the enzyme might change dynamics or conformation upon binding. For example, divalent cations, such as Mg^{2+} and Ca^{2+} play an important role in the conformational change of the DNA as seen from ECCD studies, and in the complete absence of these ions there is no hydrolysis. As has already been seen in 1H-15N TROSY spectra, the presence of these ions does perturb the structure of the protein, but not significantly. Therefore it would be interesting to

see if these ions have an effect on the architecture and mobility of the protein in the presence and absence of a DNA substrate.

1.6b Aims and hypothesis

Overall, the first aim of the study would be to confirm by biochemical kinetic studies the conformational change as the main rate-limiting factor in single-turnover rates. Next would be to quantify the flexibility of all backbone amides of hFEN1, to highlight regions of interest. Of particular interest would be to know the flexibility of the 3' binding-pocket and the arch in a solution state, for free enzyme, for a bound enzyme-substrate/product complex and with or without divalent cations. Furthermore, if exchange rates can be ascertained for certain residues that match up to a biochemical rate constant (in hFEN1's case, the single turnover rate k_{ST}) then a rate-limiting conformational change can be invoked. In this way not only can a conformational change be confirmed or disproved, but the site(s) of motion can also be elucidated.

The conformational change can be categorized into two archetypical mechanisms: conformational selection, and induced fit. In a conformational selection model, the protein (or substrate) already exhibits all the conformations necessary to achieve catalysis (i.e. RNase A). Whereas the induced fit model proposes that the substrate induces the protein to adapt a new conformation by binding in the correct way. If the types of motion and the degree of motion can be quantified by NMR spectroscopy, it would help decide whether the protein displays conformational selection or induced fit. This, potentially could help discern how the protein selects a preferred DNA structure. The hypothesis is that, (*i*) the rate of conformational change occurs on a timescale relevant to the single turnover rate (k_{ST}). Therefore the NMR dynamic measurements will quantify and locate the region at which this conformational change occurs, (*ii*) that the conformational change occurs only when the correct substrate is presented, which thereby provides hFEN1 its specificity.

Chapter 2: Materials and Methods

2.1 Protein expression and purification

2.1a Buffers

All buffers used in expression and purification of the protein are listed in bold and their composition is listed in **Table 1**.

Buffer Name	Composition
LB Agar (1 L)	10 g Tryptone, 10 g NaCl, 10 g Yeast Extract, 10 g Agarose
SOC Media (1 L)	20 g Tryptone, 0.5 g NaCl, 5 g Yeast Extract, 0.19 g KCl, 0.95 g $MgCl_2$, 3.6 g D-Glucose
2×YT (1 L)	16 g Tryptone, 5 g NaCl, 10 g Yeast Extract
TB (1 L)	12 g Tryptone, 24 g Yeast Extract, 9.4 g K_2HPO_4 and 2.2 g KH_2PO_4 pH 7.2
^{15}N-M9 Autoinduction Media	50 mM Na_2HPO_4 and 50 mM KH_2PO_4 pH 7.5, 50 mM $^{15}NH_4Cl$, 5 mM Na_2SO_4, 2 mM $MgSO_4$, 0.5% glycerol, 0.05% D-glucose, 0.2% α-lactose, 12×BME vitamins (USBiological B0110), 1×Trace Metals (Teknova 1000x Trace Metals T1001)
^{15}N,^{2}H,^{13}C-M9 Minimal Media	50 mM Na_2HPO_4 and 25 mM KH_2PO_4 pH 7.5, 18 mM $^{15}NH_4Cl$, 1% $^{13}C_6$,$^{2}H_7$-D-glucose (U-$^{13}C_6$, 99%; 1,2,3,4,5,6,6-d_7 97-98%), 0.2 mM $CaCl_2$, 5 mM $MgSO_4$, 10 mM NaCl, 0.25×BME vitamins (USBiological B0110), 0.25×Trace Metals (Teknova 1000x Trace Metals T1001)
PBS	10 mM Na_2HPO_4 and 2 mM KH_2PO_4 pH 7.4, 137 mM NaCl, 2.7 mM KCl
Buffer A (IMAC and Lysis buffer)	20 mM Tris-HCl pH 7.4, 1 M NaCl, 5 mM imidazole, 5 mM βME
Buffer B (IMAC)	25 mM Tris-HCl pH 7.4, 0.5 M NaCl, 40 mM imidazole, 0.01% Tween-20, 5 mM βME
Buffer C (IMAC)	250 mM imidazole pH 7.2, 0.5 M NaCl, 5 mM βME
Bufer D (Anion Exchange)	20 mM Tris-HCl pH 7.4, 1 mM EDTA, 20 mM βME
Buffer E (Heparin)	25 mM Tris-HCl pH 7.5, 1 mM $CaCl_2$, 20 mM βME
Storage Buffer (after glycerol addition)	100 µM hFEN1, 50 mM HEPES-KOH pH 7.5, 100 mM KCl, 0.5 mM EDTA, 5 mM DTT, 0.02% NaN_3 (50% glycerol)

Table 1. Media and Buffers for Expression and Purification. Antibiotics for growth media were all added when required and are mentioned in the text.

2.1b Protein Transformation into *Escherichia coli* (*E. coli*) cells.

Human flap endonuclease protein (hFEN1-336) was expressed from a pET29b-hFEN1-336 plasmid construct with kanamycin and chloramphenicol resistance, made previously by Dr. Brian Chapados.[27] The plasmid encoded hFEN1-336 protein and contained a human rhinovirus (HRV) type-14 3C protease-cleavable $(His)_6$-tag located at the C-terminus of the protein. It also contained a lac-operon promoter to control expression. From now on this construct and protein will both be referred to as hFEN1. Dr. David Finger prepared the hFEN1 mutants, Y40A, P188A, K93A, C163T, C182T, C235T and C311T by site-directed mutagenesis of the pET29b-hFEN1 vector using the Stratagene quick change kit. The chosen expression vector was transformed into *Escherichia coli* BL21(DE3)-RILP cells. The cells were thawed on ice, and approximately 100 ng of vector was added to 50 µL of cells. The cells were incubated on ice for 45 minutes before being heat-shocked at 42 °C for 90 seconds and subsequently quenched on ice for 5 minutes. The transformed cells were rescued by addition of 800 µL of **SOC** media and incubated at 37 °C for one hour. The cells were pelleted, the supernatant removed and resuspended in 250 µL of **SOC** media. The cells were then spread onto **LB agar** plates supplemented with 50 $\mu g.mL^{-1}$ kanamycin and 34 $\mu g.mL^{-1}$ chloramphenicol, and then subsequently incubated for 18 hours at 37 °C.

2.1c Protein expression for kinetic analyses.

Natural abundance hFEN1 and for $hFEN1_{Y40A}$ were expressed previously in terrific broth (**TB**) containing 25 $\mu g.mL^{-1}$ kanamycin and 34 $\mu g.mL^{-1}$ chloramphenicol and induced by an IPTG methodology.[27]

2.1d ^{15}N-labelled Protein expression for ^{1}H-^{15}N HSQC experiments.

^{15}N-labelled protein was expressed using **^{15}N-M9 autoinduction media** with added 400 $\mu g.mL^{-1}$ kanamycin and 34 $\mu g.mL^{-1}$ chloramphenicol.[78] A single colony was added to 5 mL of **2×YT** media with 25 $\mu g.mL^{-1}$ kanamycin and 34 $\mu g.mL^{-1}$ chloramphenicol and shaken overnight at 37 °C. 0.5 mL of the starter culture was used to inoculate a flask with 50 mL of **2×YT media** with the same concentrations of antibiotics. This was subsequently grown overnight at 37 °C. An appropriate amount of the 50 mL **2×YT** culture was taken and added to 2 L (4 × 500 mL) of **^{15}N-M9 autoinduction media** with the same concentrations of antibiotics, in order to bring the A_{600} down to 0.05 in each flask, and this was incubated at 37 °C. After 2-4 hours or until the A_{600} reached 0.5 the incubation temperature was reduced to 18 °C, and the cells expressed by autoinduction overnight.

2.1e $^{15}N,^{13}C,^{2}H$-triple-labelled Protein expression for assignment and relaxation experiments.

$^{2}H,^{13}C,^{15}N$-labelled protein was prepared using a high cell-density procedure in combination with isopropyl-β-D-thiogalatopyranoside (IPTG) induction.[79] A single colony was chosen to inoculate 5 mL of 50 % D_2O **2×YT** media supplemented with 25 µg.mL^{-1} kanamycin and 34 µg.mL^{-1} chloramphenicol (all **2×YT** media were grown with these concentrations of antibiotics) and grown for 24 hours at 37 °C. A 100-fold dilution of this saturated 50% D_2O **2×YT** culture was then used to inoculate 50 mL of **2×YT** media in 100% D_2O to allow the cells to acclimatise for growth. This was subsequently grown for another 24 hours at 37 °C. This 50 mL starter culture was then added to a 500 mL 100 % D_2O **2×YT** flask and grown for 24 hours or until a A_{600} of 5.0 was reached. At this point, the culture was pelleted by ultracentrifugation (6000×g at 4 °C for 30 min) and the supernatant was discarded. The pellet was resuspended in 500 mL of **$^{15}N,^{2}H,^{13}C$-M9 minimal media** with 400 µg.mL^{-1} kanamycin and 34 µg.mL^{-1} chloramphenicol, and grown for 2 hours more at 37 °C to clear unlabeled metabolites. Subsequently IPTG to a final concentration of 0.1 mM was added, the temperature was reduced to 18 °C, and the culture was grown for a further 72 hours until A_{600} = 10-13.

$^{2}H,^{15}N$-labelled protein was prepared analogously except that the 1% $^{13}C_6,^{2}H_7$-D-glucose (U-$^{13}C_6$, 99%; 1,2,3,4,5,6,6-d_7 97-98%) was replaced with 1% $^{2}H_7$-D-glucose (1,2,3,4,5,6,6-d_7 97-98%).

2.1f Cell harvesting and lysing

In all cases after growth, cells were harvested by centrifugation (6000×g at 4 °C for 15 min) and the supernatant was removed. The cell pellet was resuspended in ice-cold phosphate buffered saline (**PBS**), where the volume added was roughly 1:10 of **PBS** to cell culture. The cells were pelleted again at 6000×g at 4 °C for 15 minutes and the supernatant was removed. The cells were then resuspended in the same volume of **Buffer A** containing an additional, 0.1 mg.mL^{-1} lysozyme and one EDTA-free protease inhibitor cocktail tablet (Sigma S8830). Cells were incubated on ice for at least 1 hour and were frozen at −20 °C until required.

2.1g Protein Purification

To prepare the cell lysate for purification, the suspensions were thawed in cold water. The cells were then sonicated (Sonics Vibra-Cell VCX 130) on ice at 75 % intensity for 60 cycles of pulsation for 5 seconds with 30 second cooling intervals. To each lysate 5 mL of **Buffer A**

containing 1 % v/v Tween-20 was added and the insoluble cell debris was removed by ultracentrifugation (40,000×g at 4 °C for 30 min).

All purification buffers were filtered and degassed prior to use and an ÄKTA pure chromatography system (GE Healthcare Life Sciences) was used for all column purification steps. The supernatant was applied to four columns in tandem (each one ID = 1.6 cm, bed volume = 5 mL) containing Chelating Sepharose 6 Fast Flow agarose beads (GE Healthcare Life Sciences) that had been previously charged with Co^{2+} ions according to the manufacturer's protocol and pre-equilibrated with five column volumes of **Buffer A**. The column was washed with five column volumes of **Buffer A** and five column volumes of **Buffer B**. The bound hFEN1 protein was eluted with eight column volumes of **Buffer C**. Fractions containing hFEN1 protein were identified by sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE), pooled and immediately diluted with an equal volume of **Buffer D**, which was then applied to a 5-mL HiTrap Q HP column (bed volume = 5 mL, GE Healthcare Life Sciences) pre-equilibrated with the **Buffer D**. hFEN1 protein was found in the flow through, whereas nucleic acid contamination was retained by the column which was eluted using ten column volumes of **Buffer D** containing a linear gradient of 0 to 1 M NaCl. The amount of hFEN1 protein in the flow through was estimated by measuring the absorbance at 280 nm (ε_{280} = 22,920 $M^{-1}.cm^{-1}$)[80] using a UV-Vis NanoDrop spectrophotometer (ThermoFisher Scientific). Two units of Turbo3C (HRV3C protease) (BioVision) were added for every mg of protein, and the mixture was allowed to stand overnight at 4 °C to catalyse cleavage of the $(His)_6$-tag from the hFEN1 protein. The extent of affinity tag removal was assessed by SDS-PAGE, and the solution was diluted further with an equal volume of **Buffer E**. The hFEN1 protein solution was then applied to a 20-mL HiPrep Heparin FF 16/10 column (ID = 1.6 cm, length = 10 cm, GE Healthcare Life Sciences) pre-equilibrated with two column volumes of **Buffer E** and eluted using 15 column volumes of **Buffer E** containing a linear gradient of 0 to 1 M NaCl (hFEN1 protein eluted at 350 mM NaCl). Fractions containing hFEN1 protein were pooled and concentrated to less than 10 mL using an Amicon Ultrafiltration cell with a 10-kDa MWCO polyethersulphone (PES) membrane (MerckMillipore) pressurised using nitrogen gas. The protein was exchanged into the appropriate buffer (**Table 7**) using a 53-mL HiPrep 26/10 desalting column (GE Healthcare Life Sciences). The hFEN1 protein concentration was determined using the absorbance at 280 nm as described above. Unlabelled samples for kinetic measurements were exchanged into 2×**Storage Buffer**, and adjusted to a concentration of 200 µM. An equal volume of glycerol was added to each sample to allow optimal storage conditions. Samples were stored at −20 °C. Isotopically labelled hFEN1 protein samples were exchanged into the appropriate NMR buffer and concentrated to 0.5 mM using a Vivaspin 20 (10 kDa MWCO) concentrator (4000×g at 4 °C).

2.2 Kinetic studies of natural abundance hFEN1 and hFEN1$_{Y40A}$

2.2a Substrate design

All substrates for kinetic studies were ordered, synthesised using standard phosphoramidite chemistry and purified using reverse-phase high performance liquid chromatorgraphy (HPLC) from LGC Biosearch Technologies (Risskov, Denmark). Two single-stranded oligonucleotides were prepared to make the substrate construct S1-X. Each strand, F1-X and T1 were quantified by the Beer-Lambert law using a calculated extinction coefficient at 260 nm using a UV-Vis NanoDrop spectrophotometer (ThermoFisher Scientific) (**Table 2**). The calculated ε260 value was obtained using the nearest neighbor method on the Integrated DNA Technologies website (https://www.idtdna.com/calc/analyzer).[81] S1-X was prepared by heating the two strands F1-X and T1 at a ratio of 1:1.1 at 95 °C for 5 minutes in 10 mM HEPES-KOH pH 7.5, 6 mM KCl, 4 mM NaN_3 and 0.1 mM EDTA. The solution was subsequently left to anneal at room temperature. F1-X contains a 5' terminal fluorescein (FAM) phosphoramidite for subsequent fluorescence detection. The X refers to the moiety on the 2' position of the labeled adenosine ribose in **Table 2**, prepared by using a suitably modified adenosine phosphoramidite. There were a total of four different moieties on the 2' position as listed below. Each has a pKa value derived from their corresponding triphosphate varieties (**Table 3**).[82]

Oligonucleotide Strand	Sequence	ε260 (L mol^{-1} cm^{-1})
F1-X	3'-CACAGCTCGTCAGGAAC**A**TTTTT-5'-FAM	236760
T1	3'-CCTGCTGAAGCAGCAGTGTTCCTGACGAGCTGTG-5'	323200

Table 2. Oligonucleotides used for kinetic studies. The red adenosine in F1-X represents the nucleotide where the 2' position is altered.

Adenosine 2'-X Moiety	pKa value of the 3'-OH ribose position
-H	14.3
-OCH_3 (-OMe)	13.8
-OH	12.98
-F	12.74

Table 3. The four different 2'-X moieties and the corresponding pKa of the 3'-OH group.

2.2b Measuring linear free energy relationships of hFEN1 and hFEN1$_{Y40A}$

For steady-state substrate-limiting multiple turnover reactions, reaction mixtures (180 µL) were assayed at 2.5, 5, 7.5 and 10 nM substrate in 50 mM HEPES pH 7.5, 100 mM KCl, 8 mM $MgCl_2$, 0.1 mg.mL^{-1} BSA, 0.02 % NaN_3 1 mM dithiothreitol (DTT) at 37 °C. Reactions were initiated by adding hFEN1 (10 pM) or hFEN1$_{Y40A}$ (**Table 4**). The concentration of enzyme was designed to

give roughly 10% product formation after 10 minutes to avoid product inhibition or substrate depletion, thus giving initial rates (V_0). Aliquots (20 µL) of reaction were quenched in excess EDTA (50 µL of 500 mM EDTA pH 8), at 2, 4, 6, 8, 10 and 20 minutes. Reactions with the S1-OH were quenched in 200 mM EDTA and 8 M urea, pH 7, to avoid base-catalysed auto-hydrolysis. Reactions at each substrate concentration were independently repeated three times. These experiments were carried out for all 2' modified substrates.

Substrate Concentration (nM)	**S1-X Moiety and hFEN1$_{Y40A}$ concentration used for each reaction (pM)**			
	2'-H	**2'-OMe**	**2'-OH**	**2'-F**
2.5	40	10	10	10
5	40	10	10	10
7.5	40	10	1	1
10	40	10	1	1

Table 4. Protein concentrations used for the hFEN1$_{Y40}$ reaction to give true initial rates of roughly 10% product formation after 10 minutes.

The rate of product formation was assessed by denaturing high-performance liquid chromatography (dHPLC) (Wave fragment analysis dHPLC fitted with a fluorescence detector; Transgenomic, Glasgow). Roughly 65 µL of each quenched reaction was injected onto a reverse-phase ion-pairing column (DNAsep, Transgenomic). The mobile phase consisted of 0.1% acetonitrile, 1 mM EDTA, 2.5 mM tetrabutyl-ammonium bromide (TBAB) in water. A gradient of 3.6%-70% acetonitrile (**Fig. 2.i**) allowed for separation of 5'FAM labelled product (6 nt) and substrate (23 nt) as detected using an attached fluorescence detector (excitation 494 nm, emission 525 nm). Integrals of product and substrate peaks gave the percentage of the product hydrolysed at any given time, which when multiplied by the total substrate added gave the concentration of product produced, as given by equation 2.1.

$$[P] = \frac{Area(product)}{Area(product) + Area(substrate)} \times [S]_{Total} \quad (2.1)$$

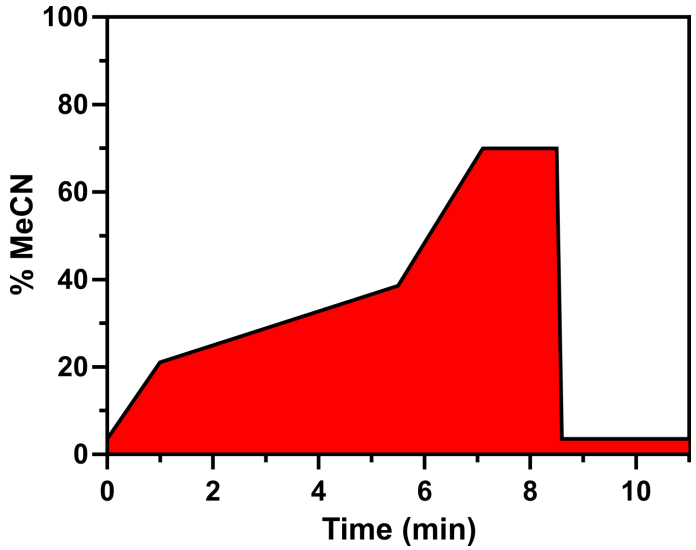

Figure 2.i. Gradient from the dHPLC system. The gradient runs as follows with percentage MeCN listed: 0 min 3.6%, 1 min 21%, 5.5 min 39%, 7.1 min 70%, 8.5 min 70%, 8.6 min 3.6% and then 3.6% until the end at 11 minutes.

Normalised initial rates were calculated by linear regression of [P] versus time and subsequently dividing by enzyme concentration used, hence giving V_0/[E]. The second order rate constant k_{cat}/K_M for each 2'-X substrate was calculated by plotting a linear regression through plots of V_0/[E] versus [S]. Brønsted plots and subsequent βlg values were calculated according to equation 2.2. All linear regression was carried out using least-squares fitting in Graph Pad Prism software. All experiments were independently repeated three times.

$$Log\left(\frac{k_{cat}}{K_M}\right) = \beta lg.pKa + A \quad (2.2)$$

For rapid-quench single-turnover reactions, enzyme (E) was mixed with S in a 1:1 volume mixture (80 µL of each) using a quench flow device (Hi-Tech Sci Ltd, Sailsbury, UK). To obtain maximal single turnover conditions, E concentrations were at least 8 times the K_M of hFEN1 while S concentrations were 1/40 times the K_M.[83] The K_M was previously determined by Dr. Jack Exell to be 151 nM.[84] Therefore [E] was 1200 nM, while [S] was 3.75 nM. 80 µL of both [E] and [S] in 55 mM HEPES-KOH pH 7.5, 110 mM KCl, 8 mM $MgCl_2$, 0.1 $mg.mL^{-1}$ BSA, 0.02 % NaN_3 and 1 mM DTT at 37 °C were prepared. After mixing and a delay of 4.5, 9, 12.1, 19.4, 27.6, 30.6, 40.8, 41.8, 57.5, 82.1, 124.2, 140.8, 240.8, 440.8, 840.8, 1640.8, 3240.8, 6440.8, 12841, 25641 or 51241 ms, the reaction was subsequently quenched with 80 µL of quench solution (8 M urea, 200 mM EDTA pH 7). Quenched reactions were analysed by dHPLC on the WAVE system as described above in order to obtain percentage of product. All reactions were independently repeated three times for all S1-X substrates. The first order rate of product formation was fit to a one phase (equation 2.3) and a two phase exponential fit (equation 2.4).

$$[P] = [P]_0\left(1 - e^{-kt}\right) \text{ (2.3)}$$

$$[P] = f_{fast}[P]_0\left(1 - e^{-k_{fast}.t}\right) + 1 - f_{fast}[P]_0\left(1 - e^{-k_{slow}.t}\right) \text{ (2.4)}$$

Where $[P]_0$ is the maximal product produced ($[P]_0$ = [ES] = [S]), f_{fast} is the fraction of the span (from 0 to $[P]_0$) accounted by the faster of the two components and k_{fast} and k_{slow} are two individual rate constants. The Brønsted plot of $Log_{10}(k_{fast})$ versus pKa of 3'-OH leaving group was plotted as described above (equation 2.2), but replacing the second order rate constant with first order single-turnover rates.

2.2c Second Order Rate Constant Determination as a Function of Viscosity

hFEN1 was assayed at different viscosities using the 5'-fluorescein (FAM) labelled bimolecular DNA oligo construct S1-H, prepared as described above in 2.2a. Substrate-limiting multiple turnover reactions reaction mixtures (180 µL) were assayed at 2.5, 5, 7.5 and 10 nM substrate in 50 mM HEPES pH 7.5, 100 mM KCl, 8 mM $MgCl_2$, 0.1 $mg.mL^{-1}$ BSA, 0.02 % NaN_3 and 1 mM DTT at 37 °C and the indicated relative viscosity. Buffer viscosity (η) was adjusted using glycerol (0-36% v/v). Glycerol relative viscosities (η/η_0) and their corresponding % v/v were calculated using temperature-corrected density calculations at 37 °C and compared to the viscosity of water (η_0).[85] Reactions were initiated by adding appropriate amounts of hFEN1 (**Table 5**), and aliquots (20 µL) of reaction were quenched in excess EDTA (50 µL of 500 mM EDTA pH 8) at 2, 4, 6, 8, 10 and 20 minutes. All reactions were repeated independently three times.

Relative viscosity (η/η_0)	Enzyme concentration (pM)
1	1
1.5	1.5
2	2
2.5	2.5
3	3
3.5	3.5

Table 5. The enzyme concentrations used in reactions with differing relative viscosity

The rate of product formation was assessed by dHPLC as described previously in **section 2.2b**. Estimates of second-order rate constants were derived from the slope of normalized initial rate ($V_0/[E]$) versus the concentration of substrate [S1-H]. The reciprocal of the mean k_{cat}/K_M values for each relative viscosity were plotted and linear regression was performed on the linear portion of the plot (η/η_0 1-2). The equation of the slope (equation 2.5) was solved for $\eta/\eta_0 = 1$ in order to calculate the second order rate constant with no viscogen present $(k_{cat}/K_M)^0$.

$$\frac{1}{\frac{k_{cat}}{K_M}} = \frac{1}{k_{on}^0}\frac{\eta}{\eta^0} + \frac{k_{off}^0}{k_{on}^0 + k_{chem}} \text{ (2.5)}$$

The normalised second order rate constant $\frac{(k_{cat}/K_M)^0}{(k_{cat}/K_M)}$ was plotted versus relative viscosity to give the relationship given in equation 2.6.[86]

$$\frac{(k_{cat}/K_M)^0}{(k_{cat}/K_M)} = \frac{k_{chem}}{k_{off}^0 + k_{chem}} \frac{\eta}{\eta^0} + \frac{k_{off}^0}{k_{off}^0 + k_{chem}} \quad (2.6)$$

2.3 NMR spectroscopic studies of isotopically labeled hFEN1

2.3a Spectrometer details

All NMR experiments were performed using standard pulse sequences on either a Bruker 600 MHz Avance DRX spectrometer equipped with a 5-mm TXI cryoprobe with z-axis gradients or a Bruker 800 MHz Avance I spectrometer equipped with a 5-mm TXI probe with z-axis gradients. Both spectrometers were operated with TopSpin 2.

2.3b NMR sample conditions

A variety of temperatures, pH and buffer molecules were screened by Dr. David Finger in order to obtain both acceptable ^{1}H-^{15}N TROSY spectra and stabilise ^{15}N-labelled 0.5 mM protein samples to remain in solution for long timescales (months). A summary of temperature and pH conditions tried can be found in **Table 6**. Following the screen, samples were henceforth always recorded at pH 7.5 and a temperature of 25 °C (298K, relative to d_4-methanol signal).

pH	Observation	Temperature °C	Observation
<7.5	Very rapid aggregation	<25	Decrease in S:N with no noticeable increase of protein stability
7.5	Mixture of stability and acceptable S:N	25	Mixture of stability and acceptable S:N
>7.5	Poor S:N due to exchange of amides	>25	Protein aggregation accelerated with no noticeable increase in S:N

Table 6. Properties of NMR experiments, pH and temperature.

The buffering molecules screened were: 10 mM Tris-HCl, 10 mM HEPES-KOH and 10 mM potassium phosphate. No apparent changes were noted between them, so HEPES was chosen at pH 7.5, due to its pKa being closest to the desired pH. KCl was added and enhanced protein stability and quality of spectra up to about 76 mM. The effect of multi-valent cations was assessed by adding 8 mM $MgCl_2$, 8 mM $CaCl_2$, 6 mM $LuCl_3$, 8 mM $ZnCl_2$ or by leaving the buffer in the presence of 0.1 mM EDTA. The different reducing agents tried were 100 mM DTT, tris(2-carboxyethyl)phosphine (TCEP) and beta-mercaptoethanol (βME). βME gave the best response

and high concentrations prevented depletion due to oxidation with air. Further oxidation prevention was achieved by blanketing samples with argon and sealing with a Precision Seal® rubber septa cap (Sigma-Aldrich Z554014). The buffer solution called **NMR buffer** (**Table 7**) contained: 10 mM HEPES-KOH pH = 7.5, 76 mM KCl, 0.1 mM EDTA, 4 mM NaN_3, 100 mM β-ME. Furthermore, D_2O (10% v/v) and 0.05 mM trimethylsilyl propanoic acid (TSP) were added for the deuterium lock and chemical shift reference standard, respectively. Samples (0.5 mL) were transferred into a 5 mm NMR tube, blanketed with argon and sealed with the rubber septa cap.

Buffer Name	Composition
NMR Buffer	10 mM HEPES-KOH pH 7.5, 76 mM KCl, 0.1 mM EDTA, 4 mM NaN_3, 100 mM βME, 10% v/v D_2O and 0.05 mM trimethylsilyl propanoic acid (TSP)
NMR Buffer Low Salt	10 mM HEPES-KOH pH 7.5, 6 mM KCl, 0.1 mM EDTA, 4 mM NaN_3, 100 mM βME, 10% v/v D_2O and 0.05 mM trimethylsilyl propanoic acid (TSP)

Table 7. NMR Buffers used for assignment, relaxation and titration experiments

^{15}N-labelled hFEN1 samples in **NMR buffer** had different small molecules added to it to improve stability of the sample or quality of the spectra. The list of small molecules added are listed in **table 8**.

Small Molecule	Concentration relative to hFEN1
Arginine/Glutamic Acid 1:1 mixture (Arg:Glu).	Always 50 mM Arg:Glu.
Deoxyribose adenosine monophosphate	10 X
3'-5' adenosine bis-phosphate	10 X
dTpdA dinucleotide	1 X
Cyclopropyl inhibitor[26]	2 X

Table 8. List of small molecules added to the NMR buffer solution. Additional molecules were added to the buffers as mentioned in the text.

2.3c Assignment of hFEN1$_{P188A}$

Comparison of ^{1}H-^{15}N TROSY spectra of ^{2}H,^{15}N,^{13}C-labelled hFEN1$_{P188A}$ with ^{15}N-labelled protein showed that all backbone amide groups had back exchanged from deuterium to protium atoms. For the backbone resonance assignment of hFEN1$_{P188A}$, a 0.5 mL sample contained 0.5 mM ^{2}H,^{15}N,^{13}C-labelled hFEN1 in **NMR buffer**. hFEN1$_{P188A}$ backbone $^{1}H^{N}$, ^{15}N, $^{13}C^{\alpha}$, $^{13}C^{\beta}$ and ^{13}C' were assigned using the standard suite of ^{1}H-^{15}N TROSY and 3D TROSY-based constant time experiments (HNCA, HN(CO)CA, HNCACB, HN(CO)CACB, HN(CA)CO and HNCO).[87] ^{1}H chemical shifts were referenced relative to the internal TSP signal, whereas ^{15}N and ^{13}C chemical shifts were indirectly referenced using nuclei-specific gyromagnetic ratios. Peak picking and frequency matching was carried out within CCPNMR Analysis version 2.4,[88] and backbone assignments were compared to the previous assignment of wild type hFEN1 (BMRB accession code 27160) to

confirm assigned residues. The secondary structure content and Random Coil Index-S^2 (RCI-S^2) prediction of wild type hFEN1 was conducted by uploading the backbone $^{1}H^{N}$, ^{15}N, $^{13}C^{\alpha}$, $^{13}C^{\beta}$ and $^{13}C'$ chemical shifts to the TALOS-N webserver.[89]

2.3d Relaxation measurements of hFEN1

Wild type hFEN1 and $hFEN1_{K93A}$ samples for ^{15}N backbone fast timescale relaxation measurements were performed using ^{2}H,^{15}N-labelled protein (**section 2.1e**) in a 5-mm Shigemi NMR microtube assembly matched with D_2O. The sample was in **NMR buffer**.

Spin-lattice ^{15}N relaxation rates (R_1), rotating frame ^{15}N relaxation rates ($R_{1\rho}$) and heteronuclear steady-state ^{15}N-{^{1}H} NOE (hNOE) values were obtained using interleaved TROSY-readout pulse sequences.[90] Temperature compensation was applied in the R_1 experiment by incorporating a spin-lock pulse placed off-resonance in the inter-scan delay, equal to the longest spin-lock time and RF power of the $R_{1\rho}$ experiment. Relaxation delays (t) of 0, 80, 240, 400, 400, 640, 800, 1200, 1760 and 2400 ms were used to calculate R_1, and delays (t) of 1, 20, 20, 30, 40, 60, 90, 110, 150 and 200 ms were used to calculate $R_{1\rho}$. The inter-scan delay was 3.5 s and the strength of the RF spin-lock field during $R_{1\rho}$ measurement was 1400 Hz at 600 MHz and 1866.7 Hz at 800 MHz. For the hNOE measurement, two interleaved experiments were acquired with relaxation delays of 10 s. For the determination of R_1 and $R_{1\rho}$ rates, the decay of backbone amide peak heights were fitted using a decaying exponential function in CCPNMR according to equation 2.7.[88]

$$Peak\ Height = A.e^{-R.t} \quad (2.7)$$

Where *A* is the height at time 0, and *R* is either R_1 or $R_{1\rho}$ relaxation rates. The error of the fit was done by a covariance method. Relaxation parameters were obtained for 192 residues because 59 residues were omitted from further analysis due to peak overlap or poor signal-to-noise ratios. For hNOE experiments, the ratio of the unsaturated to unsaturated peak height was measured and the error was taken as a percentage of the average noise level of the spectra. R_2 values were calculated from R_1 and $R_{1\rho}$ rates according to equation 2.8.

$$R_2 = \frac{R_{1\rho}}{sin^2(\theta)} - \frac{R_1}{tan^2(\theta)} \quad (2.8)$$

Where $tan\ (\theta) = \frac{\omega_1}{\Omega}$, and ω_1 is the spin lock RF field (1400 Hz at 600 MHz and 1866.7 Hz at 800 MHz) and Ω is the offset of the ^{15}N resonance of interest with respect to the ^{15}N carrier frequency.[90]

2.3e Model free analysis

Model-free analysis was performed using *relax*[91–95] on the Sheffield-WRGRID ICEBERG high performance computing cluster. Using the R_1, R_2 and hNOE values at 600 MHz and 800 MHz and the coordinate geometry of backbone amide N-H bonds as provided by the crystal structure 3Q8K,[27] model-free analysis was executed. Initial analysis provided τ_m values for all analysable residues. 16 residues from the arch region, α2-α3 loop and the C-terminus were then excluded due to lack of coordinates in the crystal structures. The remaining 176 residues, called 'structured' residues were subjected to a full analysis with each diffusion tensor being judged by an Akaike information criteria (AIC) (**Section 5.2b, Table 11**). The omitted 'unstructured' 16 residues were modeled with a spherical diffusion tensor and oblate spheroid to see which was preferable by AIC. Both analyses were conducted for the 192 residues for which relaxation data at both field strengths was available; of these, only 179 were processed fully, as 13 were removed due to large errors or computational eliminations.

2.3f Chemical shift perturbation mapping of Ca^{2+} and dAMP

A ^{15}N-labelled hFEN1 containing the K93A mutation ($hFEN1_{K93A}$) that hindered catalysis, was expressed using a custom-mutated pET29b-hFEN1-336 vector as described in **section 2.1b**. The $hFEN1_{K93A}$ protein was prepared as described for the wild type protein using the ^{15}N autoinduction method (**section 2.1d**). ^{1}H-^{15}N TROSY spectra were separately recorded in **NMR buffer**, with either 0 mM or 8 mM $CaCl_2$ added. Differences in chemical shift ($\Delta\delta_H$ or $\Delta\delta_N$) were calculated using equation 2.9.

$$\Delta\delta = \delta_{Metal\ bound} - \delta_{Metal\ free} \quad (2.9)$$

Absolute changes in weighted chemical shifts (ω) were determined using equation 2.10.

$$\omega = \sqrt{(\Delta\delta_H^2 + (\alpha.\Delta\delta_N)^2)} \quad (2.10)$$

Where the correction factor, α, for ^{15}N was 0.14, roughly the fraction of gyromagnetic ratios between ^{1}H and ^{15}N.

Absolute weighted chemical shifts were measured in the same fashion for the titration of deoxyribose adenosine monophosphate (dAMP) into hFEN1, with each titration point measured versus the original unbound chemical shift for ^{1}H and ^{15}N. A solution of 0.39 mM ^{15}N-labelled hFEN1 in **NMR buffer** with 8 mM $MgCl_2$ had aliquots of 200 mM dAMP added to it according to **table 9**. Dissociation constant (K_D) values were calculated from plots of dAMP concentration versus ω, fitted in KaleidaGraph 4.0 (Kaleidagraph, Synergy Software, Reading, PA, USA) using a quadratic binding equation (equation 2.11).[96]

$$\omega = \omega_{max} \frac{([E] + [dAMP] + K_D) - \sqrt{([E] + [dAMP] + K_D)^2 - 4[E][dAMP]}}{2[E]} \quad (2.11)$$

Where [E] and [dAMP] are the total enzyme and dAMP concentrations respectively. ω_{max} is the theoretical weighted chemical shift difference when fully bound to dAMP.

Titration	dAMP (mM)	[E] (mM)	dAMP stock added (µL)
0	0	0.39	0
1	0.39	0.39	1
2	0.78	0.39	1
3	1.6	0.39	2
4	3.1	0.38	4
5	4.7	0.38	4
6	6.2	0.38	4
7	7.8	0.38	4

Table 9. dAMP titration table with concentrations of both ^{15}N-labelled hFEN1 and dAMP, along with the volume of 200 mM dAMP stock added at each titration point.

2.3g Substrate DNA preparation for NMR studies

The DNA oligonucleotide used for NMR studies was purchased, synthesised using standard phosphoramidite chemistry and purified using reverse-phase high performance liquid chromatorgraphy (HPLC), from LGC Biosearch Technologies (Risskov, Denmark) (**Table 10**). The concentration of the dual-hairpin substrate (DHPS1) was quantified using absorbance at 260 nm and molar extinction coefficients calculated according to the 'nearest-neighbor' method the same as in **section 2.2a**.[81] The formation of the secondary structure required the oligonucleotide to be heated at 95 °C for 5 minutes in 10 mM HEPES-KOH pH 7.5, 6 mM KCl, 4 mM NaN_3 and 0.1 mM EDTA and subsequently annealed at room temperature.

Oligonucleotide Strand	Sequence	ε260 (L mol^{-1} cm^{-1})
DHPS1	5'-dTGAAAGGCAGAGCGCTAGCTCTGCCTTTCGAGCGAAGCTCC-3'	382900

Table 10. Oligonucleotide used for NMR studies.

2.3h hFEN1$_{K93A}$ substrate DNA complex formation

Initially, 500 µM ^{15}N-labelled hFEN1$_{K93A}$, ^{2}H,^{13}C,^{15}N-labelled hFEN1$_{K93A}$ or ^{2}H,^{15}N-labelled hFEN1$_{K93A}$, prepared as described above, in **NMR buffer** was titrated with sub-stoichiometric aliquots (0, 50, 100, 200, 250 µM) of lyophilized DHPS1 until the DNA was equivalent to the protein (500 µM). At 50-100 µM of DNA, the sample precipitated and the quality of the spectra decreased rapidly. To overcome protein precipitating in complex with DNA, a hFEN1$_{K93A}$-DNA complex was prepared in **NMR buffer low salt**: 10 mM HEPES-KOH pH 7.5, 6 mM KCl, 4 mM

NaN_3 and 0.1 mM EDTA in the presence of excess DNA substrate (1:1.1) under dilute conditions of protein and DNA (roughly 5 μM). Low KCl concentrations were used to slow dissociation of the complex. The sample was concentrated to ~500 μM using a Vivaspin 20 (10 kDa MWCO) spin columns and 100 mM βME, 10% D_2O (10% v/v) and 0.05 mM TSP were added (final conditions in **Table 7**). The same resonance assignment strategy, using 3D NMR spectra, was conducted with a $^{2}H,^{15}N,^{13}C$-labelled $hFEN1_{K93A}$-DNA complex prepared in this manner (**section 2.3c**). Furthermore ^{1}H-^{15}N TROSY relaxation measurements, as described previously (**section 2.3d**) were carried out on a $^{2}H,^{15}N$-labelled $hFEN1_{K93A}$-DNA complex sample.

A Ca^{2+} titration was performed on a ^{15}N-labelled $hFEN1_{K93A}$-DNA complex. ^{1}H-^{15}N TROSY spectra were acquired at 0, 0.5, 1, 2, 4, 6 and 8 mM $CaCl_2$, and chemical shift differences were recorded as described in equation 2.10.

Chapter 3: Confirmation of a conformational change in the hFEN1 catalytic cycle

3.1 Introduction to kinetic assays of human flap endonuclease

3.1a Overview of the current kinetic model of hFEN1

Kinetic studies on FENs are numerous and have revealed much about specific residues and their importance to catalysis. Such studies have compared altered enzymes and substrates with their unmodified forms.[37,38,97] Kinetic studies, and in particular non-equilibrium or pre-steady state kinetics can also reveal further information about the enzyme-catalysed reaction. This can include information about the catalytic chemistry as well as provide evidence for intermediates on the kinetic pathway. Thus one approach to elucidating rate-limiting steps such as conformational change is to perform kinetic studies.

When modeling the rate of an enzyme, it is useful to use the Michaelis-Menten relationship. In Michaelis-Menten kinetics the overall rate *v* is a combination of two processes. First, enzyme [E] and substrate [S] equilibrate reversibly, forming an enzyme-substrate complex [ES]. Second, there is a first-order rate of [ES] decay to [E] and product [P] called k_{cat}. The equation is shown below:

$$\frac{v}{[E]_0} = \frac{k_{cat}[S]}{K_M + [S]} \quad (3.1)$$

Where $[E]_0$ is assumed to be the total enzyme concentration in the system ($[E]_0 = [E]_{Free} + [ES]$), and K_M is the Michaelis constant. Using this, a simple rate equation can be drawn up for hFEN1 (**Fig. 3.i**). The mechanism of Michaelis-Menten has three main assumptions. First, the rates of formation and decay of [ES] are equal, or in other words [ES] is under a steady state condition. Hence, it is important to measure initial rates of reaction, devoid of product inhibition or substrate depletion. Second, the total substrate concentration [S] is equal to the concentration of free substrate at any given point in the reaction. Therefore the concentration of [E] << [S] to avoid a considerable percentage of bound substrate versus free substrate. Third, that the binding and dissociation happen much more rapidly than product formation, thus it is assumed that $k_{cat} << k_{off}$.

$$E + S \underset{k_{off}}{\overset{k_{on}}{\rightleftharpoons}} [ES] \xrightarrow{k_{cat}} E + P$$

Figure 3.i. The simple Michaelis-Menten relationship of enzyme kinetics. Enzyme [E] binds to substrate [S] in an equilibrium to form an enzyme-substrate complex [ES]. The [ES] complex then decays by rate constant k_{cat} to [E] and product [P].

In terms of substrate as mentioned before (**Section 1.2a**), human flap endonuclease (hFEN1) prefers to act upon a double-flapped (DF) DNA substrate with a 1nt 3' flap and any length of 5' flap.[22] The reaction of this substrate yields a short single stranded (ss) DNA flap product (P) and a nicked DNA duplex (Q). It is highly likely that the enzyme does not retain the flap P after hydrolysis has occurred; therefore its departure does not impact on the rate. However, in the case of FEN1, the nicked DNA duplex is potentially still bound to the enzyme after hydrolysis, contrary to the simple model in **Figure 3.i** (**Fig. 3.ii**).

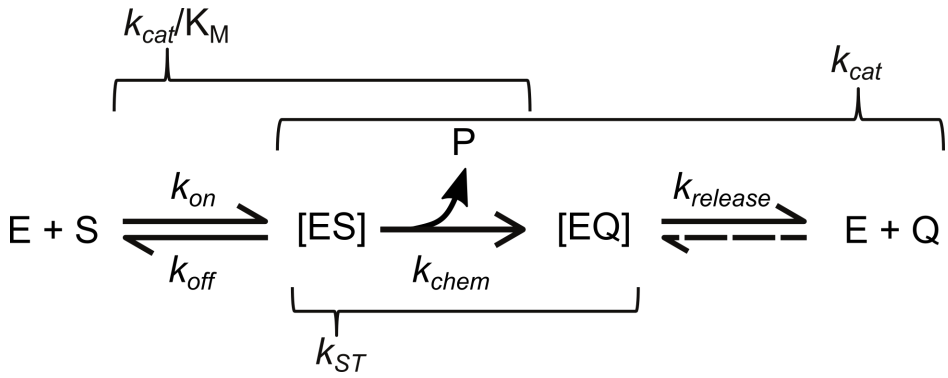

Figure 3.ii. An updated kinetic model based on the fact that a product has to be released after chemistry. Due to the nature of nucleases binding double stranded (ds)DNA, it is common to see a final release step ($k_{release}$) of said dsDNA. Meanwhile the small flap product P dissociates rapidly. Macroscopic experimental kinetic quantities and what they can measure are represented in brackets. k_{cat}/K_M measures all micro-rates up to and including the first irreversible step (k_{chem}). k_{cat} measures everything after substrate binding, and finally k_{st} measures everything after binding but before release.

Pre-steady state conditions, or single-turnover conditions, where [E] >> K_M > [S] measure the rates associated with processes after substrate is bound but before product is released. Previous studies have shown that for a multitude of FENs the single-turnover rates (k_{ST}) were faster than multiple-turnover k_{cat} values.[22,83,97,98] In hFEN1 the difference is an order of magnitude slower (k_{cat} 3 s^{-1} versus k_{ST} of 22 s^{-1}).[22] Therefore the release of nicked DNA product (Q) was determined to be rate limiting under substrate-saturating multiple-turnover (k_{cat}) conditions.

The third assumption of the Michaelis-Menten mechanism is not always accurate for a protein system and sometimes $k_{cat} \approx k_{off}$ (Briggs-Haldane)[99] or the more extreme condition where $k_{cat} >> k_{off}$ (Van Slyke-Cullen).[100] If $k_{cat} >> k_{off}$, any kinetic step after substrate association occurs much faster than dissociation, so the enzyme is said to be diffusion controlled/limited. Therefore

steady-state kinetic rates under substrate-limiting conditions (([E] < [S] < K_M, or k_{cat}/K_M conditions) will be close to that of bimolecular association (10^7-10^{-9} M^{-1} s^{-1}).[101] Indeed for T5 bacteriophage FEN,[83] and hFEN1[22] the rate has been shown to be within this range. Furthermore, increasing the viscosity of the reaction solution can allow for a slow k_{on} and k_{off}, without affecting k_{cat}, hence probing for diffusion control. Under k_{cat}/K_M conditions the T5 FEN was found to be dependent on the rate of diffusion, hence it was found to be diffusion controlled.[42]

As hFEN1 is sequestered to replication forks *in vivo*,[44] the physical limitation of diffusion is likely a minor factor. Thus, it has been suggested that the single-turnover rates are likely more biologically relevant.[8] The single-turnover reaction measures all rates after binding, and before product release, which according to **figure 3.ii** just encompasses the rate of chemistry. However, earlier studies with T5 bacteriophage FEN show no change in the k_{ST} when chemically altering the pKa of the 3' hydroxyl leaving group, suggesting that another slower rate masks the k_{ST}.[42] Applying this knowledge from older FENs, a more detailed model for hFEN1 catalysis has been proposed (**Section 1.3e** and **Fig. 3.iii**).[8] The presence of a conformational change, slower than k_{chem}, has been suggested between the initial [ES] complex and a chemically competent complex [ES]''. In this chapter the presence of this conformational change step will be confirmed for hFEN1 along with its implications for catalysis.

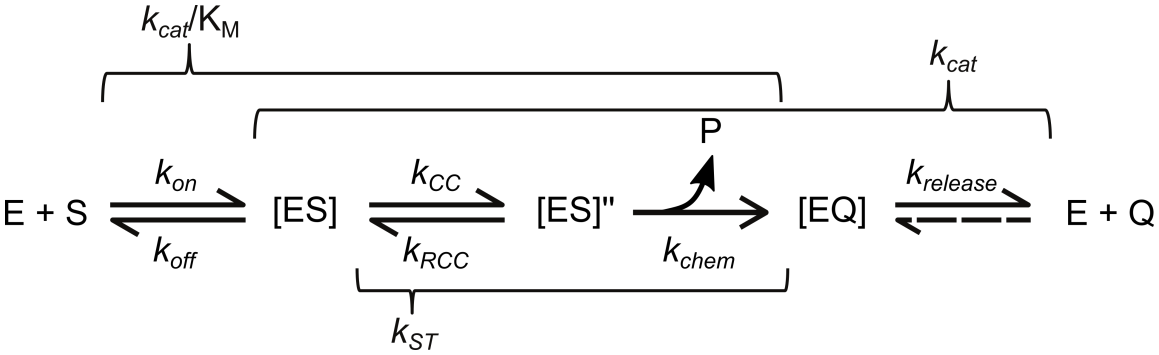

Figure 3.iii. The proposed reaction pathway of for hFEN1. The proposed conformational change rates (k_{CC} and k_{RCC}) are so slow, they mask any augments of k_{chem}. Therefore it is thought that the k_{st} reflects the rate of this conformational change. It should also be noted that the conformational change step could encompass many subsequent steps.

3.1b The theory of Brønsted analysis and previous work done on T5 FEN

In single-turnover conditions for hFEN1, there are two proposed kinetic steps: a conformational change and the chemical reaction. In order to confidently assign the single-turnover rates as a rate of conformational change, the rate of the chemical step needs to be decoupled from k_{ST}. As discussed in **section 1.1b**, phosphate diester hydrolysis occurs most likely as an in-line

associative nucleophilic reaction (comparable to S_N2 reactions with carbon). After an activated nucleophile attacks the phosphorus, a penta-covalent phosphate transition state or intermediate is formed. This collapses back to a normal tetrahedral structure when the bond forming the axial 3' bridging oxygen is broken (**Fig. 3.iv**). The breaking of this bond requires a buildup of negative charge on the 3'-OH leaving group. Thus by stabilising the negative charge on the 3'-OH the reaction can progress faster.

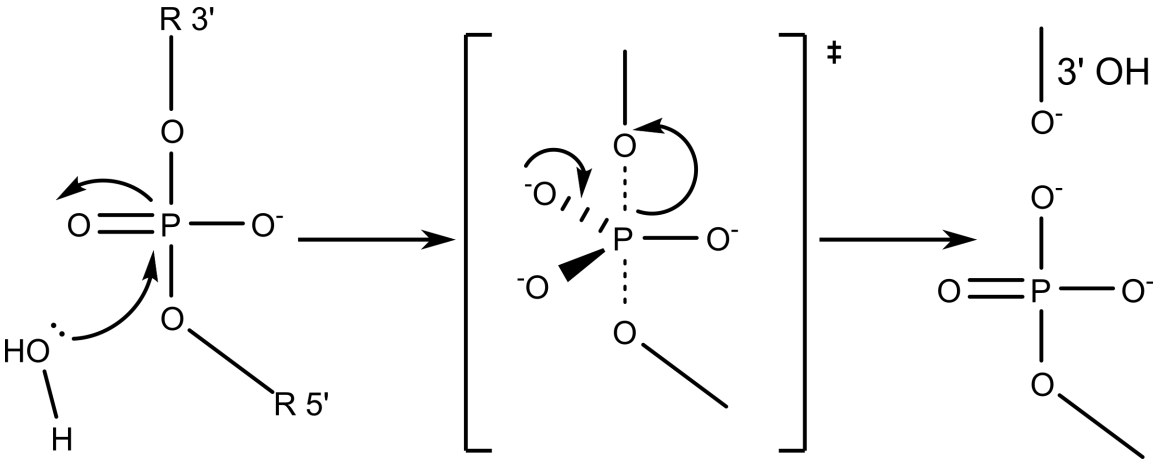

Figure 3.iv. Base-catalysed Phosphodiester hydrolysis. The water is activated by a base and attacks the phosphorus. This forms a penta-covalent intermediate. There is a buildup of negative charge on the 3'-bridging oxygen during the penta-covalent phosphorus intermediate. This causes the axial P-O bond to break.

A measure of ability of the 3'-OH to stabilise negative charge is its pKa. A lower value represents a higher proportion of conjugate base present (O^-) at equilibrium.

$$3'OH \leftrightharpoons 3'O^- + H^+ \ (3.2.a)$$

$$pKa = -Log_{10}\left(\frac{[3'O]^-[H]^+}{[3'OH]}\right) \ (3.2.b)$$

So by having a lower 3'-OH pKa value, the penta-covalent phosphate diester will collapse faster as the negative charge is stabilised on the 3'-bridging oxygen. Furthermore the negative charge on the 3'-O^- product will be stabilised more, the lower the pKa value is. Therefore base-catalysed phosphodiester hydrolysis should occur faster at lower 3'-OH pKa values. Experimentally this effect can be measured by plotting the log of rate versus the pKa of the leaving group, called a Brønsted plot (**Fig. 3.v**). A linear relationship of this plot is known as a linear free energy relationship (LFER), with the mathematical relationship shown in **figure 3.v**. The βlg value represents the sensitivity of the leaving group to pKa and *A* is the reaction specific constant. As βlg approaches zero the transition state becomes less sensitive to the pKa, and vice versa. Brønsted analysis of non-enzymatic base-catalysed methylphenylphosphate diesters shows a large βlg of -0.94.[102]

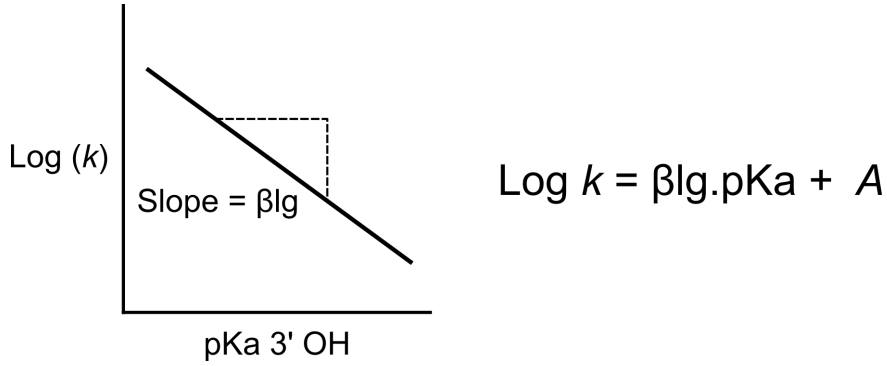

Figure 3.v. Brønsted plot of a phosphate diester. The gradient of the plot provides information as to how sensitive the rate of reaction is to the buildup of charge on the 3'-OH leaving group. The equation for the slope is given on the right.

However for enzymatic reactions the situation is more complex. Enzymes that rely on acid catalysis show small βlg values.[103] However, for two-metal ion nucleases similar to FEN1, like the bacterial alkaline phosphatase, a large βlg was observed.[104] Therefore it is more likely that hFEN1 would display a similar large βlg. Interestingly the opposite is seen for T5 bacteriophage FEN-catalysed reactions; changing the pKa of the 3'-OH leaving group does not affect rates under k_{cat}/K_M conditions. This might be unsurprising considering that the rate is diffusion controlled at low substrate concentrations. Therefore the rate of k_{chem} is masked by diffusion. However, the same result is obtained for single-turnover reactions, indicating k_{ST}, even when augmented by lower pKa, does not increase. This is evidence that some other process, a conformational rearrangement (k_{CC}, k_{RCC}) of the protein after being bound must be rate-limiting or close to rate-limiting. Furthermore, measurement of this single-turnover rate would give an idea of the timescale of which this process is happening. Similar experiments were therefore carried out on human FEN1 to see if a similar relationship could be found.

3.2 Substrates used in kinetic experiments

In order to analyse the leaving group LFER of hFEN1, four substrates were designed with a variety of 2' groups present on the departing nucleoside in the flap strand (**Fig. 3.vi**). In order of decreasing 3' OH pKa the 2' groups (X) on the adenosine sugar were hydrogen (2'-H), methoxy (2'-OMe), hydroxyl (2'-OH) and fluoride (2'-F). An optimal double flapped substrate structure was used for all reactions.[22] The overall substrate secondary structure was engineered to be a static double-flapped structure (DF1), with an interchangeable flap strand (F1-X) and a constant template strand (T1). A mismatch on the 3' upstream flap was also present so the flap structure does not equilibrate, preventing interchangeable lengths of 5' and 3' flaps in solution.

Fluorescein was attached on the 5' terminus of the flap strand for detection by DNA denaturing high performance liquid chromatography (dHPLC).

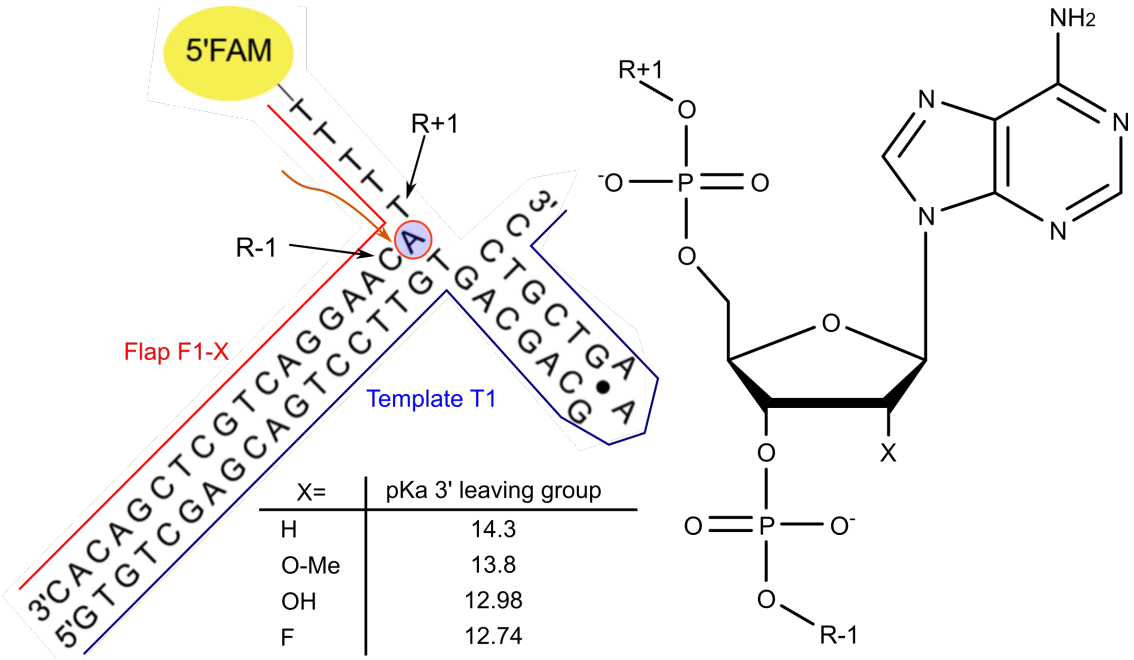

Figure 3.vi. Schematic of the double flapped hFEN1 substrate used in kinetic experiments. The template strand, T1 has a GAA hairpin resulting in a duplex upstream region with a 1nt 3' flap. This was annealed with an interchangeable 5' flap strand, F1-X with a 5' fluorescein on the 5' terminus. The expected hydrolysis site is noted with an orange arrow. The 2' position (X) of the adenosine ribose sugar (red circle) was altered according to the table. pKa values for the 3'-leaving group are listed in the table, determined as their triphosphate equivalents.[82]

3.3 Brønsted analysis to determine chemistry dependence on the rate of reaction of hFEN1.

3.3a Brønsted analysis under k_{cat}/K_M conditions

Initially LFER were assessed under substrate-limiting multiple-turnover conditions where the rate constant for the reaction is k_{cat}/K_M. Rates of reaction for the various modified hFEN1 substrates were measured by a discontinuous time assay at 37 °C. Reaction vessels with 2.5, 5, 7.5 and 10 nM of appropriate substrate were prepared and initiated with an appropriate amount of enzyme. Aliquots of the reaction removed at defined times were quenched in 500 mM EDTA pH 8 to chelate any catalytic Mg^{2+} ions, hence halting reaction. For the 2'-OH substrate a quench of 200 mM EDTA pH 7 and 8 M urea was used instead to prevent base-catalysed auto-hydrolysis of the phosphate group. Each time point was analysed by reverse phase ion pairing dHPLC with tetrabutyl-ammonium bromide (TBAB) as the interface with the solid phase. By using a gradient of acetonitrile, substrate (23nt) and flap product (6nt) were separated and detected using a fluorescence detector for enhanced sensitivity allowing pico-

moles of DNA to be detected. By using the integral of each peak and the known total substrate concentration $[S]_{Total}$ it was possible to work out the concentration of product [P] at each time point by the equation below:

$$[P] = \frac{Area(product)}{Area(product) + Area(substrate)} \times [S]_{Total} (3.3)$$

Product concentration versus time was plotted and a linear relationship indicated that initial rates (V_0) under steady-state kinetics were being measured. Least squares fits of these lines produced initial rates, which were normalised versus the total enzyme concentration $[E]_0$ then plotted versus substrate concentration. A full rate profile for each substrate concentration was repeated three times.

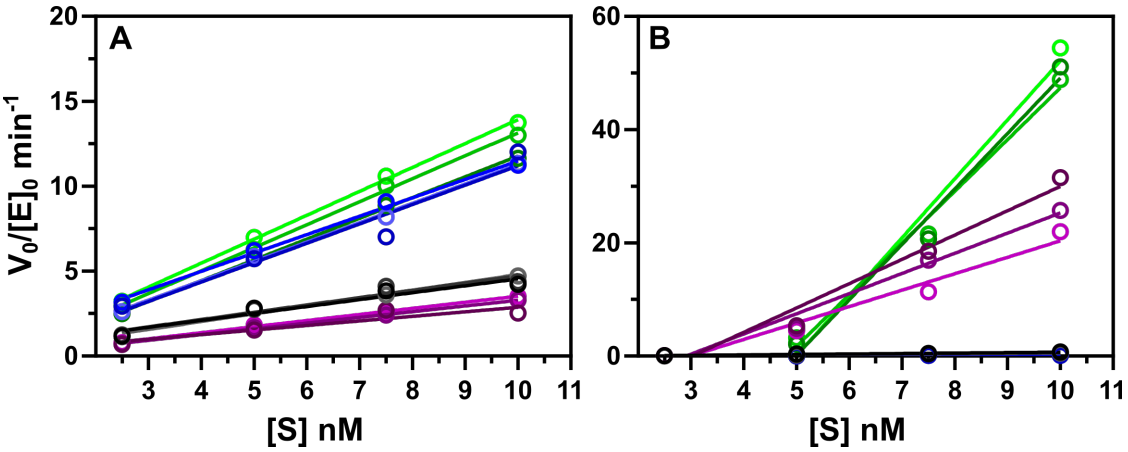

Figure 3.vii. Apparent second-order rate plots with different 2' substrates. Both plots show each replicate for substrate containing 2'H (greyshade), 2'OMe (purples), 2'OH (blues) and 2'F (greens). On the left (A) is wild type hFEN1 and on the right (B) is the mutant Y40A. Each linear regression was fitted using least squares fits.

Normalised rates of reaction ($V_0/[E]_0$) versus [S] data for each replicate were fit by linear regression to obtain k_{cat}/K_M rate constants using a least squares fit (**Fig. 3.vii**). The average k_{cat}/K_M for the 2'-deoxyriboadenosine containing substrate (0.43 nM^{-1} min^{-1}) was similar to previously reported kinetic studies.[22,27] The rates of reaction for the other substrates were all within an order of magnitude. The 2'OMe containing substrate was the closest at 0.32 nM^{-1} min^{-1}, while the 2'F (1.3 nM^{-1} min^{-1}) and 2'OH (1.1 nM^{-1} min^{-1}) were slightly higher in rate. When the log of the mean rate constant was plotted against the pKa of the leaving group, there was a shallow slope. This was fitted with a least squares fit to produce Brønsted coefficient (βlg) value of around -0.38 ± 0.1 (**Fig. 3.viii.a**).

To see if a negative LFER could be achieved with hFEN1, the same experiment was conducted but with a previously reported mutant hFEN1. The point mutation was present in the active site at the nucleobase stacking residue tyrosine 40, which was mutated to an alanine (Y40A). Initial

catalytic efficiency of hFEN1$_{Y40A}$ protein with the deoxyribo-adenosine substrate was 5-times lower than wild type protein, a bit less pronounced than previous studies.[27] However the reaction with the mutant was highly sensitive to the leaving group pKa. There was a 100-fold increase in rate for 2'OH containing substrate and a 50-fold increase for the 2'F containing substrate compared to 2'H.The Brønsted plot showed a βlg value of -1.2 ± 0.4 (**Fig. 3.viii.b**), close to that of the T5 FEN catalytic mutant K83A.[42] The 2'OMe containing substrate was problematic with high uncertainty in the results. Furthermore its rate constant was much lower than expected, which could be due to steric hindrance during reaction. Therefore it was omitted from the least squares fit of the Y40A Brønsted plot, although it would have enhanced the βlg value to -1.5 if included.

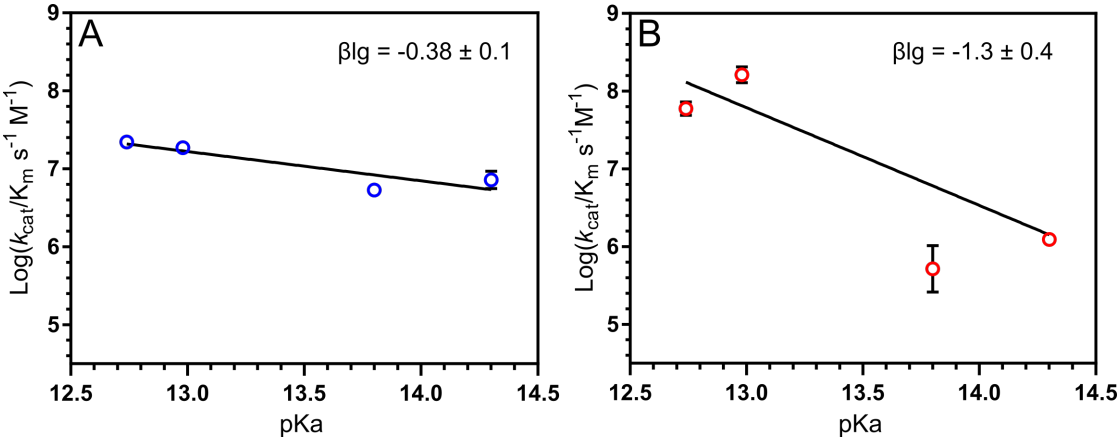

Figure 3.viii. LFER of hFEN1 and hFEN1$_{Y40A}$ mutant. The wild type (A) is in blue, and the Y40A mutant (B) is in red. Error bars are shown as the standard error of the mean from each replicate, where not shown they are smaller than the size of the point. The fit was done by linear regression using least squares, in Y40A (B) the 2'OMe data point was omitted from the fit.

The shallow LFER obtained for hFEN1 under k_{cat}/K_M conditions could be due to a few factors. Either the reaction of hFEN1 has an element of acid catalysis involved in its hydrolysis, therefore nullifying the affect of 3'OH pKa dependence. However this is would not be explained by a steep βlg for the Y40A mutant, as this residue is not implicated in contact with the phosphodiester, but is responsible for positioning of the nucleobase. It is also inconsistent with other two-metal ion enzyme catalysed systems like the bacterial alkaline phosphatase. The second option for hFEN1 is that the rate is indeed sensitive to 3'OH pKa but the measured macroscopic rate constant is masked by a much slower rate of another process, which would be consistent with T5 FEN results under similar conditions. Physical processes responsible for this slower rate could include bimolecular association between substrate and enzyme or a conformational change after substrate binding. The Y40A mutant can be rationalised by the second theory: k_{chem} becoming, or becoming close to the rate-limiting step in catalysis. Reducing the k_{chem} to within

orders of magnitude of a slower micro-rate(s) would make the macroscopic rate constant sensitive to k_{chem}. Therefore, by altering k_{chem} by altering the pKa of the 3'OH, the k_{cat}/K_M rate constant is changed. For wild type hFEN1, the k_{chem} is much faster and is masked by potentially slower kinetic steps, which results in the shallow LFER. Therefore another microrate, either bimolecular association, or a conformational change after association, is probably limiting the rate of reaction under these conditions.

3.3b Brønsted analysis to determine chemistry dependence at single-turnover conditions.

The LFER of hFEN1 was assessed under maximal single-turnover conditions to eliminate the effects of diffusion. Rates of reaction for the same modified hFEN1 substrates were measured at 37 °C at single-turnover conditions where [E] was at least eight times K_M and [S] was 1/40 times K_M. These reaction conditions were used previously to assess the maximal single single-turnover rate, as further addition of enzyme yielded no further increase in rate.[83] Under these conditions [E] and [S] were instantaneously mixed and left to react at set times (a range of 10 ms to 50 s, **section 2.2b**) before being quenched in a solution of 8 M urea and 200 mM EDTA pH 7. This was done on a quench-flow apparatus, mixing 1:1:1 volumes of [E], [S] and quench solutions. Each time course for a substrate was repeated three times. Aliquots of the quenched reaction were taken and analysed by dHPLC as described previously (**section 3.3a**). Analysis of the integrals of peaks showed the percent product produced versus time (**Fig. 3.ix.a**).

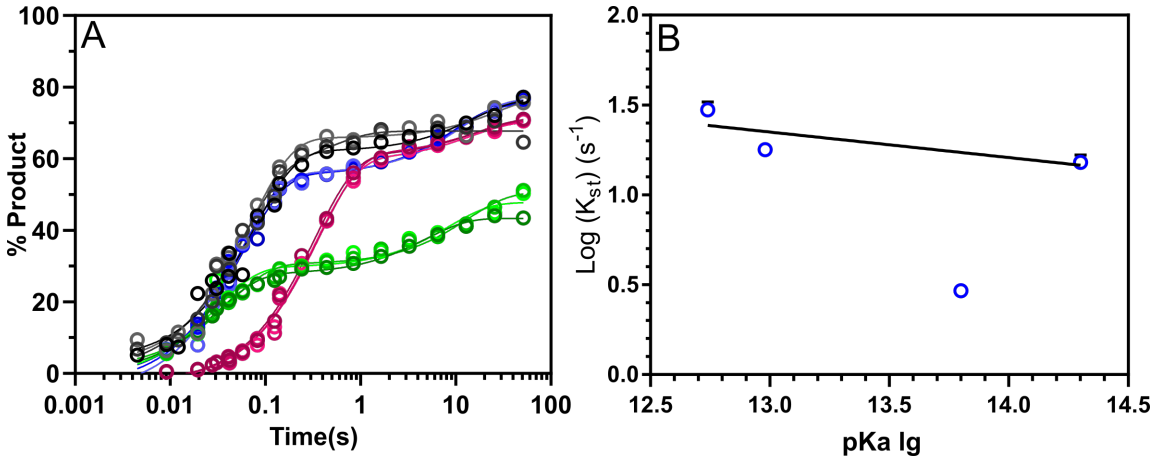

Figure 3.ix. Single Turnover Brønsted analysis of hFEN1. (A) The single turnover two-phase exponential (by least-squares) fits of each replicate. The substrates were 2'-H (greyshade), 2'-OMe (purples), 2'-OH (blues) and 2'-F (greens). (B) The plot of log(k_{ST}) versus 3'-OH pKa of hFEN1 protein with error bars visible except when smaller than the data points. The fit was a linear regression using least squares fit through all but the 2'-OMe point.

It is expected for the rate of reaction to be first order, as it is measuring the rate of production of product following the equation: $[ES] \rightarrow [EQ] + [P]$. Therefore product should rise

exponentially by the equation $[P] = [P]_0(1 - e^{-kt})$ (where $[P]_0$ = [ES] = $[S]_{Total}$). In reality however, the data is observed to have a two-phase exponential relationship. Comparing fits of single and two phase exponentials by an extra sum of squares F test found the two-phase preferable in all cases. This could be interpreted by a percentage of the substrates failing to react, and could be due to a few reasons. Firstly non-productive complex formation at high enzyme concentrations could occur, whereby an enzyme is bound to the substrate at an inactive location (i.e. the blunt ended downstream region, **Fig. 3.vi**), meaning the enzyme would first need to dissociate before another bound in the correct position. This could be especially true if multiple enzymes at high concentration bind to one substrate. Secondly, enzyme denaturation could occur when injected under high pressures in the quench-flow apparatus inactivating a proportion of the bound [ES] complexes. On another note, the plateau at higher time points should converge towards 100% product formation (i.e., $[S]_{Total}$), however most the plots only reached around 70% product (**Fig. 3.ix.a**). The anomalous results for the 2'-F substrate shows a very low percentage (50%) of product formed. It is not known why the 2'-F substrate would display this unusually low plateau.

When log(k_{ST}) values were plotted versus 3'-OH leaving group values there was a near-zero β_{lg} of -0.14 ± 0.1 (**Fig. 3.ix.b**). The k_{ST} of the 2'-OMe substrate is much lower than expected and has been omitted from the fit. This could be due to steric clashes when bound to the enzyme, causing a much slower than expected reaction. This has been noted in the previous study with T5 bacteriophage FEN.[42] The βlg value is comparable to that under k_{cat}/K_M conditions, showing that the measured k_{ST} is insensitive to 3'-OH leaving group pKa. This further suggests that k_{chem} is a much faster rate than other slower rate determining processes. One such process that is eliminated at maximal single-turnover conditions is the bimolecular association step, as all enzyme is assumed to be bound to substrate from the start of the reaction. Therefore the slow microrate masking k_{chem} is likely to occur once enzyme is bound to substrate. This further lends weight to the T5 FEN model, whereby a conformational change, or series of conformational changes is rate-limiting.

3.4 Viscosity effects on enzymatic reactions

3.4a Diffusion-limited reactions and the effect of viscosity on bimolecular association rates (k_{on})

Changing the viscosity of the reactant medium influences the rates of enzymatic catalysed reactions. In particular it has been used to investigate the role the rate of bimolecular

association (k_{on}, **Figs. 3.i-iii**) has in the overall enzyme-catalysed reaction. It can be shown using diffusion coefficients that the rate of association is inversely proportional to the viscosity (η) of the solution.[86] For a simple Michaelis-Menten relationship (**Fig. 3.i**) under steady-state conditions it can be shown that rate of product formation (EP) (k_{prod}) is given by:

$$k_{prod} = \frac{k_{chem}\, k_{on}}{k_{off}+k_{cat}} \quad (3.4)$$

Assuming that viscosity only affects the bimolecular diffusion rates, and noticing the inverse correlation, it is possible to derive the equations:

$$k^0_{on}\, \eta^0 = k_{on}\, \eta \quad (3.5.a)$$

$$k^0_{off}\, \eta^0 = k_{off} \eta \quad (3.5.b)$$

Where the superscript 0 indicates conditions where there is no viscosity-enhancing agent present. Therefore by substituting equations (3.5.a-b) into (3.4), and taking the inverse of k_{prod} it is possible to obtain:

$$\frac{1}{k_{prod}} = \frac{1}{k^0_{on}}\frac{\eta}{\eta^0} + \frac{k^0_{off}}{k^0_{on}+k_{chem}} \quad (3.6)$$

This means that in theory by plotting the inverse of rate versus the relative viscosity (η/η^0) it is possible to retrieve the rate of association as the inverse of the gradient. Experimentally it is easier to represent this as a ratio of $k_{prod}{}^0$ over k_{prod}:

$$\frac{k^0_{prod}}{k_{prod}} = \frac{k_{chem}}{k^0_{off}+k_{chem}}\frac{\eta}{\eta^0} + \frac{k^0_{off}}{k^0_{off}+k_{chem}} \quad (3.7)$$

In practice for hFEN1 the measurement of k_{prod} was measured under k_{cat}/K_M conditions. This provided kinetic parameters up to and including the first irreversible step of catalysis, which for hFEN1 is assumed to be hydrolysis (k_{chem}). This presents two extremes, one where diffusion is much faster than the hydrolysis ($k_{off} >> k_{chem}$) and one where the diffusion is much slower than hydrolysis ($k_{off} << k_{chem}$). In equation 3.7 if k_{off} dominates, or if the reaction is not diffusion-limited then the gradient will be 0. However, if k_{chem} dominates, or if the reaction is diffusion-limited then the equation above will show a slope of 1. Previous studies of T5 bacteriophage FEN showed that it is diffusion-limited under k_{cat}/K_M conditions, with a normalised rate versus relative viscosity slope of near 1.[42] Therefore the same viscosity studies on hFEN1 were performed to see if it displayed similar characteristics.

3.4b The effect of viscosity on the k_{cat}/K_M rate constant of hFEN1.

hFEN1 was subjected to different glycerol concentrations during substrate-limiting multiple-turnover reactions. The same discontinuous time assay was used as in **section 3.3a** except

different glycerol concentrations were used in the reaction. The same double-flapped substrate as in **section 3.2** was used, with a deoxyribose adenosine. Viscosities of glycerol solutions at 37 °C (η) were calculated relative to water (η_0) (**Section 2.2c**).[85] After analysis of peak integrals by dHPLC, the resulting plots of initial rates versus substrate concentration show a marked decrease in the slope (k_{cat}/K_M) upon addition of more viscogen (**Fig. 3.x**). The resulting plot of the reciprocal of k_{cat}/K_M versus relative viscosity however shows much greater viscosity dependence than anticipated (equation 3.6). Therefore only the first three points were analysed by linear regression to solve for $(\eta/\eta_0) = 1$, to calculate $1/(k_{cat}/K_M)^0$. From equation 3.6 in theory the slope should be the inverse of k_{on}^0, and in fact the association rate obtained is 8.2×10^7 M^{-1} s^{-1}, placing it close to diffusion and comparable to the reported values for T5 FEN.[42]

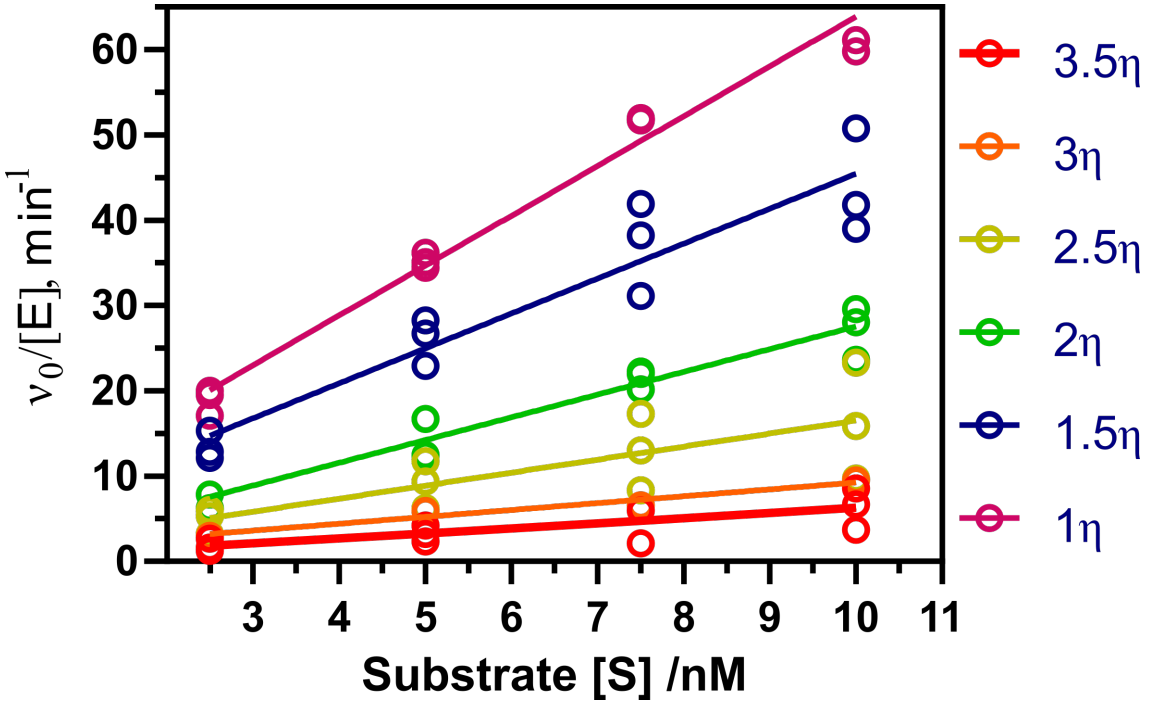

Figure 3.x. Initial rate graphs of hFEN1 under different relative viscosity conditions. Each relative viscosity (listed on the right) was repeated three times. The linear regression by least squares was performed on the entire dataset to give k_{cat}/K_M values for six relative viscosity values.

The resulting plot of $\frac{(k_{cat}/K_M)^0}{(k_{cat}/K_M)}$ versus η/η_0 is expected to give a slope of 1 for reactions controlled by diffusion (equation 3.7).[42,86] The derived data for hFEN1 shows that for relative viscosities of 1 to 2, the slope was approximately 1, as expected, compared to T5 FEN data. However, a much greater impact on relative rate was observed with hFEN1 with η/η_0 values greater than 2; this appears to almost be exponential with a slope closer to 2 (**Fig. 3.xi.b**). This could be due to a few reasons: First, an assumption of the equations above is that only bimolecular association is affected by viscosity changes, thus the enhanced slope could be due

to more than one kinetic process being affected by viscogen, for example a conformational change. Second, it could be due to the glycerol stabilising an inactive conformation of hFEN1, with its osmolyte properties. This would have the effect of augmenting the viscosity dependence with an apparent inhibition. The second situation is unlikely as the same experiment done by Dr. Jack Exell show the viscogen sucrose also display a similar viscosity dependence. Nevertheless, the data suggests that some other process is being slowed by viscosity, causing a much greater decrease in rate than expected, questioning whether hFEN1 is solely diffusion limited.

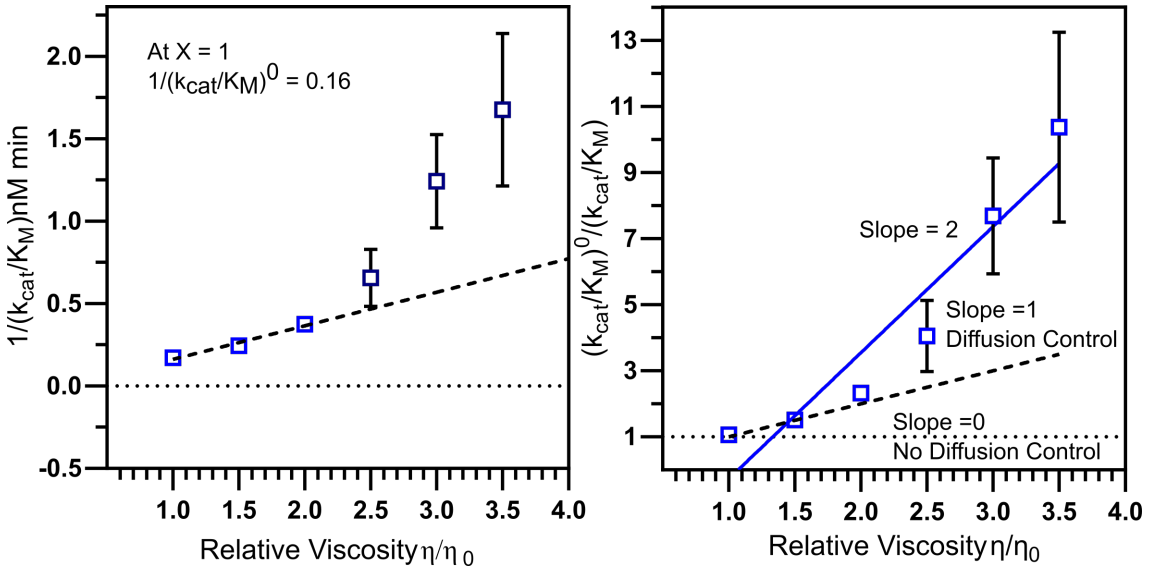

Figure 3.xi. Assessment of diffusion control for hFEN1. Inverse second order rate constants plotted versus relative viscosity (left). The linear regression by least squares was only done with the first three points. The second order rate constant at relative viscosity of 1, or $(k_{cat}/K_M)^0$, was calculated from the line equation according to equation 3.6. The reciprocal of the relative second order rate constants versus relative viscosity are present on the right. The linear regression of the experimental data is represented as a blue line, and has a slope of around 2. Slopes of 1 for diffusion limited enzymes and 0 for no diffusion limitation are represented as black dashed and dotted lines respectively.

3.5 Discussion

Using LFER it can be shown that for the WT hFEN1 reaction chemistry is not slow enough to impact rates under multiple or single-turnover conditions. The chemical rate k_{chem} was not enhanced significantly by decreasing the pKa of the 3'-OH leaving group in the wild type hFEN1 protein. Under single-turnover conditions there was a minimal LFER, but the 2'-OMe group skews the results, possibly due to steric clashes. More altered 3'-OH pKa substrates would be preferable to fully confirm this result. But it does suggest that the chemical rate must be an order of magnitude faster than another single-turnover rate. Thus, a proposed conformational change is rate-limiting or close to rate-limiting under single-turnover conditions.

The mutant hFEN1$_{Y40A}$, while not occurring naturally, has a significant sensitivity to 3'-OH leaving group pKa. It displays a 100-fold rate enhancement at lower 3' pKa values. This in effect means that catalytic mutants, including naturally found mutations such as E160D, of hFEN1 could preferentially target RNA over DNA.[105] One important location *in vivo* where this might have implications is the replication fork. Okazaki fragments are capped on their 5' termini by the RNA primers laid out by the Pol-α primase complex. Therefore the activity of chemically impeded hFEN1 mutants may be faster for RNA primers compared to DNA in the replication fork *in vivo*. Changing the efficiency by which hFEN1 processes the RNA primer and DNA portions of Okazaki fragments could cause defects during Okazaki fragment maturation.[106]

Viscosity studies on multiple-turnover rates show that the rate is affected in a non-linear fashion with respect to the relative viscosity of glycerol. This is most likely due to the effects of bimolecular diffusion and one or more physical process working together. Similar diffusion curves have been noticed previously for an enzyme that folds upon binding.[107] The viscogen could be affecting the kinetics of a folding event or similar conformational change in hFEN1 during catalysis. Indeed it has been noted that the presence of viscogens can slow conformational rearrangements.[108] The presence of viscogen can even slow rates for enzymes with conformational changes present in their catalytic cycle.[109] Furthermore, the effect is only seen for microviscogens like glycerol, but it is not noticed in the presence of macroviscogens like bovine serum albumin (BSA). For hFEN1, recent studies using single-molecule techniques have shown that glycerol or sucrose inhibits an apparent single-turnover rate, which is independent of diffusion.[40] However, the same study with the macroviscogen polyethylene glycol-8000 showed no inhibition of rate. The microviscogen slows the conformational rearrangement of domains by solvent friction, hence the dependence on viscosity, whereas the macroviscogen cannot. This further lends weight to the idea of a conformational change playing a role in the catalytic process of hFEN1. Therefore, the non-linear effect of viscogen on the relative second order rate constants observed in this chapter was likely due to a combination of two processes. Firstly viscogen affects the rates of bimolecular association and secondly, it slows subsequent conformational changes. The rate of the conformational change therefore plays an important role in catalysis. The main two questions remaining are, therefore, how fast is this conformational change occurring, and in what regions of the protein is it happening? The remainder of the thesis will attempt to answer these two questions.

Chapter 4: Optimisation of the hFEN1 protein system for assignment and subsequent NMR experiments

4.1 NMR spectroscopic studies of enzymes

NMR spectroscopy is a very powerful technique, used for decades on small molecules to characterise their structure, rates of reaction or interaction with each other. For macromolecules, such as proteins, the technique is slightly more complex. As the molecular weight of a compound increases so does the R_2 relaxation rate, allowing less and less time for acquisition of coherent data. Thus, line broadening leads to a decrease in the signal to noise ratio (S:N). A spectrum also becomes much more crowded, with hundreds of peaks rather than tens. Using specific techniques such as multi-dimensional spectra, TROSY pulse sequences and selective labeling of proteins, it is possible to nullify most of the problems associated with macromolecules and garner useful information about a protein system in solution. This chapter will cover the techniques for assembling a hFEN1 protein construct for NMR spectroscopic studies. It will also cover the screening procedures applied to acquire TROSY spectra with acceptable S:N, and the assignment strategy of a novel protein system.

4.2 Preparation of protein for NMR spectroscopic studies

4.2a Protein expression and purification

Human FEN1 (hFEN1) is a 380-amino acid residue protein that consists of a nuclease core domain (amino acid residues 1-336), which is sufficient for catalysis as shown by single-turnover rate constants.[27] The remainder is an extended C-terminus that contains the nuclear localization signal and a PCNA interacting peptide box (PIP box).[110] The extended C-terminus has also been implicated as necessary for interaction with several other DNA metabolic proteins.[111] Studying the protein dynamics relevant to catalysis was the main aim of the project; therefore, the hFEN1-336 construct used previously for X-ray crystallography was used.[27] A plasmid construct for hFEN1-336, complete with a C-terminal $(His)_6$-tag that was removable using the human rhinovirus 3C protease was produced previously by Dr. Brian Chapados.[27] Thus, the final hFEN1-336 construct, which is 38 kDa, contains 336 native amino acids with an additional six residues (LEVLFQ, see **Fig. 4.i**) present at the C-terminus from the protease recognition site. Herein, the hFEN1-336 construct will simply be called hFEN1.

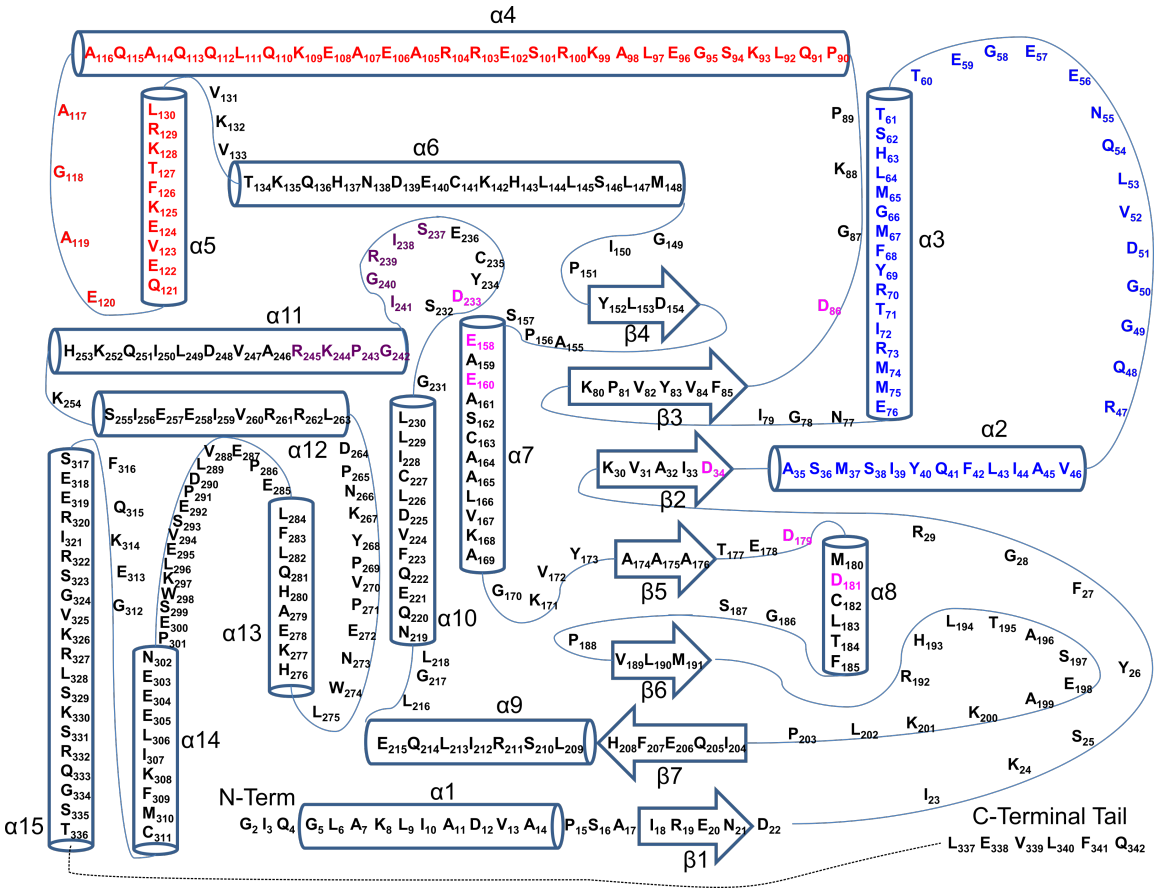

Figure 4.i. A secondary structure map of the hFEN1-336 construct. The primary sequence of hFEN1-336 is listed, with secondary structures mapped and α helices and β strands listed. It should be noted that the secondary structure elements are not spatially aligned. Magenta residues are active site carboxylates, purple residues are part of the H2tH motif, blue residues are part of the 3' flap binding pocket and red are part of the arch.

Isotopically enriched hFEN1 protein was produced in BL23(DE3)-RILP *E. coli* cells (**Section 2.1**), and purified using three chromatography procedures: Imobilised metal affinity chromatography (Co^{2+}-IMAC), anion-exchange and heparin-affinity. The IMAC column allows capture of the $(His)_6$-tagged protein from the cell lysate. Anion-exchange chromatography removes any DNA that is non-specifically bound to hFEN1. Purification attempts without this step resulted in large amounts of DNA contamination present (**Fig. 4.ii.a,b**). Finally the heparin column separates out any proteins that do not bind to DNA. Following purification, the protein samples were buffer exchanged into the appropriate NMR buffer (**Chapter 2: Table 7**). Overall this procedure produced pure samples of hFEN1 in sufficient yield for NMR spectroscopy samples (**Fig. 4.ii.c**). For ^{15}N labeled hFEN1 a cell grow-up of 1L resulted in 1.5 mol of protein, or approximately 57 mg. For triply labeled ^{2}H, ^{15}N, ^{13}C hFEN1 a cell grow-up of 500 mL resulted in 1.1 µmol or 40 mg of protein.

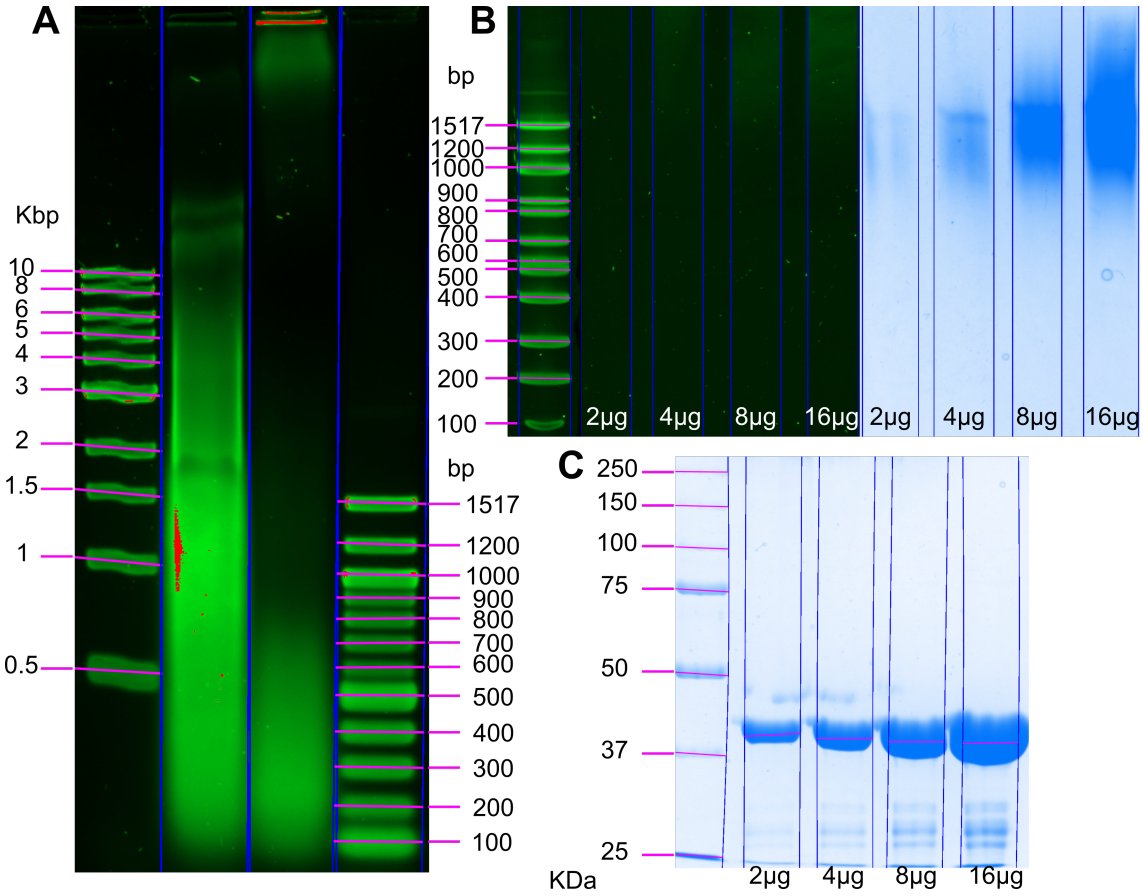

Figure 4.ii. Agarose gel electrophoresis and SDS PAGE of purified hFEN1. (A) Agarose gel, stained with SYBR safe of a protein sample taken after the purification procedure with no anion exchange column. There are large smears of DNA contamination, despite other purification steps. (B) Agarose gel of subsequent purification procedure with an anion exchange column step added after the IMAC column, the gel was stained with SYBR safe (left) for DNA and SYPRO ruby (right) for protein. (C) A SDS polyacrylamide gel showing the final purity of the sample. Bands at the bottom of the gel are potential degradation products from heating SDS samples at 95 °C, subsequent samples heated at 65 °C for 10 minutes do not show these bands.

4.2b Protein stability optimisation using different reducing agents

My contribution to the work began with an attempt to improve conditions of hFEN1 in solution under high concentrations, mainly in an attempt to inhibit precipitation. hFEN1 protein contains 6 cysteines (C141, C163, C182, C227, C235 and C311), with 4 of them containing buried thiol side-chains (163, 182, 235 and 311) while the other two are partially exposed (141, 227). The oxidation of cysteines have been linked to protein degradation,[112] and also hFEN1 is more catalytically active under reducing conditions due to it being a nuclear protein. Therefore the reducing agent dithiothreitol (DTT) (1-10 mM) was always included into the buffer for the spectra above. To ensure that the long-term instability of the protein was not caused by exhaustion of the reducing agent, high concentrations (100 mM) of DTT, tris(2-carboxyethyl)phosphine (TCEP) and beta-mercaptoethanol (βME) were supplemented to three separate samples. Surprisingly, 100 mM βME was the most effective at preventing precipitation

in the time necessary for dynamic experiments to be conducted (> 1 month). The effect of βME is likely due to its osmolyte-like properties in addition to its reducing capacity. Samples were also placed under argon to retard the oxidation of βME as well as preventing paramagnetic oxygen from dissolving in solution. Despite all of these measures, the protein still aggregated over the course of one month, although acceptable spectra could be obtained on this timescale.

4.3 Optimisation of ^{1}H-^{15}N TROSY spectra by mutagenesis and small molecule addition

4.3a Observations of the wild type ^{1}H-^{15}N TROSY spectra

Initial ^{1}H-^{15}N TROSY spectra taken previously by Dr. David Finger with perdeuterated samples in appropriate optimised buffer were promising, and yielded acceptable S:N for assignment purposes (**Section 1.5b**). Inspecting the ^{1}H-^{15}N TROSY assignment spectra of ^{2}H, ^{13}C, ^{15}N hFEN1 protein shows the scatter of chemical shifts was that of a well-folded protein (**Fig. 4.iii**). Most amide peaks fit within 105-133 ppm for ^{15}N and 6.6-9.5 ppm for ^{1}H. There were large differences in peak height, most peaks exhibited a range of 1-10 times intensity compared to the weakest peak (E257), with some reaching 25 to 35 times. This suggested that not all peaks were relaxing uniformly. There were 292 observable secondary amide peaks (excluding side chain amides from glutamines and asparagines and tryptophan N-H peaks). The backbone amides of prolines are indefinable due to their lack of an amide proton and account for 15 missing peaks in the ^{1}H-^{15}N TROSY spectrum. The initiator methionine is not present due to removal by N-terminal amino peptidase during expression, so M1 is not observed. However, there was still a discrepancy between the 292 detectable amide peaks and the 324 theoretically detectable peaks. This inconsistency could be due to broad peaks or peak overlap in crowded regions of the spectra.

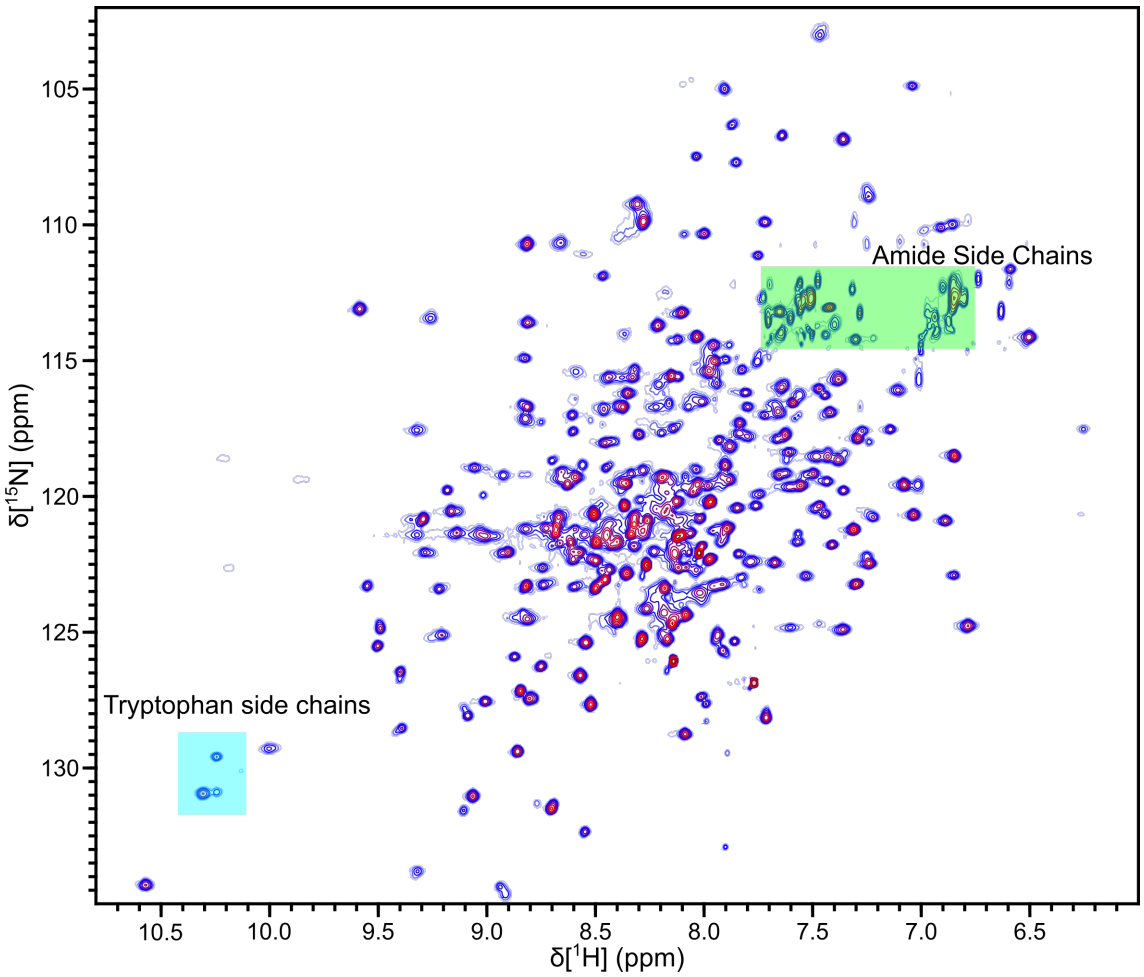

Figure 4.iii. ^{1}H-^{15}N TROSY spectrum of ^{2}H, ^{13}C, ^{15}N hFEN1 for assignment purposes. Each peak corresponds to a ^{1}H-^{15}N correlation, with most backbone amides showing as a peak, alongside sidechains of asparagines, glutamines (green box) and tryptophans (cyan box). The scatter is indicative of a well folded protein structure. The peaks are much sharper due to perdeuteration of the protein sample. The sample was done in the presence of 0.1 mM EDTA

Peak overlap is caused by two or more amide groups sharing similar chemical shifts and can result in confusing protein fingerprinting. Changing the chemical environment of each residue, usually by altering the tertiary protein structure, can separate peaks. Furthermore, triple resonance experiments can usually separate out overlapped peaks due to the carbons having distinctive chemical shifts. Inherent peak broadening in large molecules due to reasons discussed in **section 1.5a** can result in a reduction in the S:N to the point where a peak has vanished into the noise. Furthermore, peaks exchanging between two or more distinct chemical states could cause peak broadening. If the rate of exchange is approximately equal to the chemical shift difference, peaks will broaden out and can be lost in the noise. Finally, although the protein sample was optimised for stability it still aggregated over weeks. Protein aggregate will cause problems with magnet shimming, as well as differing conformations being sampled as the pulse sequences are acquired. This will result in a lowered S:N as well as undesirable conformations, which are unrepresentative of the natural protein structure. Therefore two

potential solutions to this problem were undertaken. Firstly, small molecules were added in an attempt to stabilise the protein and potentially shift peaks in crowded regions. Secondly, residues were mutated on the protein, in order to enhance stability and/or change the structure enough to notice a change of exchange regimes.

4.3b Addition of small molecules to enhance spectral quality and sample stability

Previous NMR studies have shown that small molecules can also increase protein stability at high concentrations.[113] A 1:1 arginine/glutamic acid mixture, various mono- and di-nucleotides, and known inhibitors of hFEN1 were added to determine if these molecules would improve spectra and stabilize the sample. A 50 mM 1:1 arginine/glutamic acid mixture did not significantly help the sample in stability or spectral quality. Addition of 3'-5' adenosine bis-phosphate (10 x protein concentration), 1x dTpdA dinucleotide (1 x protein) and a previously reported cyclopropyl inhibitor (2 x protein),[26] all caused chemical shift changes in hFEN1 (**Fig. 4.iv**). Most of the shifts were observed in or around active site regions, such as the N-terminus of the protein or proximal to active site carboxylates. There were more dramatic changes with the inhibitor than with the bis-phosphate or dinucleotide. This is consistent, as it is expected all these molecules would occupy the active site to a certain degree, but the inhibitor is known to bind to the active site with nM affinity.[26,114]

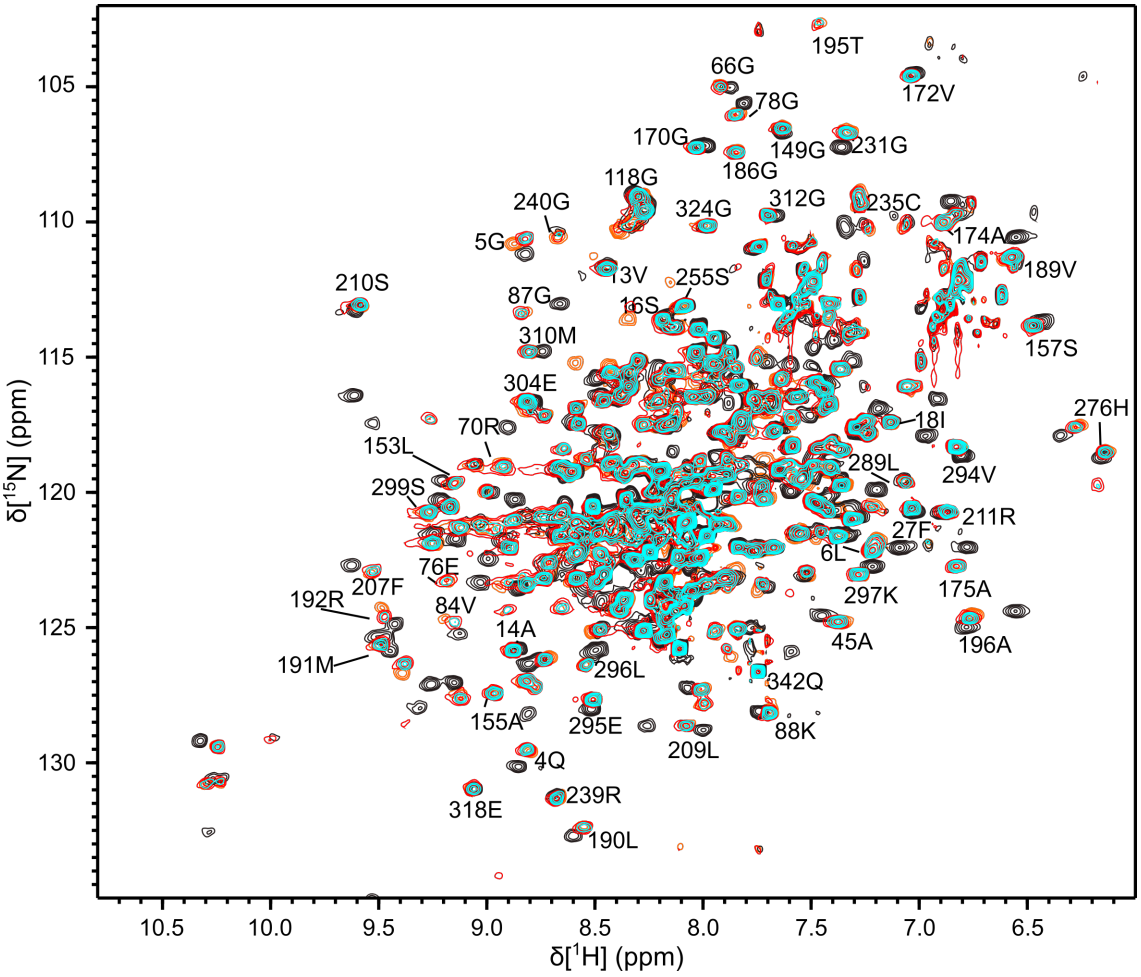

Figure 4.iv. ^{1}H-^{15}N TROSY spectra of hFEN1 with different small molecules added. The spectra showed are wild type protein (red), hFEN1 with 1:1 dTpdA dinucleotide (cyan), 10:1 3'-5' bis-phosphate (orange) and 2:1 cyclopropyl inhibitor (dark grey). The spectra are all recorded in the presence of 8 mM $CaCl_2$.

Equally dramatic changes were observed with addition of deoxyribose adenosine monophosphate (dAMP). Upon addition of dAMP, changes were observed in and around the active site regions, like with the previous inhibitor and nucleotides. However there was also an unexplained change in dynamics of certain residues associated with the arch and 3' binding pocket of hFEN1 (**Fig. 4.v**). To understand these exchange processes better, a titration of dAMP was performed in the presence of 8 mM $CaCl_2$. Concentrations of dAMP (0, 0.39, 0.78, 1.6, 2.8, 4.7, 6.2 and 7.8 mM) caused steady shift changes in active site regions until reaching saturation around 6.2 mM, indicating a fast exchange process. One example was S157 (**Fig. 4.v.d**), which is close to the active site carboxylates E160 and E158, that reports a K_D of 4.5 ± 1.4 mM. Other fast exchanging residues were A155 (**Fig. 4.v.c**) (K_D 3.4 ± 0.85 mM), Q4 (**Fig. 4.v.c**) (K_D 1.6 ± 0.55 mM) and G78 (**Fig. 4.v.a**) (K_D 3.0 ± 1.2 mM), all within a similar weak mM binding state. The binding was not uniform however, as A45 displayed a slow exchange pattern (**Fig. 4.v.e**), while G118 disappeared completely, suggesting an intermediate exchange process (**Fig. 4.v.b**) and G87 was more fast-intermediate (**Fig. 4.v.b**).

To test whether the binding was linked to the active site, 8.1 mM EDTA was added to the sample to chelate all Ca^{2+} ions. All peaks that were shifted by the addition of dAMP were shifted back to their original EDTA position, and G118 reappeared (**Fig. 4.v.a**). This strongly suggests that the interaction location of the dAMP is in the active site and furthermore, is mediated by the divalent cations present there. Overall, although shifts were observed for various residues upon the addition of nucleotides and FEN inhibitors, none of the small molecules significantly improved long-term sample stability and so were not chosen to use for assignment purposes.

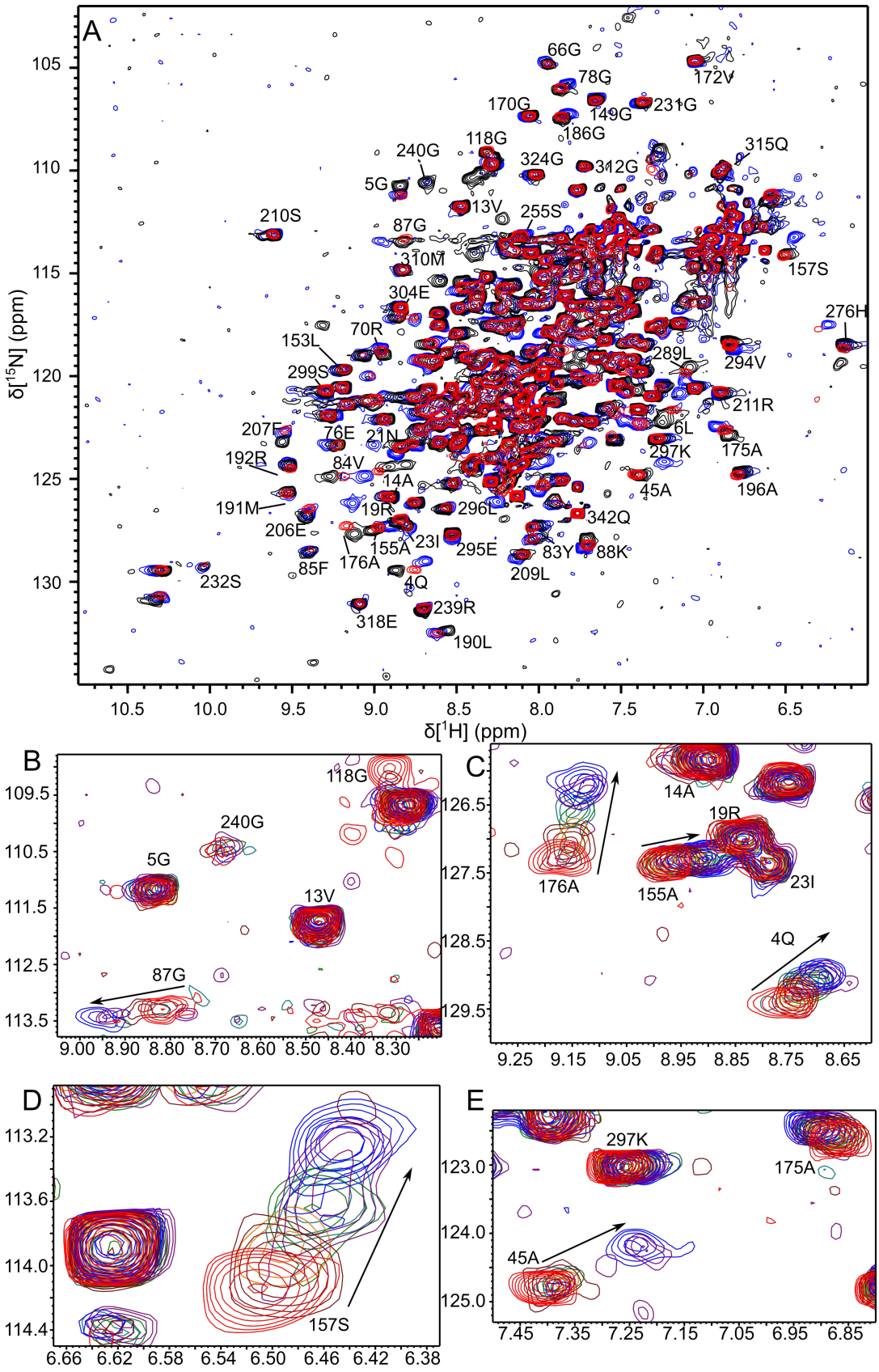

Figure 4.v. dAMP titration of hFEN1. (A) A ^{1}H-^{15}N TROSY spectra showing wild type hFEN1 in 8 mM $CaCl_2$ (red), hFEN1 in 8mM $CaCl_2$ with 20.7 X dAMP added (blue) and hFEN1 in 8.1 mM EDTA with 20.7 x dAMP present (black). Below that are panels (B-E) highlighting specific residues during the dAMP titration. dAMP concentraitons are 0 mM (red), 0.39 mM (maroon), 0.78 mM (orange), 1.6 mM (green), 2.8 mM (cyan), 4.7 mM (teal), 6.2 mM (purple) and 7.8 mM (blue).

4.3c Protein mutation choices in order to enhance spectral quality and sample stability

A potential target for mutation was one of the six cysteine residues in hFEN1, as thiol moieties are known to oxidise when exposed to reactive oxygen or reactive nitrogen species.[112] Oxidation of the thiols could result in much bulkier hydrophilic groups, and might change the conformation of the protein, perhaps unfolding it. As the reducing agent βME enhanced the stability of samples, it was decided to investigate whether removal of cysteine side chains could improve the spectra. Mutation of these residues to a threonine was selected, as this would mimic a hydrogen bonded buried residue, but eliminate the potential of oxidation. ^{15}N labeled proteins of C163T, C182T, C235T and C311T were cloned, expressed and prepared by Dr. David Finger. Comparing the ^{1}H-^{15}N TROSY spectra of the mutants (**Fig. 4.vi**), the main structure of the protein remained unchanged. However, individual mutations failed to stop aggregation. Attempts to express a triple and quadruple cysteine mutated protein resulted in the formation of inclusion body at protein expression stages, perhaps due to incorrect protein folding or degradation.

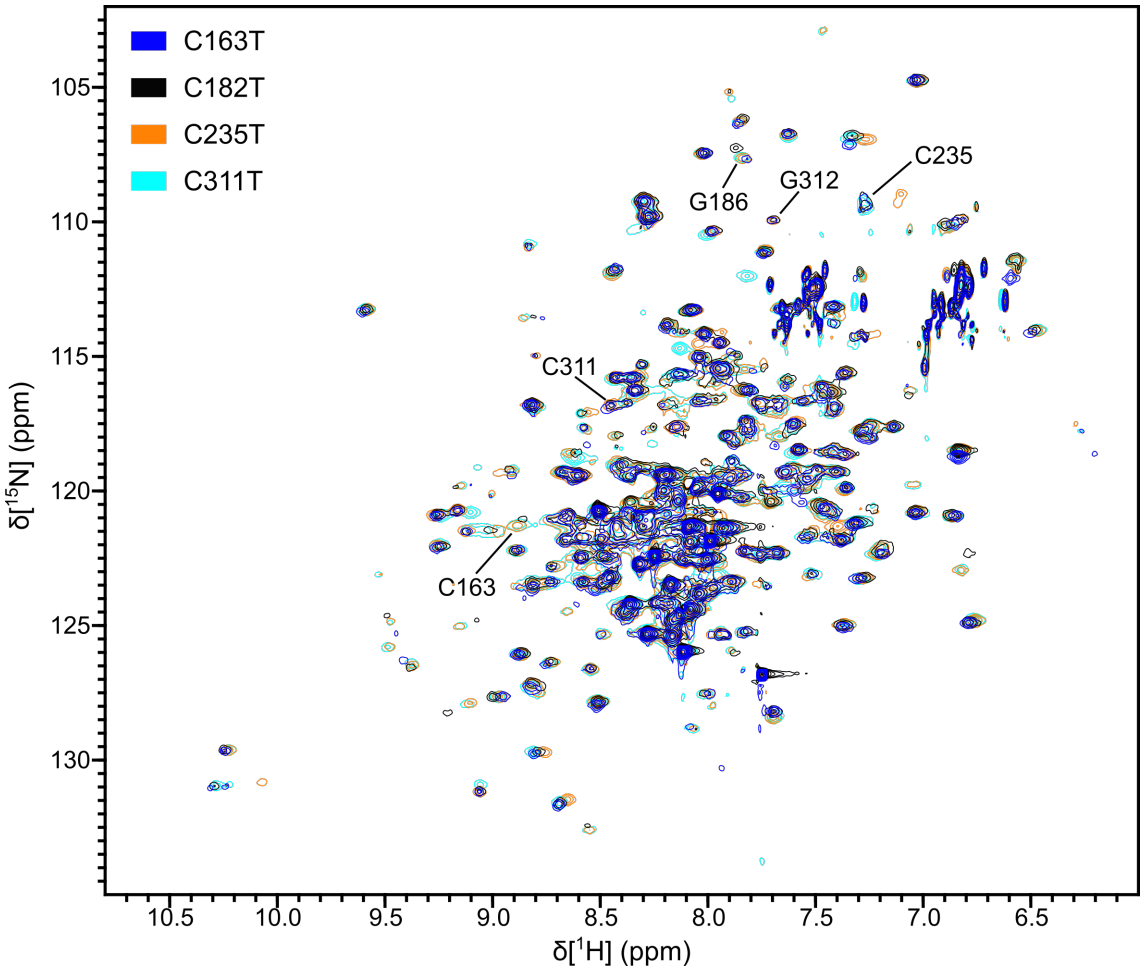

Figure 4.vi. ^{1}H-^{15}N TROSY spectra of four cysteine mutants. Cysteine mutants show very similar overall protein architectures. There are however, local differences present at or near mutated residues, which

are indicated on the spectra. C182 cannot be assigned, but G186 highlights a good change in chemical shift from its normal position.

Another potential choice for mutation was a proline residue, as it can help introduce a new peak in resonance-transferred experiments. A second reason for choosing proline is its ability to isomerise its ω-bond between *cis* and *trans* forms. Prolyl isomerisation has been known to be important in a wide scope of biochemical processes, including protein folding and stability.[115] Furthermore, the residue S187, which is buried in a hydrophobic pocket has been reported to be phosphorylated by Cdk-Cyclin complexes.[116] One possibility is that the adjacent proline 188 might exist in several conformers, allowing the kinase to access the serine for phosphorylation. The isomerisation allows for conformational change in the region around the proline, perhaps causing its residues to be in intermediate exchange, hence a reduction in signal. Furthermore, due to the week timescales of aggregation, a prolyl isomerisation could be the cause of such a slow forming event. Changing the residue to a alanine, which will only reside in the *trans* form, would abolish prolyl-isomerisation, aiding the S:N, and may help separate out peaks in crowded regions.

4.4 Assignment of a novel mutant of hFEN1

4.4a Assignment strategy of P188A mutant hFEN1.

Triply labeled $hFEN1_{P188A}$ protein with ^{15}N, ^{13}C and ^{2}H was required to perform a fresh assignment. $hFEN1_{P188A}$ protein was prepared from plasmids constructed by Dr. David Finger in the same method as reported (**Section 2.1**). The purification procedure allowed for sufficient back-exchange of amide deuterons to protons, as ^{1}H-^{15}N TROSY spectra for both showed no change upon a triple-labeled sample preparation (**Fig. 4.viii**). The yield of ^{15}N, ^{13}C, ^{2}H triply labeled protein from a 500 mL grow-up was 1.4 µmol, or approximately 50 mg of protein. Deuteration of non-labile protons in the protein allowed for reduced dipole-dipole interactions between detected amide protons and adjacent C-H and N-H moieties, resulting in sharper peaks. Six TROSY triple-resonance experiments were performed: HNCO, HNCA, HNCACB, HN(CA)CO, HN(CO)CA and HN(CO)CACB as described in **Section 2.3c**.

4.4b Confirming the wild type hFEN1 assignment by using an independent $hFEN1_{P188A}$ assignment

The data was analysed using CCPNMR analysis software.[88] The data was displayed as strips of specific ^{15}N frequencies (z-axis) with the carbon peaks spread vertically (y-axis) at the correlating

proton frequency (x-axis). Examining the directed experiments reveals what the carbon frequencies for *i*-1 would be, but detected on *i*'s amide residue. Links were found by searching the spectra for non-directed carbon shifts, which match up with the directed experiments but on a different amide residue. Using unique ^{13}C chemical shifts (e.g. alanine or glycine) for both Cα and Cβ, it was possible to identify unique sequences of amino acids which appear only once or twice in the hFEN1 protein sequence (i.e. AAA, GA, AG). Using this technique 73% of the theoretically detectable peaks in the protein were assigned (**Fig. 4.vii**) (BMRB accession code: 27403, **Appendix A1**).

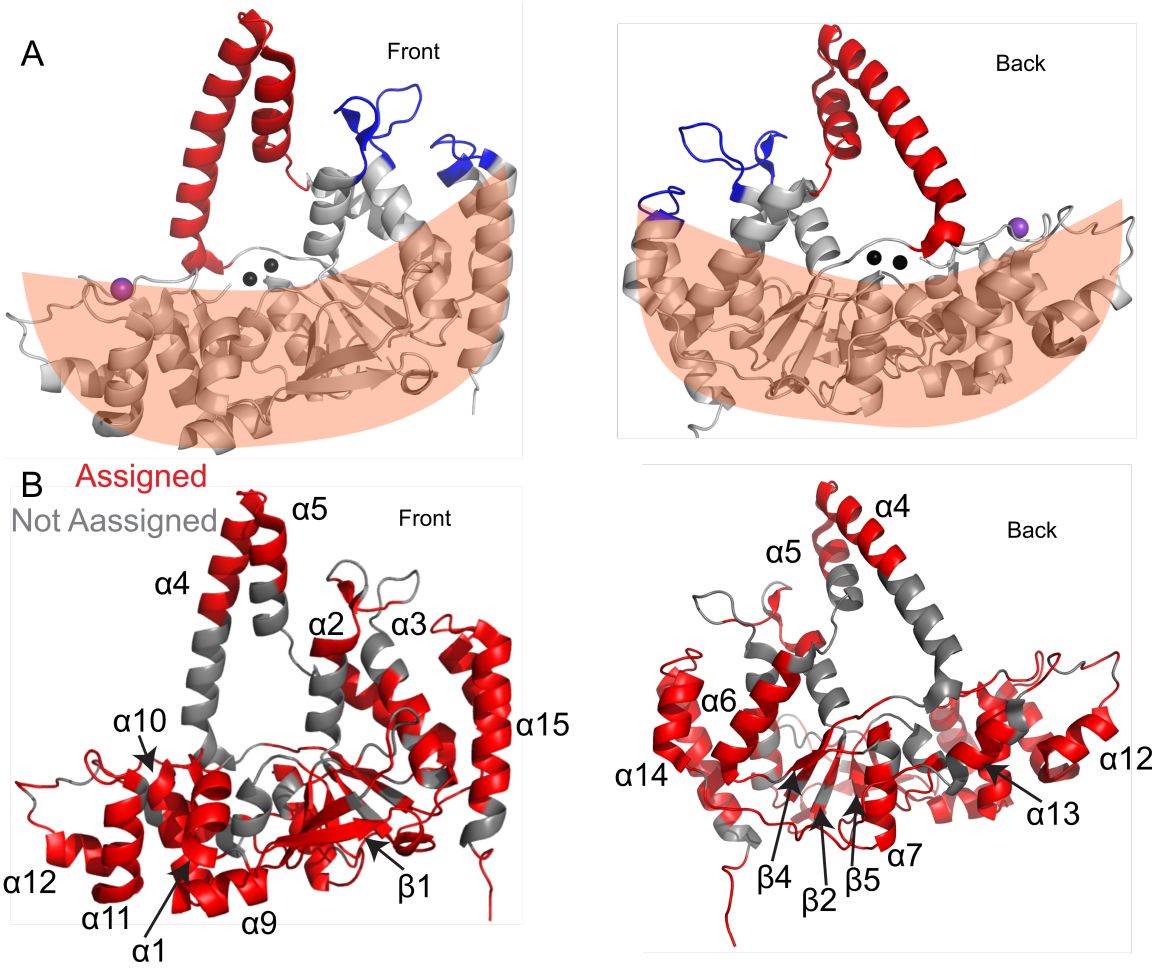

Figure 4.vii. Cartoon representation of the hFEN1$_{P188A}$ molecule to illustrate location of assigned residues. (A) A cartoon showing the 3Q8K structure with important regions highlighted, the arch (red), the α2-α3 loop (blue), the saddle region (orange saddle), the positions of the two catalytic metal ions (black spheres) and the potassium ion in the H2tH motif (purple sphere). (B) The assigned (red) and unassigned (grey) residues are coloured onto the product crystal structure 3Q8K. Secondary structure elements are labeled for clarity.

The hFEN1$_{P188A}$ protein assignment sought to confirm the links from the previous assignment done on WT hFEN1, as well as improve on the 70% assignment, especially in regions of interest (bottom of α4, and residues known to be in the active site **Fig. 4.vii.a**). A similar quantity of assignment was achieved (73%), but a higher degree of assignment confidence on the WT

hFEN1 protein was obtained. New residues were assigned at the middle of the α4 arch (E106 and A105), but the bottom of the arch still remained unassignable. Additionally S210 was assigned at the top-left of the spectrum (**Fig. 4.viii**). Some peaks have been shifted so that they are more dispersed and allow for easier and less ambiguous assignments. Conversely, other peaks like H276, upfield in proton, changed exchange regime and showed a reduced peak signal, especially in the triple resonance experiments.

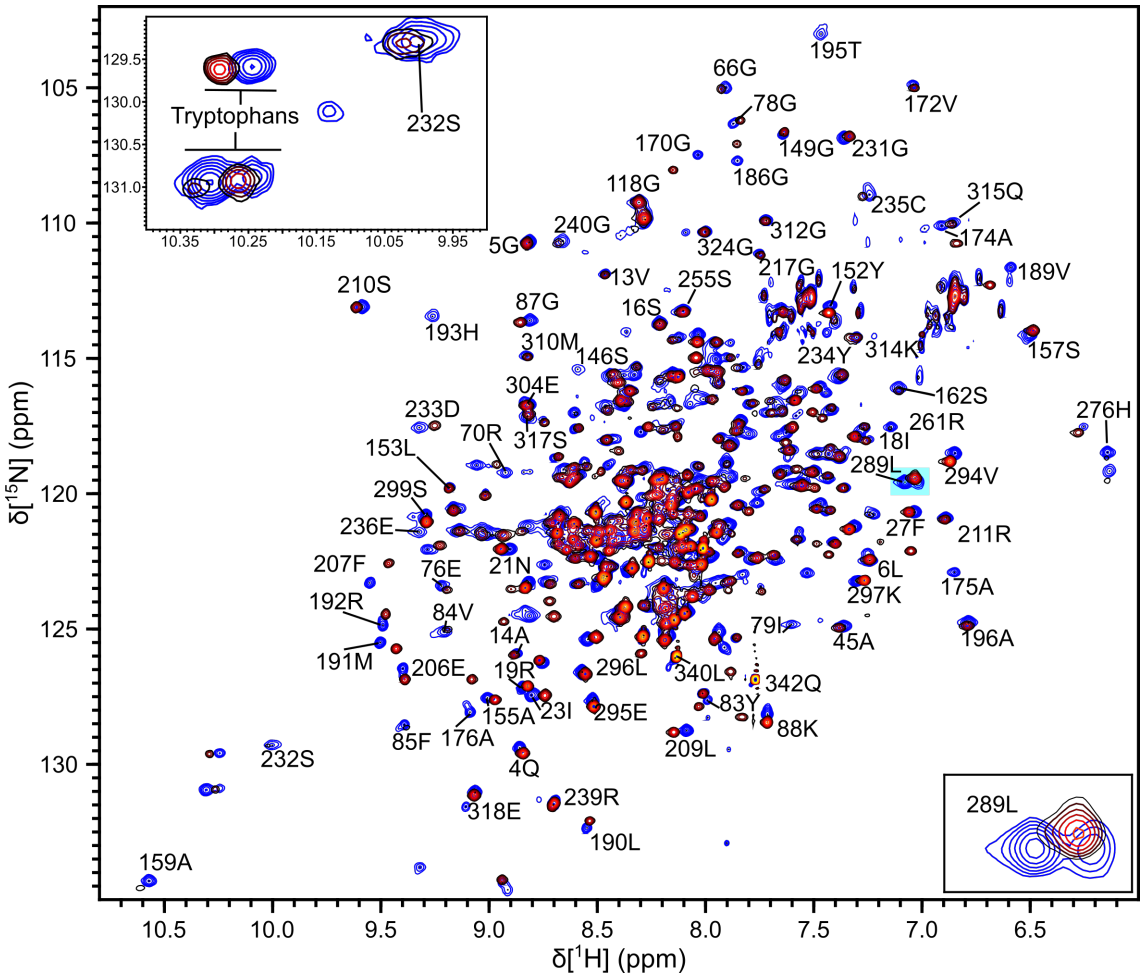

Figure 4.viii. ^{1}H-^{15}N TROSY spectra of hFEN1$_{P188A}$ and wild-type hFEN1. Wild type hFEN1 protein spectrum (blue) overlaid with hFEN1P188A protein spectrum (red). The two tryptophans are highlighted in the top-left panel and 289L is highlighted in the bottom-right panel. Both proteins are perdeuterated.

Overall the P188A mutant helped confirm the assignment already achieved, identified some new peaks on the WT spectrum, as well as identifying regions of interest that are affected by the mutation. Residues such as 289L or tryptophan side chains exhibit distinct slow exchange patterns, resulting in differently populated peaks between WT and P188A protein (**Fig. 4.viii**). This shows signs of conformational motion that suggests these residues were sampling two rotameric states affected by flexibilities elsewhere in the protein. Mutating the P188 residue to alanine mainly resulted in a change of the local conformations of residues in close proximity to

the mutation and did not affect the remainder of the protein structure. Similarly it did not allow for significantly new assignments to be found, yielding no new peak separations. Finally the $hFEN1_{P188A}$ protein has shown that the proline and the conformations associated with it have little to do with the stability of the protein, as it remained equally prone to aggregation as the WT hFEN1 protein. Despite the P188A mutant yielding a new assignment, it did not solve solubility issues. This fact, combined with the issue of unrepresentative conformational dynamics means that the WT hFEN1 protein was chosen for further studies.

4.5 Chemical shift analyses of previously assigned wild type hFEN1 protein

The majority of unassigned peaks were found in the bottom of α4, bottom of α5, the active site, α2 and α3. A peak height analysis showed that the interfaces between assignable and unassignable regions had peaks of decreasing intensity (**Fig. 4.ix**). This implies that the missing peaks were undergoing intermediate exchange, and hence could not be seen in the ^{1}H-^{15}N TROSY spectra. The mutation P188A had no impact on these dynamics, and they are likely an intrinsic property of the protein. In contrast, residues in the α2-α3 loop, some residues in the arch region and the C-terminal tail had greatly increased peak heights. This suggests that these residues are moving on a different timescale to the remaining portion of the protein.

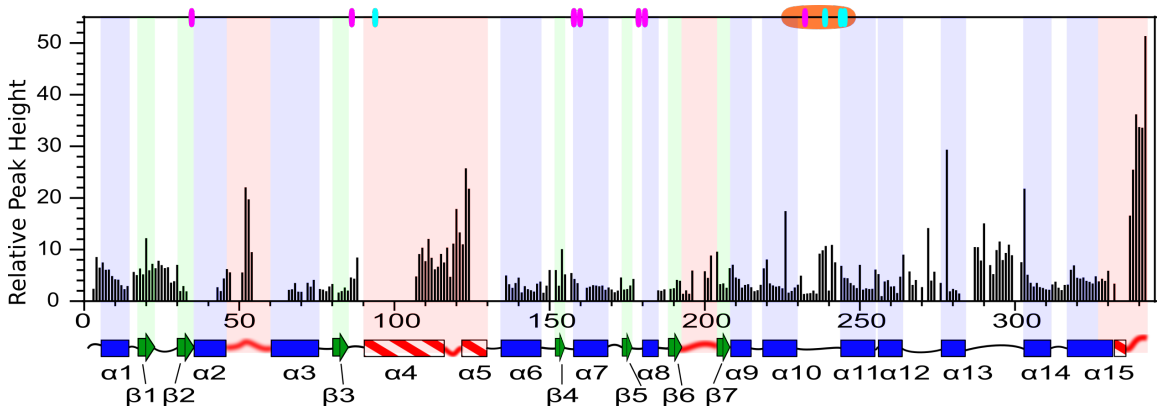

Figure 4.ix. Relative peak height versus residue number. Peak height is relative to the lowest intensity assigned peak (E257) obtained from the wild type hFEN1 protein ^{1}H-^{15}N TROSY spectrum. Secondary structure from the 3Q8K crystal structure is mapped below, with red regions displayed as areas of structural heterogeneity between crystal structures 3Q8K and 1UL1. Active site carboxylates are labeled on the top of the plot in magenta, basic residues that contact DNA in cyan, and the H2tH motif as an orange box.

Chemical shift analysis of all assigned backbone nuclei (^{15}N, $^{1}H^{N}$, $^{13}C^{\alpha}$, $^{13}C^{\beta}$ and $^{13}C'$) was also conducted using TALOS-N.[89] The algorithm predicts the backbone torsion angles ψ and ϕ based on chemical shifts compared to a database of known tripeptide sequences and further filters using an artificial neural network. Thus, it can predict where the residue falls in a Ramachandran

plot, and therefore compute a probability score for α-helix, β-strand or absence of both (loop). The hFEN1 prediction agreed with the dsDNA binding region (saddle region) in crystal structures for both product bound (3Q8K) and apo-hFEN1 (1UL1_X) protein (**Fig. 4.x**). This confirmed further the assignments made on hFEN1 protein.

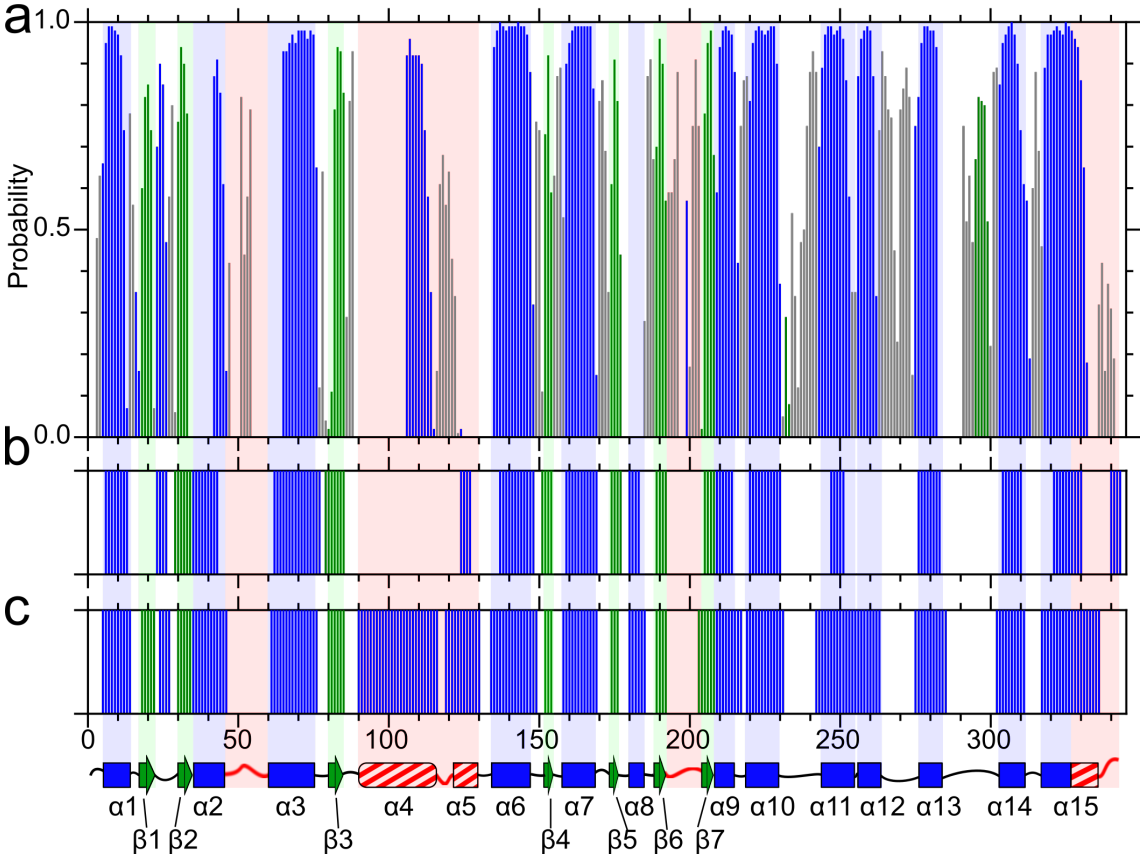

Figure 4.x. TALOS-N predictions of secondary structure. TALOS-N predictions are shown in (A) with the probability of the predictions being correct for α-helix in blue, β-strand in green and absence of both (loop) in grey. The secondary structure mapping of crystal structures 1UL1 (subunit X) (B) and 3Q8K (C) are in the plots below. The secondary structure mapping below is from 3Q8K with heterogeneous sections highlighted in red.

Differences within the prediction and crystal structures were however present. E295-S299 were predicted by TALOS-N to display β-strand ψ and ϕ angles. Upon further inspection of the crystal structures (3Q8K and 1UL1_X), the dihedral angles do in fact loosely agree with β-strand, but crystal structure models display it as a loop. Likewise the β1-β2 loop actually contains one turn of an α-helix, which was so small it was not considered to be of helical nature, but it is present in both the TALOS-N prediction and both crystal structures. The largest unanswered discrepancies are present in the arch region of α4-α5. In apo-hFEN1 protein, there is a lack of density in this arch region, whereas in complex with DNA, there is distinct α-helical nature. The TALOS-N prediction agrees with the DNA bound protein for residues A106-A111, with a

decreasing probability of α-helix for Q112-Q115. However, for the remaining A116-E124 the TALOS-N prediction does not agree with the DNA bound protein X-ray structure; these residues are predicted to not exist in any secondary structure. Thus, only the N-terminus of the arch was consistent with the structure of DNA bound hFEN1, whereas the C-terminal portion of the arch was not.

4.6 Discussion

hFEN1 protein, could be produced and purified in large quantities suitable for NMR spectroscopy, but is a very unstable enzyme, especially at concentrations higher than 100 µM. Optimised purification procedures allowed highly pure samples devoid of nucleic acid contamination. Many procedures were tried in order to afford acceptable spectra while mitigating protein aggregation, and to some degree these were successful. Buffer optimisation gave ideal conditions, the main factors being addition of 100 mM βME, at around 100 mM ionic strength, at pH 7.5, which allowed samples to remain in solution for a month at 25 °C. However, all of the mutations designed to give greater stability, including both cysteine mutants and the P188A mutant, failed to achieve this. Nevertheless, $hFEN1_{P188A}$ protein did help achieve a greater confidence in the assignment.

The assignment of the protein, both WT hFEN1 and P188A, were achieved to a large degree of confidence, although most of the unassigned regions are present in regions of interest (arch and active site residues). This could be due to an intrinsic property of the enzyme, as suggested by peak height analysis, whereby these residues are exchanging in the intermediate regime, resulting in poor S:N. Other peaks, especially residues in the arch, α2-α3 loop and the C-terminal tail are much higher compared to the average of the protein, indicating enhanced conformational dynamics in these regions. The TALOS-N secondary structure prediction further highlights the arch region of the protein as a curious region. While the N-terminal half of the arch (top of the traditional α4) displayed α-helical nature, and thus agreed with DNA bound crystal structures of hFEN1 (3Q8K, 5UM9), the C-terminal half (the traditional α5) did not. This brings into question the nature of this region of the apo-protein in solution, both structurally and dynamically.

Addition of active site probing small molecules, such as hFEN1 inhibitors or dAMP showed that the structure of the protein is changed upon binding. Regions around the active site, such as the protein N-terminus and residues near active site carboxylates changed chemical shift

considerably. More interestingly is the ability of dAMP to change the dynamics of the top of the arch (G118) to an intermediate exchange regime, but only when divalent ions are present in the active site. This suggests a level of connectivity between the active site and the top of the arch, further adding to the question of how the arch is linked to catalysis, despite being quite far away in space.

The arch is a striking feature of hFEN1. Its role in in terms of substrate recognition through threading, and the necessity for its structural integrity from the impact of proline scanning mutagenesis are well-documented.[37] Establishing what structure the arch adopts in solution, or what dynamic movements it has are crucial to understanding the catalytic conformational change involved in the rate-limiting stage of catalysis. Equally, the elusive α2-α3 loop is thought to be involved in recognition of a 1nt 3' flap and must somehow confer this to the active site when in the correct conformation. While predicted to be in the correct secondary structure by TALOS-N, the α2-α3 loop is similar to the arch in terms of peak height and so may display analogous rapid dynamics. This might link the two regions together, allowing for an allosteric control of activity. In both cases however, a more rigorous NMR spectroscopic dynamic measurement of the peaks was needed to elucidate further information.

Chapter 5: Nuclear spin relaxation and model-free analysis of hFEN1

5.1 Protein Dynamics and Exchange using NMR spectroscopy

The large peak height and the lack of predicted TALOS-N secondary structure present in the arch region and the α2-α3 loop indicate these regions of FEN1 are of dynamic interest (**Section 4.5**). Furthermore, crystal structures of both apo-hFEN1 and DNA-bound hFEN1 protein highlight a degree of structural heterogeneity in these two regions that is difficult to explain by a static structure. Looking at B-factors of each crystal structure highlight further the increased root mean displacement of the α2-α3 loop and arch region (**Fig. 5.i**). Furthermore, one other region involved in template DNA binding, the β-pin (β6-β7) is also seen to have higher B-factors, but does not seem to have higher-than-average peak heights. B-factors are not an adequate measure of protein dynamics, as they could be interpreted not only by local mobility in solid crystals, but also by errors involved in fitting the X-ray diffraction pattern. With all this in mind, a more rigorous dynamic analysis of hFEN1 protein in solution is necessary to provide further information about a conformational change in the catalytic cycle. This chapter will cover the experiments performed to produce order parameters for most detectable residues and hence measure the flexibility of most residues in hFEN1.

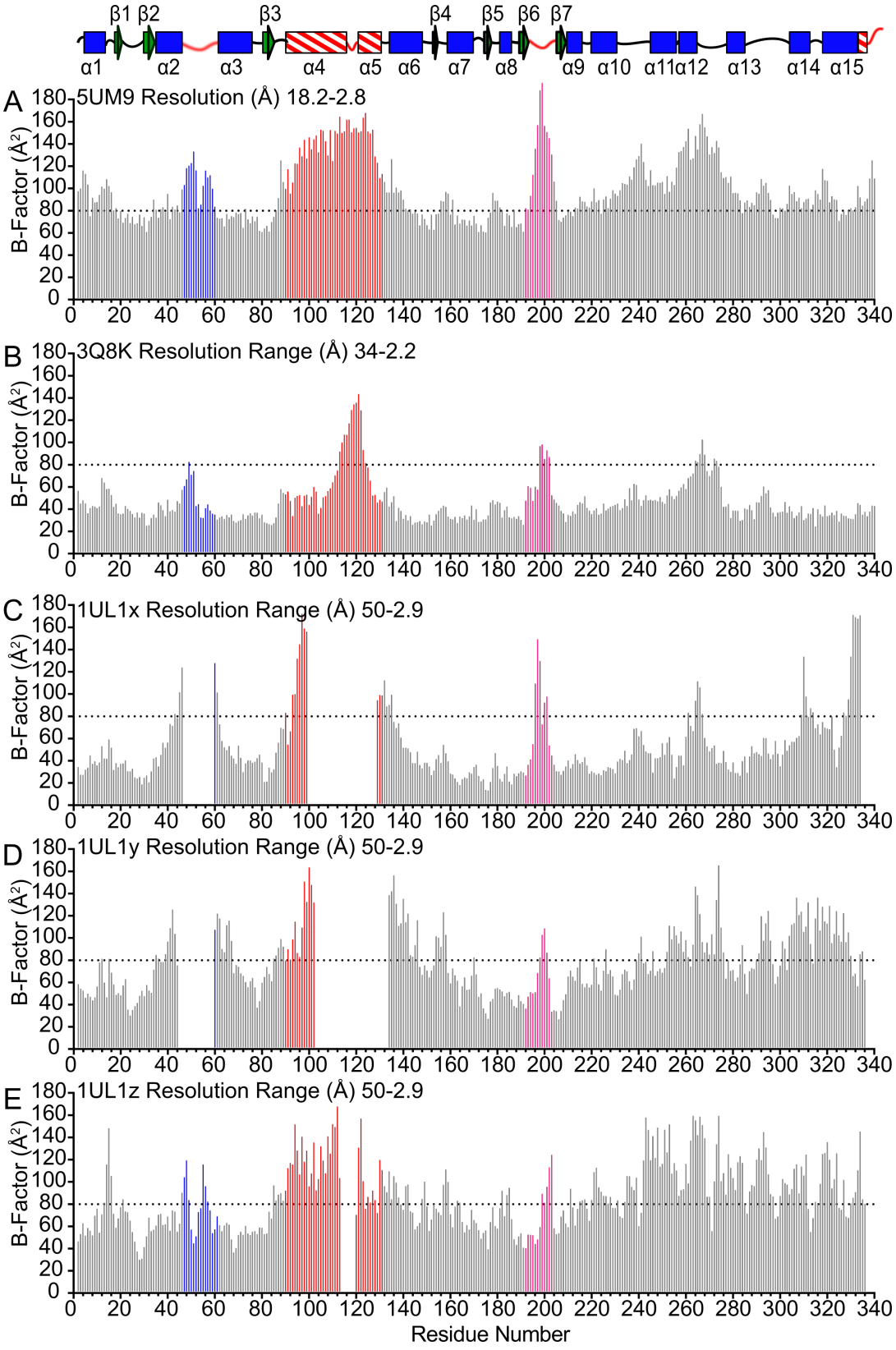

Figure 5.i. Plots of hFEN1 protein sequence highlighting regions of higher than average B-factors. The secondary structure map is taken from the crystal structure 3Q8K, α-helices are blue cylinders, β-strands are green arrows and loops are black lines. Regions which have structural heterogeneity between 3Q8K and 1UL1 are highlighted by red lines and red striped cylinders. The B-factor plots are of (A) 5UM9,[29] (B) 3Q8K,[27] (C) 1UL1x, (D) 1UL1y, (E) 1UL1z.[110] Regions of structural heterogeneity include the α2-α3 loop (blue), the arch region (red) and the β-pin (magenta). A B-factor of 80 Å^2 is highlighted as a dotted line to illustrate a root mean square displacement of 1 Å.

5.2 Nuclear spin relaxation of apo-hFEN1 protein

5.2.a Relaxation measurements of R1, R2 and hNOE values

Using nuclear spin relaxation (NSR) and subsequent model-free analysis NMR spectroscopy techniques, the fast picosecond-nanosecond motions of backbone amides were characterised. First, doubly labeled protein with ^{2}H and ^{15}N was expressed and purified using the same methodology as the triply labeled protein (**Section 2.1**). TROSY based experiments for ^{15}N spin lattice (R_1) and rotating frame ($R_{1\rho}$) relaxation rates, as well as TROSY {^{1}H} heteroNOE (hNOE) were performed at two field strengths (600 MHz and 800 MHz) at 298 K. The R_2 values were calculated from R_1 and $R_{1\rho}$ (**Section 2.3d**). Out of 251 residues that were assigned originally, a full set of R_1, R_2 and hNOE values were obtained for 203 peaks at 600 MHz and 197 peaks at 800 MHz. The reason for the elimination of the data is partially due to the complexity of the hFEN1 protein, and also its unique relaxation profile. Peaks that were weak initially gave very poor exponential fits, likely due to intermediate exchange broadening and some peaks were merged with other peaks, affording ambiguous and mixed relaxation rates. Therefore, these peaks were eliminated from further analysis, as they were unrepresentative of their actual relaxation values. Model-free analysis needs all three values of R_1 R_2 and hNOE at two field strengths, and therefore many incomplete datasets were rejected from the analysis.

The data however, show a clear pattern (**Fig. 5.ii**). Figures from the 600 MHz data showed that there is an average relaxation rate or hNOE value, which is shared by most of the residues of the protein ($R_1 \approx 0.5\ s^{-1}$, $R_2 \approx 30\ s^{-1}$ and hNOE ≈ +0.8). Using the average R_2/R_1 ratio shared by most of the residues of about 60, an overall estimation of the τ_c was found to be around 25 ns,[117] which is consistent with a protein of its size (38 kDa).[118] Loops like the β-pin (β6-β7) (46 – 38), α12-α13 (55 – 42), α10-α11 (54 – 46) and the large α13-α14 loop (52 – 27) display a lower than average R_2/R_1 ratio. The lowest R_2/R_1 ratios were found in the arch region (24 – 3.7), the α2-α3 loop (41 – 8.4) and the non-native C-terminal tail (13 – 1.6). While the values were different, the exact same patterns were found in the 800 MHz data, although the data was noisier. The regions with lowered R_2/R_1 ratios and more negative hNOE values correspond to regions with much higher B-factors, or missing electron density in crystal structures, and regions of higher peak intensities in the assignment.

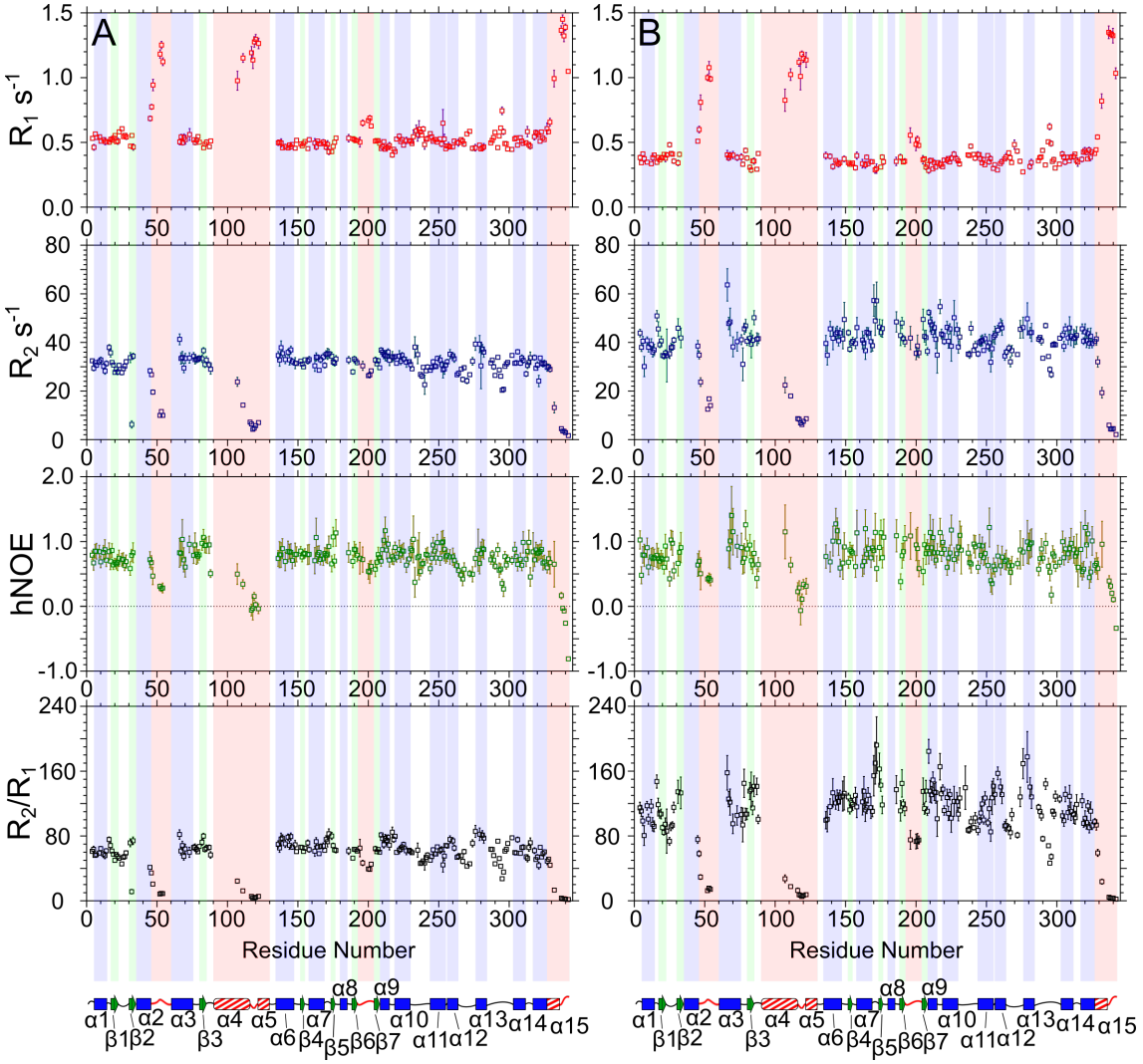

Figure 5.ii. Experimentally determined relaxation values and hNOE values versus residue number with secondary structure map below. (A) Shows the data from 600 MHz and (B) the data from 800 MHz. Spin-lattice (R_1 in red), Spin-spin (R_2 in blue) and ^{15}N-{^{1}H} NOE (hNOE) values are plotted at the two different field strengths. Also derived from the data are the R_2/R_1 values below in black.

5.2.b Model-free analysis using *relax*

With 197 complete relaxation datasets, the data was subjected to model-free analysis using the software *relax*.[119] The software calculates the diffusion tensor and the correlation time (τ_c) for the whole protein, while in parallel also calculating the effective correlation time (τ_e) for each residue. Both parameters are calculated together, and are then validated using statistical analysis, resulting in a more reliable and accurate calculation of model-free parameters. Model-free parameters are fit to four different diffusion tensor models: isotropic spherical, prolate spheroid, oblate spheroid and ellipsoid. Since amide bond angles were needed for the spheroidal tensors, the 3D coordinates of nitrogen and proton were provided using the 3Q8K protein structure. Calculations excluded a further 18 residues, most likely due to high errors in the data. Initially, the calculations ran into problems and the predictions failed to converge to a

solution. This was proposed to be due to the fact that the arch, α2-α3 loop and C-terminal tail were moving on a timescale so different from the bulk of the protein, the diffusion tensor model failed to agree on a unified solution.

With this in mind, the 16 residues that had no secondary structure in the 1UL1 (x subunit) were omitted from calculations. This included residues in the arch (109-129), the α2-α3 loop (46-59), the C-terminal tail (337-342) and the β-pin (199, 200). Subsequently *relax* managed to converge on all diffusion tensors, and chose, by the lowest Akaike information criteria (AIC), the oblate spheroid tensor (**Table 11, Fig. 5.iii**). The tensor seems to fit with the overall shape of the hFEN1 molecule and best describes the tumbling of the molecule in solution. The remaining 16 residues with no discernable secondary structure were run through *relax* again, with both the oblate spheroid and spherical diffusion tensors tested, and the spherical diffusion tensor was chosen overwhelmingly by AIC (**Table 11**). This hybrid approach allowed all 179 residues to be analysed, despite the wildly different dynamic properties of certain residues.

Structured Diffusion Tensor	AIC Value	Unstructured Diffusion Tensor	AIC Value
Spherical	1483.9	**Spherical**	182.15
Prolate Spheroid	1508.3	-	-
Oblate Spheroid	1479.2	**Oblate Spheroid**	1535.3
Ellipsoid	1509.7	-	-

Table 11. Diffusion tensor statistical selection process using AIC. Analysis includes the 163 residues in discernable structured regions of the 1UL1-x subunit crystal structure (left), and the 16 remaining unstructured residues (right). There is a slight preference for oblate spheroid over complete sphere in the structured region, whereas there is overwhelmingly a spherical selection for the 16 unstructured residues.

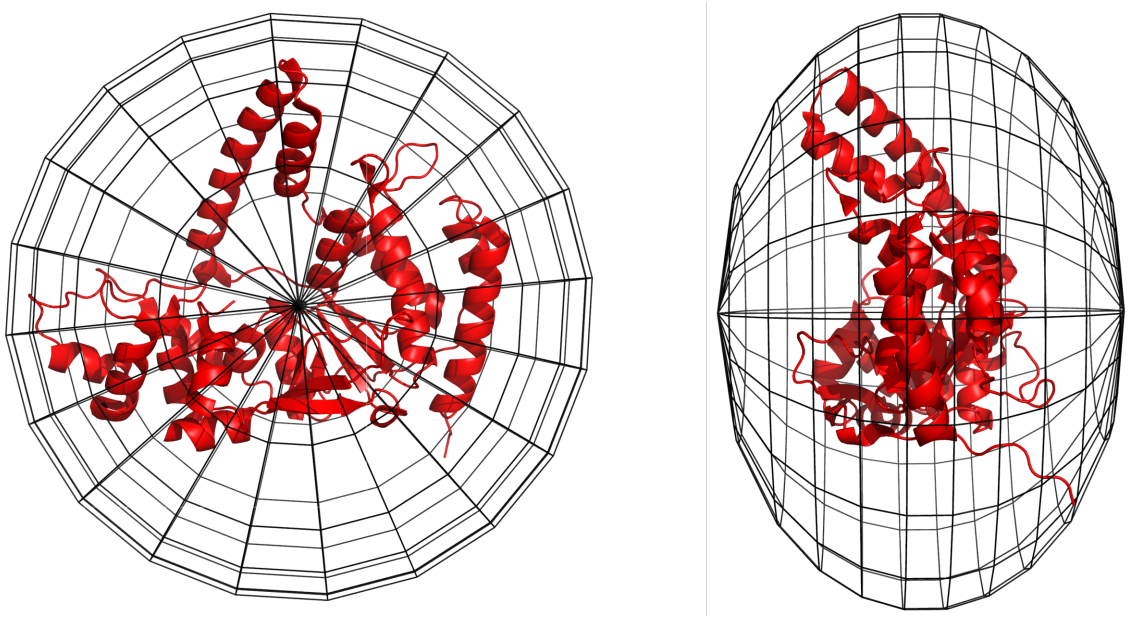

Figure 5.iii. View of the diffusion tensor overlaid upon the hFEN1 crystal structure 3Q8K. The tensor shows good agreement with the structure of the protein, suggesting that the fast tumbling of the molecule are towards the X and Y-axes (long axes).

5.1.c Analysis of Model-free parameters for the majority of the protein

Order parameters (S^2) were obtained from the residues and showed a distinct pattern (**Fig. 5.iv**). Most backbone amides were well described by the overall tumbling of the molecule with S^2 values of around 0.8. Loops, especially in the C-terminal half of the protein (215-342) displayed on average lower S^2 values. There were minor drops to around 0.7 for example in α6-β4 and the β-pin. Much lower and noisier S^2 values of 0.6-0.2 were seen for loops like α12-α13 and α13-α14. The errors on this data are most likely due to errors in the *relax* fitting, as the errors on the raw data are no worse for these residues. It should be noted that the β-pin displays medium R_2/R_1 values of 65 – 39, but the model-free analysis indicates it is not moving on a ps-ns timescale. Furthermore a large number of residues have R_{ex} terms, which account for 71% of all fitted residues (**Fig. 5.v, Table 12**). The R_{ex} parameters are designed to account for chemical exchange in the model-free analysis and offset larger than expected R_2 values. Thus, for 71% of the analysed residues, some kind of chemical exchange is expected, and probably accounts for some kind of relatively slow motion occurring in the protein on the millisecond timescale. Overall, this implies that hFEN1 has a much more complicated dynamic dsDNA interacting region, with millisecond motions present in loops.

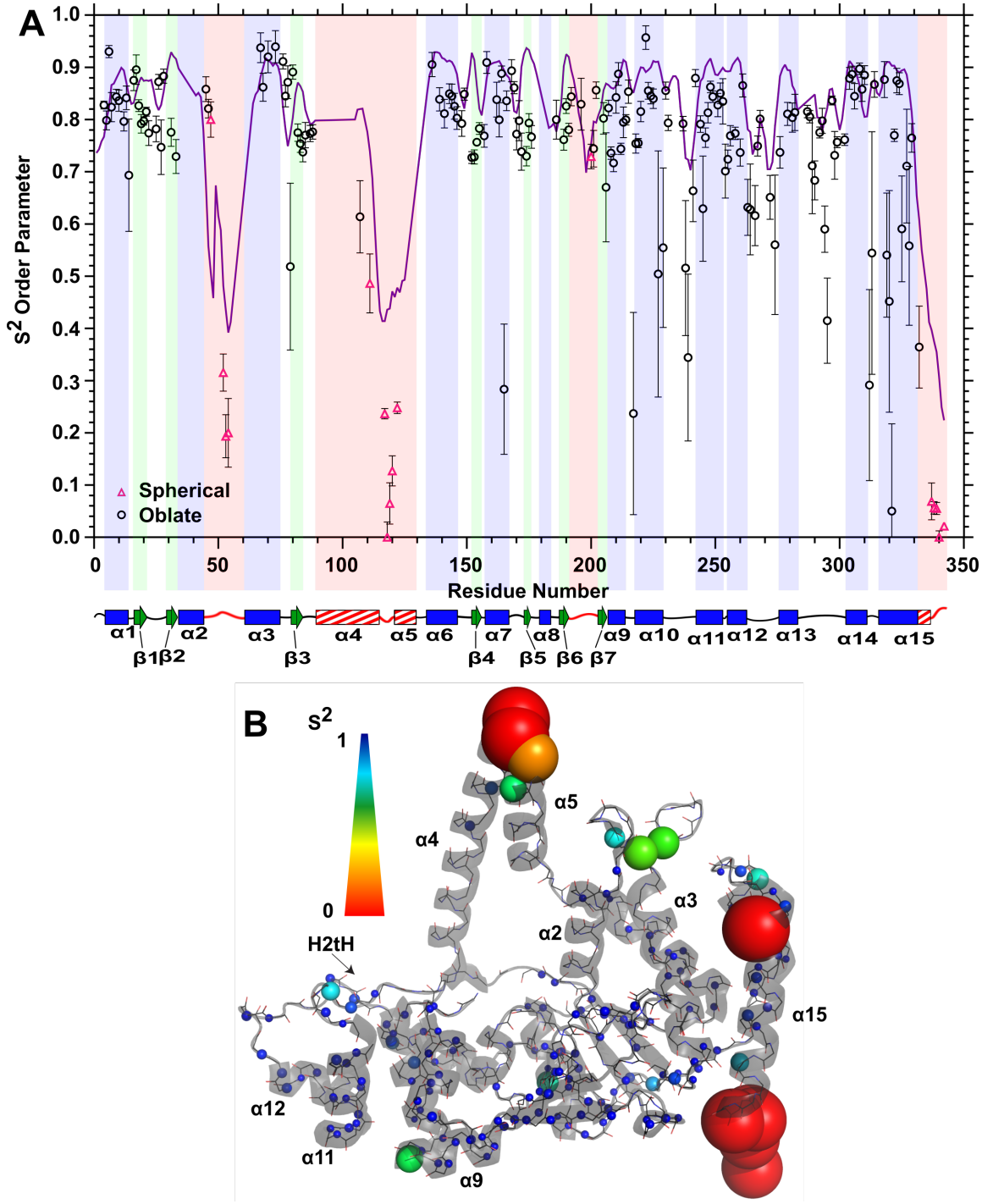

Figure 5.iv. Model-free analysis of the relaxation data of hFEN1 showing regions of increased flexibility. (A) The order parameter (S^2) values versus residue number with a secondary structure map below as seen in previous figures. Black circles represent residues analysed fully by *relax*, while pink triangles are residues analysed by an isotropic spherical diffusion tensor. The purple line represents predicted S^2 values from the TALOS-N random coil index prediction.[89] (B) S^2 values plotted on the backbone amide nitrogens of the hFEN1 crystal structure (3Q8K). The spectrum bar on the left represents the colour and size of each residue as determined by the S^2 value.

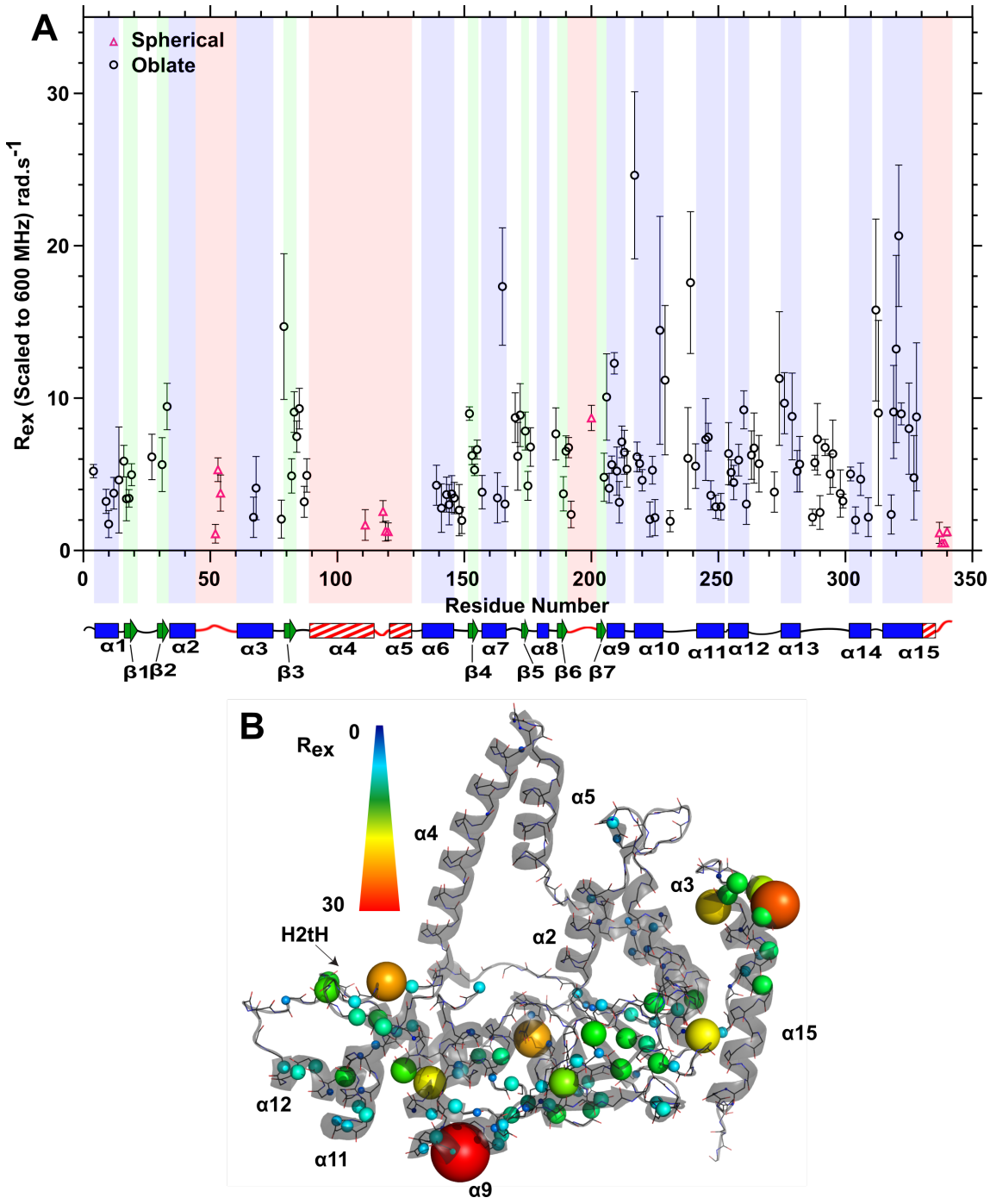

Figure 5.v. Model-free chemical exchange parameters (R_{ex}). (A) The chemical exchange (R_{ex}) values versus residue number with a secondary structure map below as seen in previous figures. Black circles represent residues analysed fully by *relax*, while pink triangles are residues analysed by an isotropic spherical diffusion tensor. (B) S^2 values plotted on the backbone amide nitrogens of the hFEN1 crystal structure (3Q8K). The spectrum bar on the left represents the colour and size of each residue as determined by the R_{ex} value. 70% of all analysed residues had a R_{ex} value assigned to them. The highest values were present in the saddle region of the protein indicating a slow millisecond timescale of motion for the majority of residues.

Model-Free Models	Number of Residues Selected for Each Model
m1 = {S^2}	21
m2 = {S^2, τ_e}	11
m3 = {S^2, R_{ex}}	62
m4 = {S^2, τ_e, R_{ex}}	24
m5 = {S^2, S^2_f, τ_s}	18
m6 = {S^2, τ_f, S^2_f, τ_s}	2
m7 = {S^2, S^2_f, τ_s, R_{ex}}	28
m8 = {S^2, τ_f, S^2_f, τ_s, R_{ex}}	13
TOTAL	179

Table 12. Model-free model selection for all 179 analysed residues. 70% of all residues were fitted with a R_{ex} term present, and roughly one third of all residues were in the m3 model.

5.2.d Intrinsically Disordered Regions in arch, α2-α3 loop and C-terminal tail

The lowest reliable S^2 parameters however are present in the α2-α3 loop, the top of α4 and α5 and the C-terminal tail. These regions display a noticeable difference in the timescales of motion from the majority of the protein, consistent with the relaxation data. The mobility of the C-terminus of the protein (339-342) can be rationalised by the fact that it is a non-native tail region, devoid of secondary structure and exposed to solvent. There are only three data points for the loop, but they are all indicate a S^2 value of ~ 0.2. It could be that a large, solvent exposed loop could result in faster motions for these residues. Finally the S^2 values in α4 drastically decrease from about 0.6 to > 0.1 and is at its lowest at the small loop at the top of the arch, then further increases to 0.2 at the top of α5. Having such low S^2 values means that these residues are very poorly described by the motion of the molecule and are much better described by the fast picosecond-nanosecond internal motions. These regions are therefore postulated to be disordered regions in apo-hFEN1 in solution.

To confirm the experimental data, chemical shift analysis using Random Coil Index-S^2 (RCI-S^2) values were predicted using TALOS-N.[89] There was good agreement with the model-free analysis, although the prediction underestimates the lower S^2 values (**Fig. 5.iv**). This further confirms the theory that these are intrinsically disordered regions, as coded by the primary sequence of the protein.

5.2.e Classification of Intrinsically Disordered Regions (IDR)

To further classify the intrinsic disorder present in these regions, the differences between experimental chemical shift values and predictions using the neighbor corrected random coil chemical shift library for intrinsically disordered proteins (ncIDP) was investigated.[120] The ncIDP predictions of the entire of hFEN1 protein for ^{15}N and Cα, show, for most of the structured

region of the protein, a large discrepancy from the experimental values. However, the C-terminal tail, arch region and the α2-α3 loop have chemical shifts close in value ncIDP predicted chemical shifts, which is a further indication that these loops are disordered in the absence of substrate. Although disordered, a closer look at ^{15}N and Cα resonances can be a powerful indicator of local dynamic preferences for certain secondary structures. **Figure 5.vi** shows that the residues in the α2-α3 loop and the C-terminus are different between experimental and predicted values of ≥1, indicating a good agreement with an extended conformation. However, the arch region is more complex. The N-terminal α portion of α4, from residue A107-Q113 exhibits differences of ≤1, which indicates that this region prefers α-helical φ-ψ space. There is a gradual decrease in α-helical propensity from Q110-A114, and from residues A114-A120 there is a switch to extended conformation. Furthermore AGADIR predictions of percent helicity using primary amino acid sequences (**Fig. 5.vi.c**)[121] and the TALOS-N data (**Fig. 4.x**) of secondary structure prediction both agree with this data. This suggests that α4 is poised to form α-helical structure, whereas α5 is not.

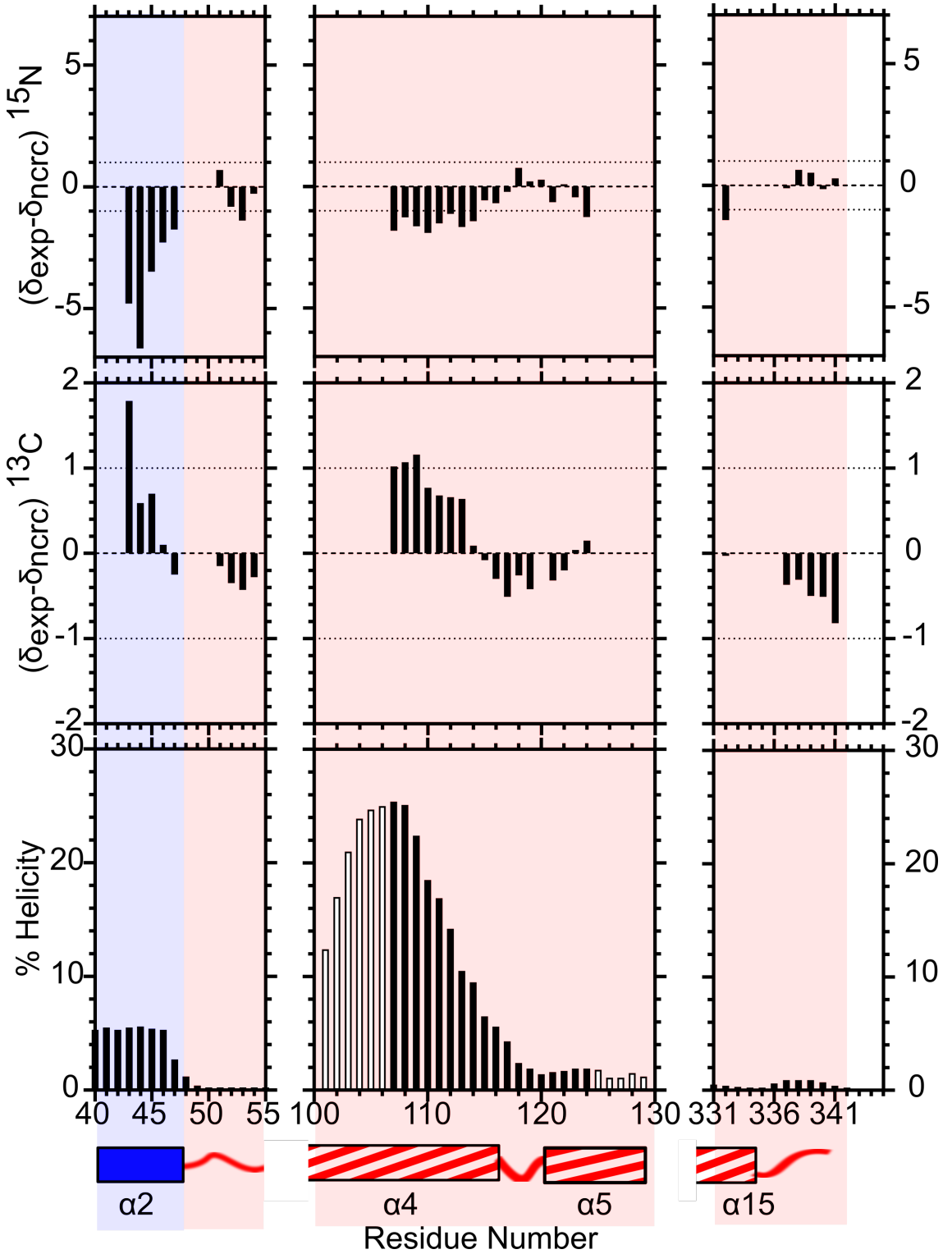

Figure 5.vi. Chemical shift analyses and AGADIR predictions for predicted disordered regions. Differences between neighbour corrected random coil and experimental chemical shifts for ^{15}N (A) and ^{13}Cα (B). Values of less than one are in good agreement of extended unstructured regions, whereas negative values for 15N and positive values for ^{13}Cα indicate helical structure. C) An AGADIR prediction using protein primary structure to determine % helicity.[121] Hollow bars indicate information where there is no NMR data present. α2 dramatically decreases its helical propensity as it reaches the C-terminal portion, whereas α4 starts with much higher propensity then decreases equally quickly. Both α5 and the final C-terminal tail have no α-helical nature.

Further analysis of the arch region was done using an algorithm called the classification of intrinsically disordered ensemble regions (CIDER). The software uses charge distribution of the primary sequence to predict how extended or collapsed the disordered region is.[122] Analysis of residues P90-L130 using CIDER showed that this region was a polyampholytic coil or hairpin. The linear sequence distribution of charged residues is assessed by a value of self-assembly κ, with a value of 1 indicating full collapse and 0 indicating no self-assembly. The arch region of hFEN1 has a very low κ value of 0.12, suggesting that the loop prefers to exist in an extended conformation in solution.

5.2.f Implications of having intrinsically disordered regions

The arch and the α2-α3 loop are moving on timescales completely separated from the tumbling of the protein. At first glance it seems counterintuitive to have disordered sections of an enzyme involved in catalysis. The structure of the enzyme when bound to DNA would indicate that active site residues in the arch, such as R100 and K93, have to be positioned correctly in order for reaction to occur. In other words the structure of the enzyme relates to its function as a nuclease. Positional ambiguity for these essential catalytic residues should slow down the reaction, as seen with mutants where these key catalytic residues are mutated to alanines.[37,38] However, there may be benefits to having such disordered regions present in the protein.

Firstly, since the arch most likely exists in an extended conformation, it allows for a large capture area of any potential 5' flaps. It is difficult to envision a 5' ssDNA flap going through the small gap presented in the ordered crystal structure in the absence of an energy source. By having the gap much larger, and disordered there is a higher chance of the threading process happening spontaneously. Thus, when the DNA is bound, a disordered arch increases the probability of a successful threading event occurring thereby hastening the reaction. Furthermore, it allows for accommodation of DNA with some degree of secondary structure, which is consistent with kinetic studies with a gapped substrate.[22,38] This might be vital in for DNA replication, as it allows for a degree of flexibility in the type of flap substrate processed by hFEN1.

Once the DNA is captured an ordering event must occur to place active site residues into the correct position, as well as forcing the DNA to adopt a conformer amenable for hydrolysis (i.e. conformational change). A disadvantage of having a disorder-to-order transition into a catalytically active state is that this has a high entropic penalty. Thus, the assembly of the entire

arch would be expected to have a high energy cost (large energy barrier), therefore it would be expected to be a slow process. However, this would be inconsistent with the speed of the single turnover rate of hFEN1 (milliseconds). The entropic cost of ordering the entire arch would be very high; however having α4 already transiently sampling α-helical conformations reduces the penalty. This allows for a much smaller conformational shift of ensembles with low energy barriers, allowing for rapid folding. Such triggers for a conformational shift could be the formation of the mini-hydrophobic core at the top of the arch, and contacts with the template DNA with residues like R104. The α5 is much smaller and can be potentially induced to form a α-helix when then entropic penalty is balanced by enthalpic gains from external contacts. Such contacts could include interactions from the ordered α2-α3 loop (R47) and contacts with the template DNA (K125, K128 and R129). Thus using a combination of conformational ensemble shifts (α4) and induced helix formation (α5), a finely tuned fast ordering process with low energy barriers can be achieved.

Finally, there is the question of specificity. The induction of α5 is only energetically favourable when the cues from the correct substrate are present. The positioning of the 1nt 3' flap into α2-α3 forms an omega loop (Ω-loop) formation. This non-regular secondary structure, is frequently involved in catalysis, folding and stability, just like in the example of TIM (**Section1.5e**).[39] Positioning of the template portion of the DNA and the 5' flap forms contacts in the arch and help induce a change in α5. Having the whole system as an induced fit would result in great specificity, but the energetic penalties would result in very slow folding. Adopting a dual conformational ensemble change and induced fit system allows the specificity to remain high enough, while keeping the energetics of folding as flat as possible. It also allows for rapid unfolding and release of substrate that is non-ideal. Furthermore, it would have a low affinity for the catalytically-active conformation which would allow for fast dissociation of product DNA, and hence aid catalytic turnover.

5.2.g Reasons that relaxation dispersion of hFEN1 is currently unattainable.

One of the aims was to analyse the rates of motion between a potential ordered and disordered state of the protein, and to link this rate to the single turnover rate of hFEN1. Model-free analysis identified regions of disorder, but cannot provide rates of motion between an ordered and disordered state. R_{ex} terms for 70% of the protein indicate that some kind of exchange is taking place, so relaxation dispersion was thought to be a valid experiment to quantify the exchange processes taking place. Therefore to screen for any residues undergoing a potential

two state exchange process, differences in ^{15}N chemical shift were measured using HSQC and HMQC experiments at 600 MHz and 800 MHz. This procedure can be applied to ascertain residues undergoing exchange.[123] However the experiments were inconclusive showing difference in the signs of chemical shift difference between the two field strengths. This could be due to poor spectral resolution. Furthermore, measuring relaxation dispersion by CPMG or $R_{1\rho}$ is time-consuming, during which the hFEN1 protein was liable to precipitate. These two factors combined prevented acquisition of relaxation dispersion profiles.

5.3 Discussion

The arch region and the α2-α3 loop were both confirmed as regions of intrinsic disorder. The intrinsic disorder is though to aid in capture of a 5' ssDNA flap and hence aid threading. Analysis of the disordered arch shows that the α4 portion is transiently forming α-helix, while α5 is not. This allows for a mix of a conformational selection of helix in α4 combined with an induced fit of helix in α5. These ordering processes are both made energetically favorable by the correct substrate. These both help explain the catastrophic rate reductions in arch proline mutants of hFEN1 protein.[37] Another induced fit event potentially accounts for the molding of the α2-α3 Ω-loop around a 3' flap. It is known that the 3' flap enhances the reaction rate considerably.[22] The disorder provides low-energy barriers to folding, hence expediting catalysis whilst maintaining the ability to fold only when bound to a correct substrate, hence enhancing specificity. Finally, having disorder as the default state of the hFEN1 protein would allow easier unfolding and thus dissociation from the bound dsDNA, hence assisting turnover. The utility of intrinsic disorder for the highly structure specific nuclease hFEN1 protein can therefore be seen.

Direct evidence of the timescales of motion for the arch and α2-α3 loop could not be obtained through relaxation dispersion. The residues in question simply do not display adequate spectra in order to obtain the data. There is also the lack of data for residues of interest flanking the disordered regions, presumably due to intermediate exchange, which prevents analysis of key regions of interest. The dsDNA binding region of the protein is exhibiting millisecond timescale motions, as is seen with the ubiquity of R_{ex} terms. It is unknown what effect this has on the protein but it is consistent with the tendency to precipitate and the very uncooperative unfolding of hFEN1 in urea.[37] For the moment however it can be assumed that the arch and α2-α3 loop are the main regions that are linked to the proposed conformational change involved in catalysis.

There are several other factors involved in hFEN1 catalysis, including the addition of divalent cations and the DNA conformational changes that must occur. While unable to inspect potential exchange states using relaxation dispersion, looking directly at a DNA-hFEN1 bound complex will highlight dynamics when bound to the substrate. Furthermore, all NSR experiments have been conducted in the presence of 0.1 mM EDTA. However, there are chemical shift changes in ^{1}H-^{15}N TROSY spectra upon binding of either Mg^{2+} or Ca^{2+} ions, which are essential for catalysis and DNA conformational changes.[28] By looking further at hFEN1-DNA complexes and the role that divalent cations play in the whole of the protein, further information can be gained about the dynamical tuning of hFEN1 protein.

Chapter 6: Investigation of substrate bound hFEN1 complexes by NMR spectroscopy.

6.1 Analysing changes in dynamics upon binding DNA.

The ordering of the highly flexible arch and the α2- α3 loop are thought to occur when all substrate feature selection steps are passed. The main questions are therefore, how fast does this ordering occur, and at which timescale are the residues moving when bound to the substrate? In an attempt to answer these questions a protein-DNA complex was analysed by NMR spectroscopy. In particular, a DNA titration of hFEN1 was desirable in order to pinpoint residues that are affected by binding of substrate. Of particular interest are the arch region and the α2- α3 loop. Do they change timescales in the presence of substrate DNA? If so then it might become apparent by chemical shift perturbation mapping the DNA titration. Furthermore, if a stable complex can be made, then the protein could be assigned to further studies. A subsequent nuclear spin relaxation (NSR) analysis could be performed on the hFEN1-DNA complex, assessing the fast flexibilities of the entire protein. This could be compared to the flexibilities of apo-hFEN1, in particular the outcomes of the arch and α2-α3 loop. Although untenable for the apo-protein, relaxation dispersion might be able to be used in order to quantify μs-ms motions of exchange for residues on a protein-DNA complex. In this chapter, necessary conditions to form a DNA-protein complex will be discussed, as well as the implications of dynamics of the arch and α2- α3 loop when bound to DNA.

6.2 Assignment and chemical exchange dynamics of a substrate bound hFEN1 complex.

6.2a hFEN1 K93A mutant and DNA titration in the presence of EDTA.

To study DNA binding to hFEN1 protein, it was necessary to form catalytically inert complexes. hFEN1 is notoriously promiscuous when it comes to DNA hydrolysis, and over NMR experimental timescales would probably hydrolyse even non-ideal DNA samples. Therefore a mutant of hFEN1 protein was chosen to hinder hydrolysis, while keeping the structure and function of the protein the same as wild type. Lysine 93 (K93) is an essential amino acid present in the bottom of α4, and has been observed to interact with the scissile phosphate of the duplex in the active site. Mutated to an alanine, the protein $hFEN1_{K93A}$ shows a dramatic decrease in both single- and multiple-turnover reactions.[37,38] The protein however shows little change in its interaction with DNA. FRET experiments to detect DNA bending report minor changes in the

K_{Bend} of $hFEN1_{K93A}$ compared to wild type protein (K_{Bend} of 8 ± 1 for K93A vs 7 ± 1 for wild type). Equally, exciton coupled circular dicrhoism (ECCD) experiments that measure DNA conformational change on addition of metal ions report similar signals to that of wild type, suggesting that the protein can deform the DNA into the active site.[28] A ^{1}H-^{15}N TROSY spectrum of $hFEN1_{K93A}$ in the presence of 0.1 mM EDTA (**Fig. 6.i**) was taken that shows minor chemical shift changes in residues proximal to Lysine 93, for example, residues like K88 and C235. Leucine 289, which is in slow exchange in wild type hFEN1, moves predominantly to one conformer. The remainder of the protein is unchanged compared to wild type. Due to its similar ability to accommodate DNA conformational changes but single-turnover rate of almost 2000 times lower than wild type, $hFEN1_{K93A}$ was chosen for all protein-DNA complex experiments.

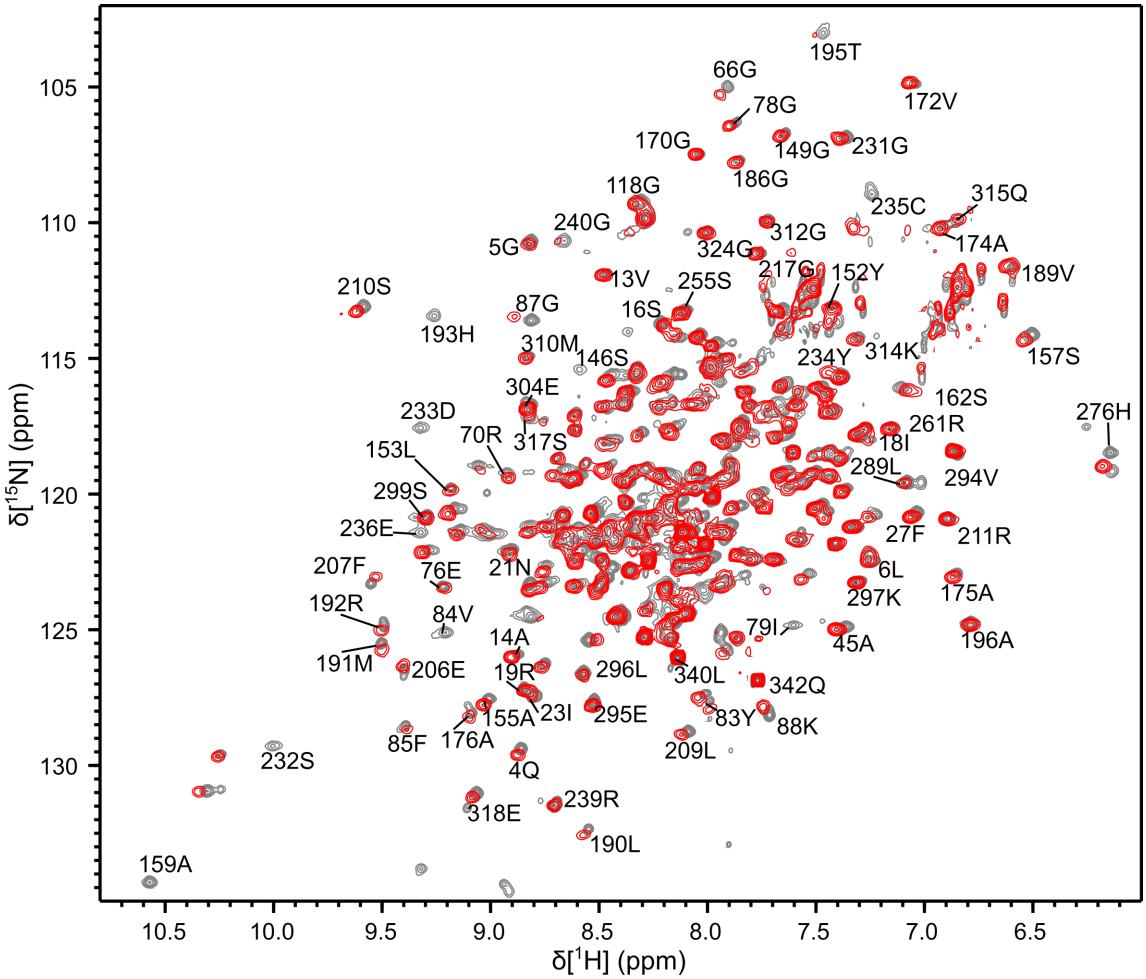

Figure 6.i. ^{1}H-^{15}N TROSY spectra highlighting differences between wild-type hFEN1 and $hFEN1_{K93A}$. Wild-type protein is shown in grey and $hFEN1_{K93A}$ is shown in red. Both spectra were recorded in the presence of 0.1 mM EDTA. A selection of more dispersed peaks are displayed on the spectrum of the wild-type protein. It should be noted that the grey spectra is taken with perdeuterated protein.

A DNA titration was performed with $hFEN1_{K93A}$ to see structural and dynamic changes upon substrate binding. The DNA chosen to bind to $hFEN1_{K93A}$ was an ideal double-flapped substrate

with two hairpins on the upstream and downstream regions (dual hairpin substrate: DHPS1, **Fig. 6.ii**). A unimolecular substrate construction ensured the protein-DNA interactions were limited to the bent junction, as there is only one 5' terminus and 3' flap present. The ^{15}N-labeled hFEN1$_{K93A}$ protein was prepared and purified as described in **section 2.1**. From 1 L of media the yield of ^{15}N hFEN1$_{K93A}$ protein was 1.4 µmol or approximately 53 mg. DHPS1 was titrated into a 500 µM ^{15}N-labeled hFEN1$_{K93A}$ protein sample in the presence of 0.1 mM EDTA and ^{1}H-^{15}N TROSY spectra were taken at each titration point. The overall DHPS1 concentration went from 0 µM to 500 µM (0, 50, 100, 200, 250 and 500 µM). Peak chemical shift changes were observed for a large portion of the protein, with most of the peaks undergoing a fast exchange regime. Further addition of DHPS1 caused rapid precipitation of the sample and the spectra deteriorated in quality as 500 µM DNA was reached. This is could be due to the formation of large bodies of protein-DNA aggregates with multiple proteins binding to one DHPS1 molecule. The precipitation of the protein prevented fitting of K_D values and also prevented confident transfer of the assignment from the apo-protein to a DNA complex spectrum.

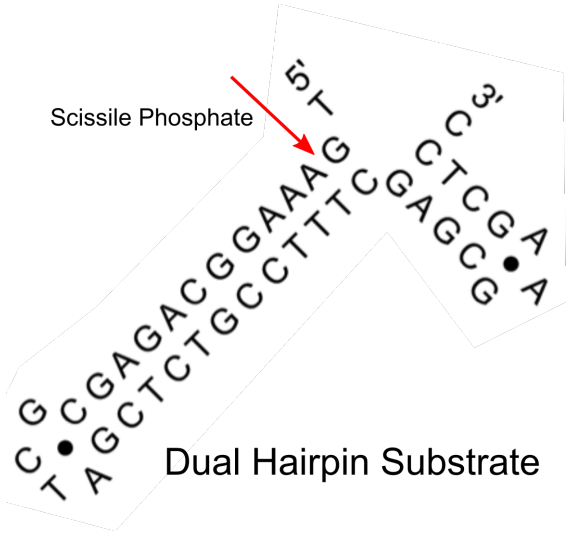

Figure 6.ii. Cartoon representation of the dual hairpin substrate (DHPS1). A unimolecular DNA construct with two hairpins present and mismatches on the 5' and 3' ends to produce flaps.

Peaks associated with the arch, α2- α3 loop and β-pin disappeared after the first or second addition of sub-stoichiometric DNA (50-100 µM) (i.e., A45, A196 **Fig. 6.iii.b**, G118 **Fig. 6.iii.c** and 53L **Fig. 6.iii.d**) indicating that these peaks are undergoing fast-intermediate exchange. The addition of DHPS1 therefore changes the dynamics of certain flexible DNA binding regions from a very fast ps-ns timescale to a much slower µs-ms timescale. The peaks from the α2- α3 loop and arch region did not re-appear after further addition of DHPS1, suggesting that these residues stay in an intermediate regime. The timescales of motion are likely to be 10 – 1000 s^{-1}, which is typical of intermediate exchange at 600 MHz. This change of exchange regime further

confirmed the conformational change of dynamics upon binding of substrate, but cannot place a definitive number on the timescale.

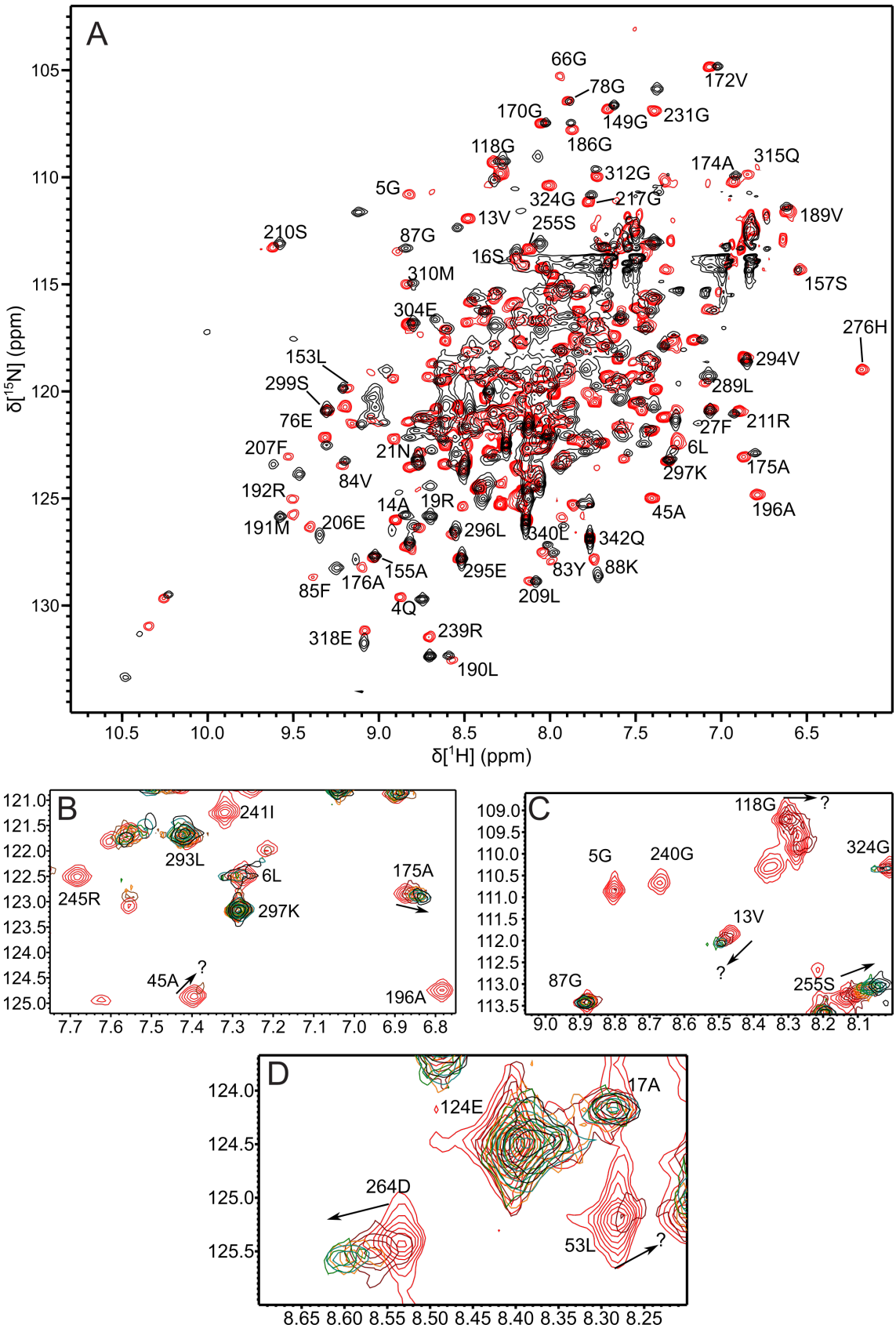

Figure 6.iii. Titration of substrate DNA into a hFEN1$_{K93A}$ protein sample in the absence of divalent cations. (A) ^{1}H-^{15}N TROSY spectra highlighting differences between hFEN1$_{K93A}$ bound and unbound to

substrate DNA. $hFEN1_{K93A}$ protein unbound is shown in red and the stable $hFEN1_{K93A}$-DHPS1 complex shown in black (see **section 6.2b**). Both spectra were recorded in the presence of 0.1 mM EDTA. A selection of more dispersed peaks are displayed on the spectrum of the unbound protein. (B-D) ^{1}H-^{15}N TROSY spectra taken during the titration of DNA into a $hFEN1_{K93A}$ sample with specific regions highlighted. DHPS1 DNA concentrations are 0 µM (red), 50 µM (maroon), 100 µM (orange), 200 µM (green), 250 µM (teal) and 500 µM (black). The quality of the spectra deteriorated rapidly during the later concentration points, so later chemical shifts are not accurate.

6.2b $hFEN1_{K93A}$-DNA stable complex formation and assignment

A method to produce a stable $hFEN1_{K93A}$-DHPS1 complex in order to produce an assignment and measure NSR on a complex was optimised. ^{13}C, ^{15}N, ^{2}H triply labeled $hFEN1_{K93A}$ protein was produced and grown using the method described before in **Section 2.1**. A growup in 500 mL of media yielded 0.6 µmol or 23 mg of protein. A 1:1.1, protein:DNA complex was pre-made under dilute conditions (1-10 µM) at lower salt (6 mM KCl) concentrations in the presence of EDTA (**Section 2.3h**). This sample was then concentrated to 500 µM protein. Once made, the sample was significantly more stable than the apo-protein and it would remain in solution for longer (> 1 month) (**Fig. 6.iii.a**). This could be due to the stability of the protein being enhanced by its substrate bound in a preferential location. The number of observable peaks was 230, lower than the 292 in the wild type hFEN1 spectrum. This suggests a loss of peaks to intermediate exchange or an increased R_2. Assignment of the protein complex was performed by the same methodology as apo-hFEN1 protein, using triply labeled ^{13}C, ^{15}N, ^{2}H protein in the presence of 8 mM $CaCl_2$. The assignment was hampered by poor correlation in the ^{13}C dimension, and only 132 peaks were assigned by this methodology, which is around 40% of theoretical peaks, and 57% of detectable peaks (**Fig. 6.iv.b**) (BMRB accession code: 27404, **Appendix A2**).

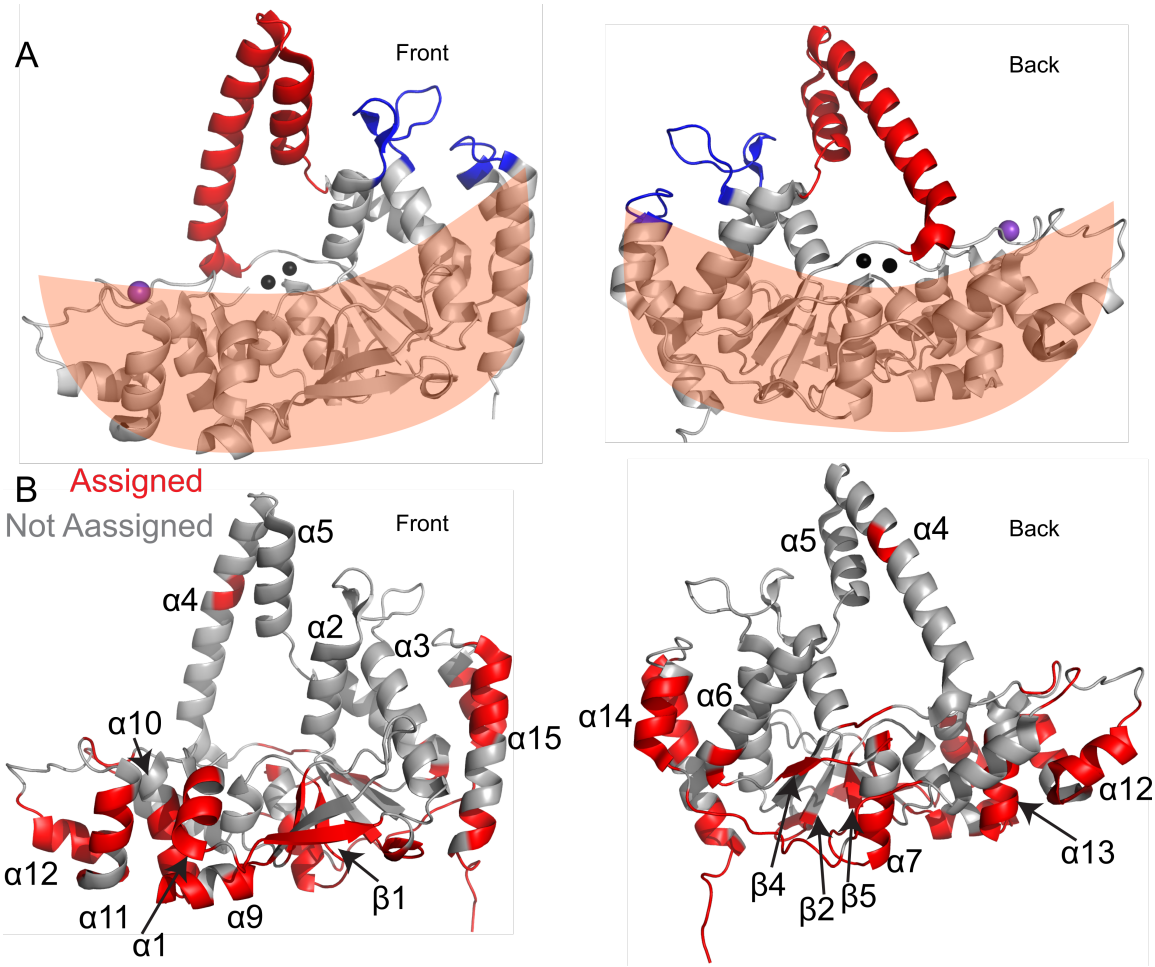

Figure 6.iv. Cartoon representation of the hFEN1$_{K93A}$ molecule to illustrate location of assigned residues. (A) A cartoon showing the 3Q8K structure with important regions highlighted, the arch (red), the α2-α3 loop (blue), the saddle region (orange saddle), the positions of the two catalytic metal ions (black spheres) and the potassium ion in the H2tH motif (purple sphere). (B) The assigned (red) and unassigned (grey) residues are coloured onto the product crystal structure 3Q8K. Secondary structure elements are labeled for clarity. Substrate DNA is present on the protein during assignment, but omitted on the figure for clarity.

Chemical shift differences between apo-hFEN1$_{K93A}$ protein and the hFEN1$_{K93A}$-DNA complex were plotted for known residues (**Fig. 6.v**). Regions that displayed high chemical shift differences included the H2tH motif (e.g., G231, I238, R245, and D248), the active site (e.g., G87, K88, A175 and R192), the 3'-flap binding pocket (e.g., S317, E318 and G324) and the hydrophobic wedge (e.g., G66). These chemical shift differences were mostly localised in regions known to interact with the DNA as seen from previous crystal structures bound to substrate and product (5UM9 and 3Q8K respectively).

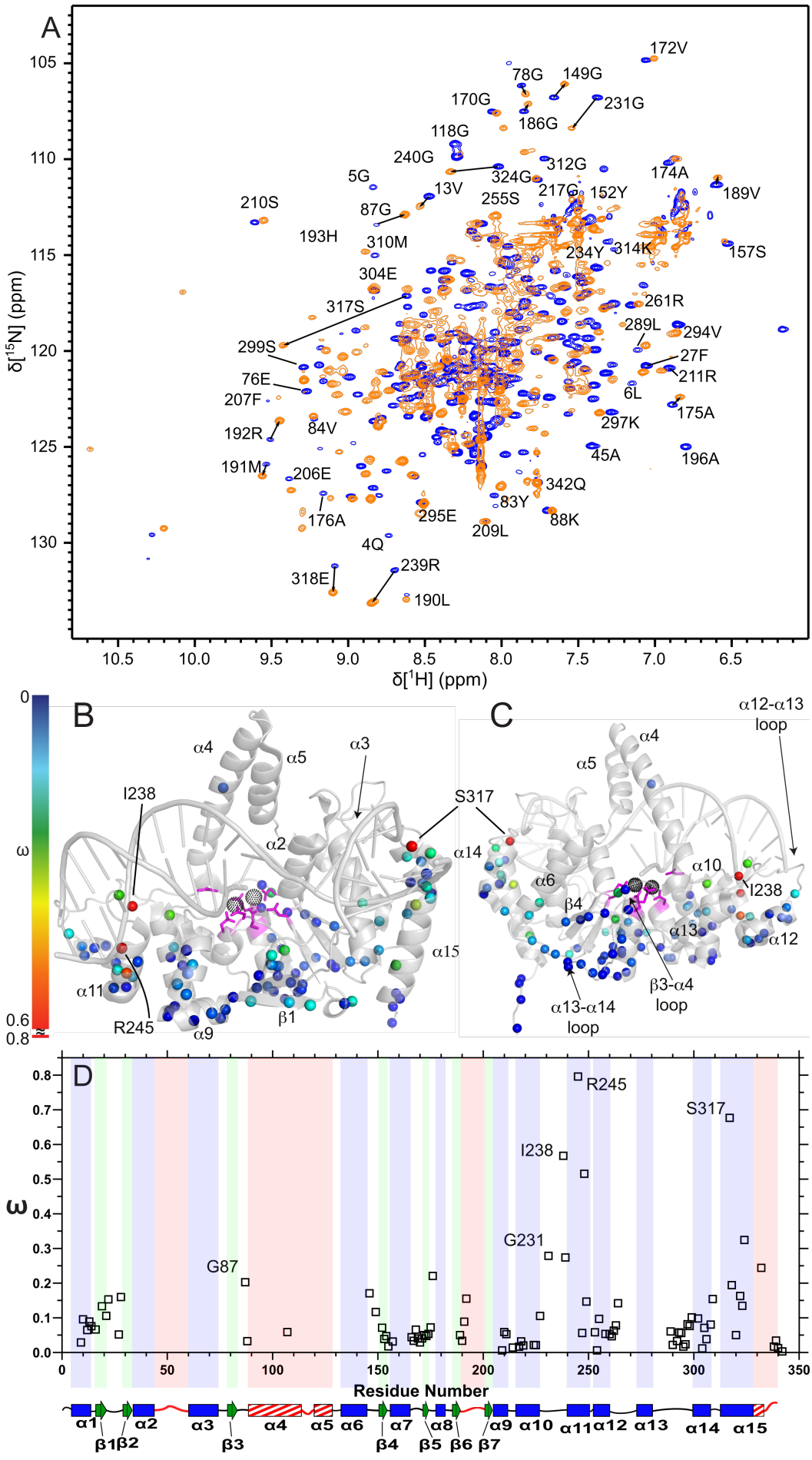
A
δ[15N] (ppm)
δ[1H] (ppm)
172V
78G
149G
170G
231G
186G
118G
240G
312G
5G
324G
174A
13V
255S
217G
152Y
189V
210S
87G
193H
310M
234Y
314K
157S
304E
317S
261R
289L
294V
299S
27F
76E
211R
207F
6L
175A
192R
84V
297K
45A
196A
191M
206E
342Q
83Y
176A
295E
88K
209L
4Q
239R
318E
190L
B
C
α4
α5
α3
I238
S317
α2
α14
α12-α13 loop
α10
α6
β4
α13
α12
α15
α11
R245
β1
α9
β3-α4 loop
α13-α14 loop
ω
D
R245
S317
I238
G231
G87
Residue Number
α1 α2 α3 α4 α5 α6 α7 α8 α9 α10 α11 α12 α13 α14 α15
β1 β2 β3 β4 β5 β6 β7

Figure 6.v. Differences in chemical shift between protein bound and unbound to DNA. (A) ^{1}H-^{15}N TROSY spectra highlighting differences between $hFEN1_{K93A}$ bound and unbound to substrate DNA. $hFEN1_{K93A}$ protein unbound is shown in blue and $hFEN1_{K93A}$-DHPS1 complex shown in orange. Both spectra were recorded in the presence of 8 mM $CaCl_2$. A selection of more dispersed peaks are displayed on the spectrum of the unbound protein. Small arrows highlight clear chemical shift differences. Front (B) and rear (C) views of chemical shift differences represented on the protein product structure 3Q8K. For every traceable peak pair there is a sphere represented on the amide residue. The magnitude of chemical shift difference is represented by color on each sphere as seen by the spectrum-bar. Secondary structure elements, and key residues are labeled. Active site carboxylates are represented in magenta and the positioning of the divalent cations in black mesh. (D) A graph showing chemical shift perturbation (ω) measured in ppm plotted versus residue number with secondary structure elements shown below for clarity.

Relaxation experiments were performed on this $hFEN1_{K93A}$-DHPS1 complex using the same methodology as before (**Section 5.2a, Section 2.3d**). Relaxation data had low signal-to-noise and was missing important residues in the arch and α2-α3 loop. A particular problem was the quality of the ^{1}H-^{15}N HSQC spectrum, hindering the ability to convert $R_{1\rho}$ values to R_2 reliably. Despite this, some preliminary R_1 data was obtained at 600 MHz (**Fig. 6.vi**). The R_1 values show a decrease from the apo-protein, consistent with the increase in molecular weight compared to apo-$hFEN1_{K93A}$ (52 kDa from 38 kDa). The majority of the residues were relaxing around 0.4 s^{-1}, with the C-terminus of the protein displaying much higher values, which is consistent with the apo-protein. Other than the C-terminus of the protein there are no anomalously high R_1 values. The arch, α2-α3 loop and β-pin remain unassigned due to being in an intermediate exchange regime and so there is no information on these key regions.

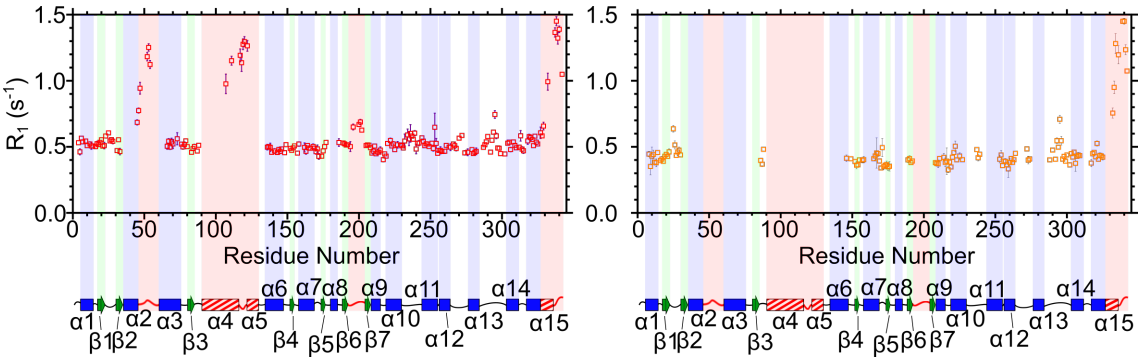

Figure 6.vi. Comparison of spin-lattice (R_1) values of apo-hFEN1 and $hFEN1_{K93A}$-DHPS1 complex. The apo-hFEN1 protein R1 values average around 0.5 s^{-1} (left in red), whereas the $hFEN1_{K93A}$-DNA complex R_1 values average around 0.4 s^{-1} (right in orange). Secondary structure mapping are displayed below the plots.

6.3 Influence of calcium ions on both apo-hFEN1$_{K93A}$ protein and hFEN1$_{K93A}$-DHPS1 complexes

6.3a Catalytically inert calcium ions are a good replacement for natural magnesium.

An essential co-factor for hFEN1 to perform hydrolysis and facilitate DNA conformational change is Mg^{2+} ions.[124,125] An analogous replacement for Mg^{2+} ions is Ca^{2+} ions, which facilitate the same DNA conformational change, but do not allow for hydrolysis.[28,126] In **section 1.5c**, the effect of divalent cations was noted on hFEN1 in terms of both quality of NMR spectra and stability of the sample. It concluded that Mg^{2+} and Ca^{2+} had no effect on stability and minor chemical shift perturbations of peaks next to the active site occurred. Differences between samples containing Mg^{2+} and Ca^{2+} showed localised changes near the active site, in particular the N-terminus of the protein. This is likely due to the differences in ionic radii and differences in coordination geometry. Therefore Ca^{2+} ions are suitable for studies with DNA, as they do not allow for hydrolysis, while mimicking the protein structural changes that Mg^{2+} ions produce *in vivo*. Therefore the role of Ca^{2+}, both in apo-hFEN1$_{K93A}$ and DNA bound hFEN1$_{K93A}$, can be studied. Ca^{2+} ions are similar in function but catalytically inert, hence can be used to reveal structural changes throughout the hFEN1$_{K93A}$ protein when substrate DNA is bound.

6.2b Calcium ions in Apo-hFEN1$_{K93A}$

Comparison of ^{1}H-^{15}N TROSY spectra with substrate-free hFEN1$_{K93A}$ in the absence and presence of 8 mM Ca^{2+} showed several chemical shift differences (**Fig. 6.vii.a**). The main differences of chemical shift were located in regions next to active site carboxylates (D34, D86, E158, E160, D179, E181, D233) and the N-terminus of the protein (**Fig. 6.vii.b,c**). Minor changes were present in the β5, α3, α7, β4-α7 loop, the β3-α4 loop and the βpin, which are all regions near the active site residues. There were also slight chemical shift perturbations present in the β strands, perhaps propagated structural changes from the active site. Furthermore, the addition of Ca^{2+} ions did not change peak heights associated with the arch and α2-α3 loop, suggesting that it does not change dynamics. Overall the changes in chemical shift are minor but located mainly in regions next to active site carboxylates.

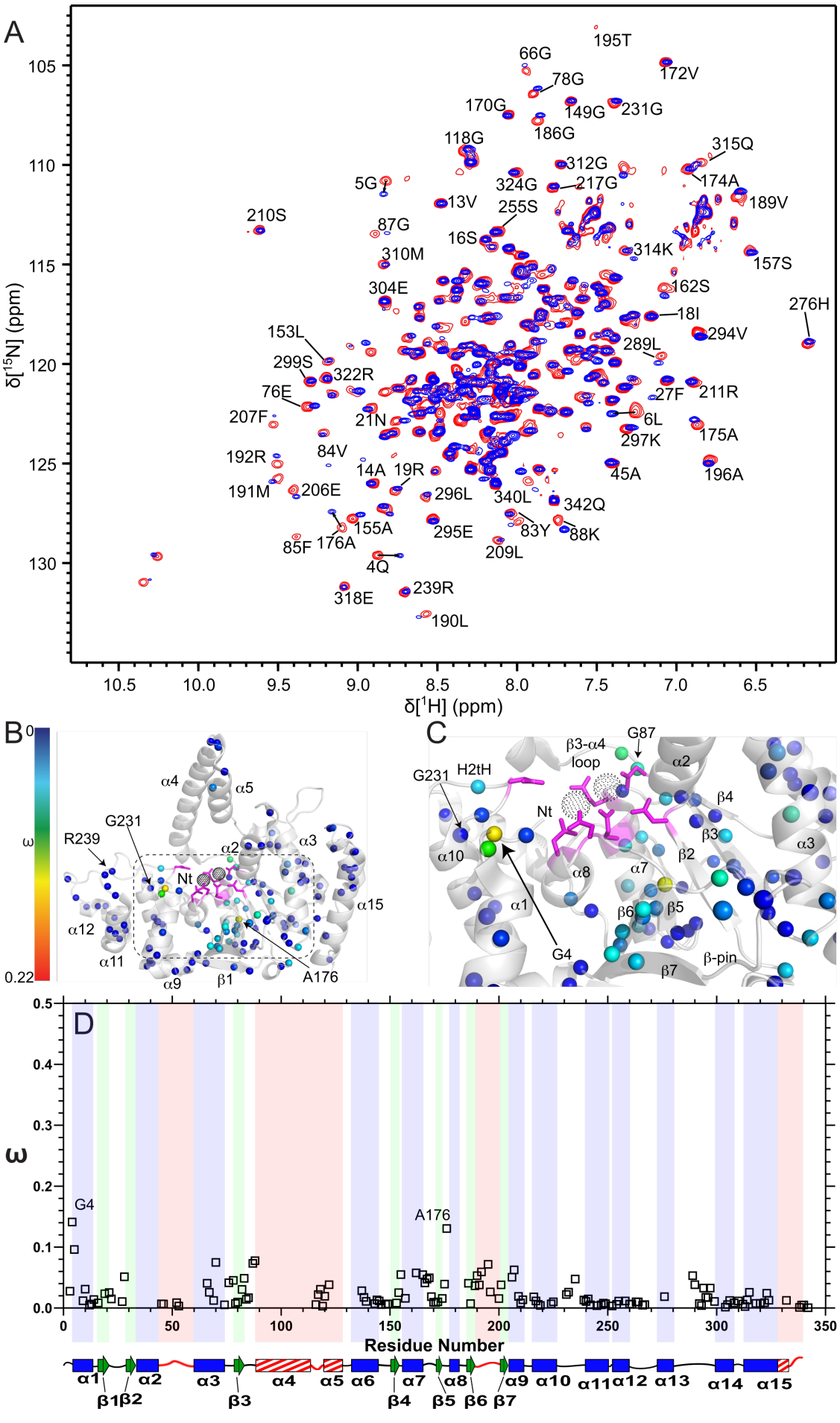
A
δ[15N] (ppm)
δ[1H] (ppm)
195T
66G
78G
172V
170G
149G
231G
186G
118G
315Q
312G
5G
324G
217G
174A
13V
255S
189V
210S
87G
16S
314K
157S
310M
304E
162S
276H
18I
153L
294V
289L
299S
322R
76E
27F
211R
207F
21N
6L
297K
175A
192R
84V
14A
19R
45A
196A
191M
206E
296L
340L
342Q
155A
295E
83Y
88K
85F
176A
209L
4Q
318E
239R
190L
B
ω
0
0.22
α4
α5
G231
R239
α2
α3
Nt
α15
α12
α11
α9
β1
A176
C
β3-α4
loop
G87
α2
G231
H2tH
β4
Nt
β3
α3
α10
β2
α8
α7
α1
β6
β5
G4
β-pin
β7
D
ω
G4
A176
Residue Number
α1
β1β2
α2
α3
β3
α4
α5
α6
α7
β4
α8
β5
β6
β7
α9
α10
α11
α12
α13
α14
α15

Figure 6.vii. Chemical shift differences between $hFEN1_{K93A}$ in the absence and presence of Ca^{2+} ions. (A) ^{1}H-^{15}N TROSY spectra of both $hFEN1_{K93A}$ in the presence of 0.1 mM EDTA (red) and 8 mM $CaCl_2$ (blue). Clearly dispersed peaks are labeled. (B) Chemical shift difference represented on the protein product structure 3Q8K. For every traceable peak pair there is a sphere represented on the amide residue. The magnitude of chemical shift difference is represented by color on each sphere as seen by the spectrum-bar. Secondary structure elements, and key residues are labeled. Active site carboxylates are represented in magenta and the positioning of the divalent cations in black mesh. (C) A close-up of the active site region within the dotted rectangle in (B). (D) A graph showing chemical shift perturbation (ω) measured in ppm plotted versus residue number with secondary structure elements shown below for clarity.

6.2c Calcium ions in the $hFEN1_{K93A}$-DHPS1 complex

A titration of 0-8 mM $CaCl_2$ was added to a 500 mM $hFEN1_{K93A}$-DHPS1 complex and known peaks were monitored. The peaks that changed chemical shift were similar to those that changed upon addition of Ca^{2+} ions to the $hFEN1_{K93A}$ protein sample, and included peaks around the active site. This suggests that the Ca^{2+} ions were bound in a similar location. However, the N-terminal residues Q4, G5 and L6 all disappeared upon addition of Ca^{2+} when the DNA was present (**Fig. 6.viii**, **Fig. 6.ix.c**). Q4 for example, broadened out greatly and shifted slightly upfield in ^{1}H after the first addition of Ca^{2+}, and subsequent titrations broadened it until it eventually disappeared. This shows that N-terminal peaks were changing their exchange regime to intermediate when Ca^{2+} was added to the $hFEN1_{K93A}$-DNA complex. The residue S317 was assigned in the $hFEN1_{K93A}$-DHPS1 complex in the presence of Ca^{2+} but at lower Ca^{2+} concentrations the peak disappears (**Fig. 6.ix.e**). Therefore suggesting that it changed to an intermediate exchange regime in the absence of Ca^{2+}. Comparing $hFEN1_{K93A}$ and $hFEN1_{K93A}$-DHPS1 complex, the weighted ^{1}H-^{15}N chemical shift perturbations (ω) value for S317 is 0.68 ppm (**Fig. 6.v.d**). This is consistent with the backbone amide of S317 making a hydrogen bond with a phosphate group on the upstream DNA, as observed in the 3Q8K crystal structure. Regardless of the presence or absence of Ca^{2+} ions, the arch residues and the α2-α3 loop do not reappear, suggesting that they are still in an intermediate exchange regime. Overall, dynamics of the protein changed when Ca^{2+} was added to a $hFEN1_{K93A}$-DNA complex, suggesting a change in conformation of N-terminal residues and S317, but not the arch or α2-α3 loop.

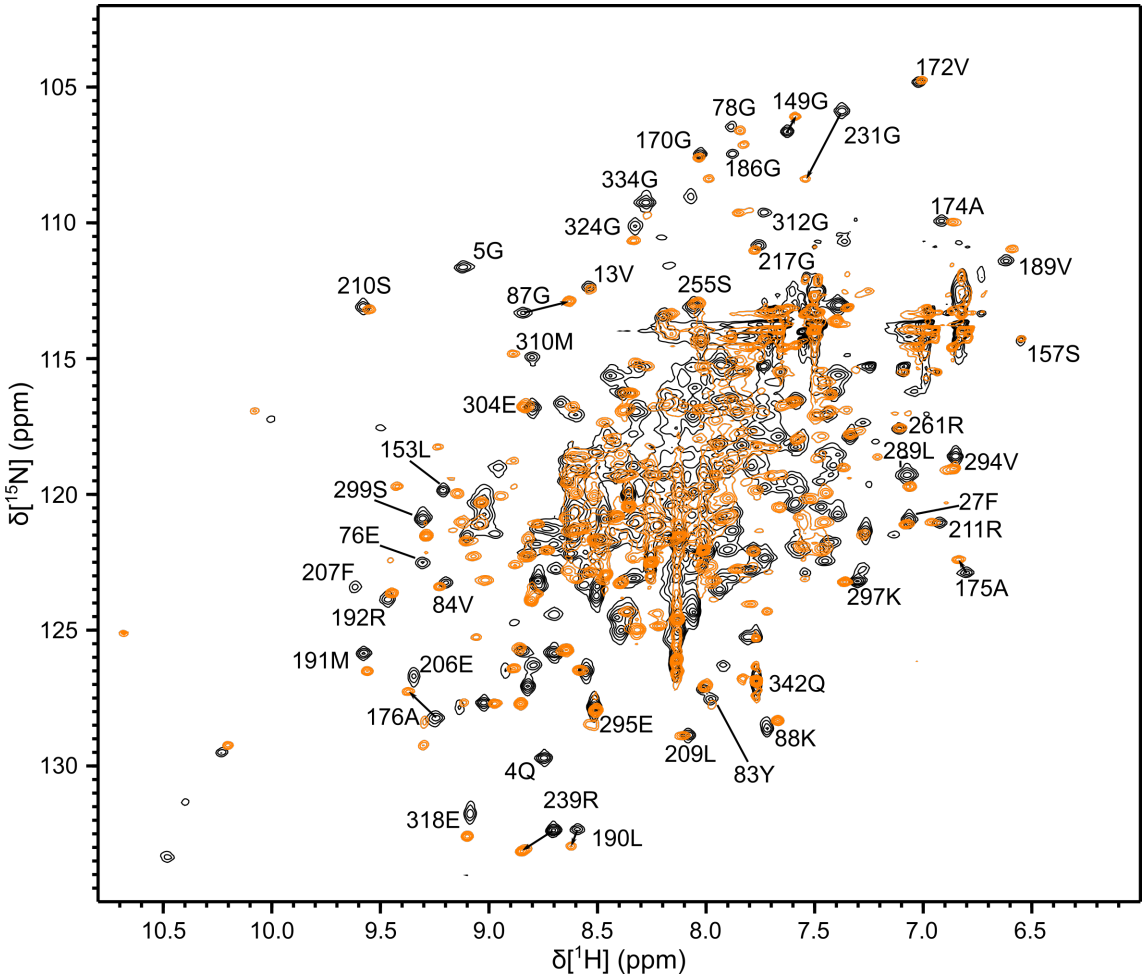

Figure 6.viii. ^{1}H-^{15}N TROSY spectra highlighting differences between hFEN1$_{K93A}$-DHPS1 complex in the presence and absence of Ca^{2+} ions. hFEN1$_{K93A}$-DHPS1 in the presence of 0.1 mM EDTA is shown in black. The hFEN1$_{K93A}$-DHPS1 complex in the presence of 8 mM $CaCl_2$ is shown in orange. A selection of more dispersed peaks are displayed on the spectrum of the unbound protein. Small arrows highlight clear chemical shift differences.

The chemical shift values of a large number of peaks removed from the active site changed when Ca^{2+} was titrated into the hFEN1$_{K93A}$-DNA complex. An example is G87 (**Fig. 6.viii**), which is next to the active site carboxylate D86, and has a large increase in chemical shift perturbation (ω) when Ca^{2+} ions are added in the hFEN1$_{K93A}$-DNA complex (0.22 ppm in the presence of DNA compared to 0.07 ppm without). In the 5UM9 hFEN1-substrate crystal structure the 5' flap passes through the arch near to G87, suggesting the Ca^{2+} ions mediate contact with it. Indeed, the presence of divalent metal ions has been shown to be required for efficient DNA threading.[126] Large chemical shift perturbations were also found at G231 and R239 (**Fig. 6.viii** and **Fig. 6.ix.d**,**c**) when Ca^{2+} ions were added to the hFEN1$_{K93A}$-DHPS1 complex. G231 is present at the N-terminal end of the H2tH motif and R239 is also in the H2tH motif contacting the template DNA. When Ca^{2+} ions are added in the absence of DNA, G231 has an ω value of 0.022 ppm, however in the presence of DNA, the chemical shift difference was greatly enhanced to

0.39 ppm. Likewise, addition of Ca^{2+} ions causes a shift for R239, which goes from an ω value of 0.013 ppm in the absence of DNA, to 0.17 ppm in the presence of DNA. Residues in the upstream binding pocket (α14-α15 loop and α6) also showed greater chemical shift differences than in apo-hFEN1$_{K93A}$. This shows that the H2tH motif, which is around 20 Å from the active site and distant residues in the upstream binding region responded to the Ca^{2+} metal occupancy of the active site.

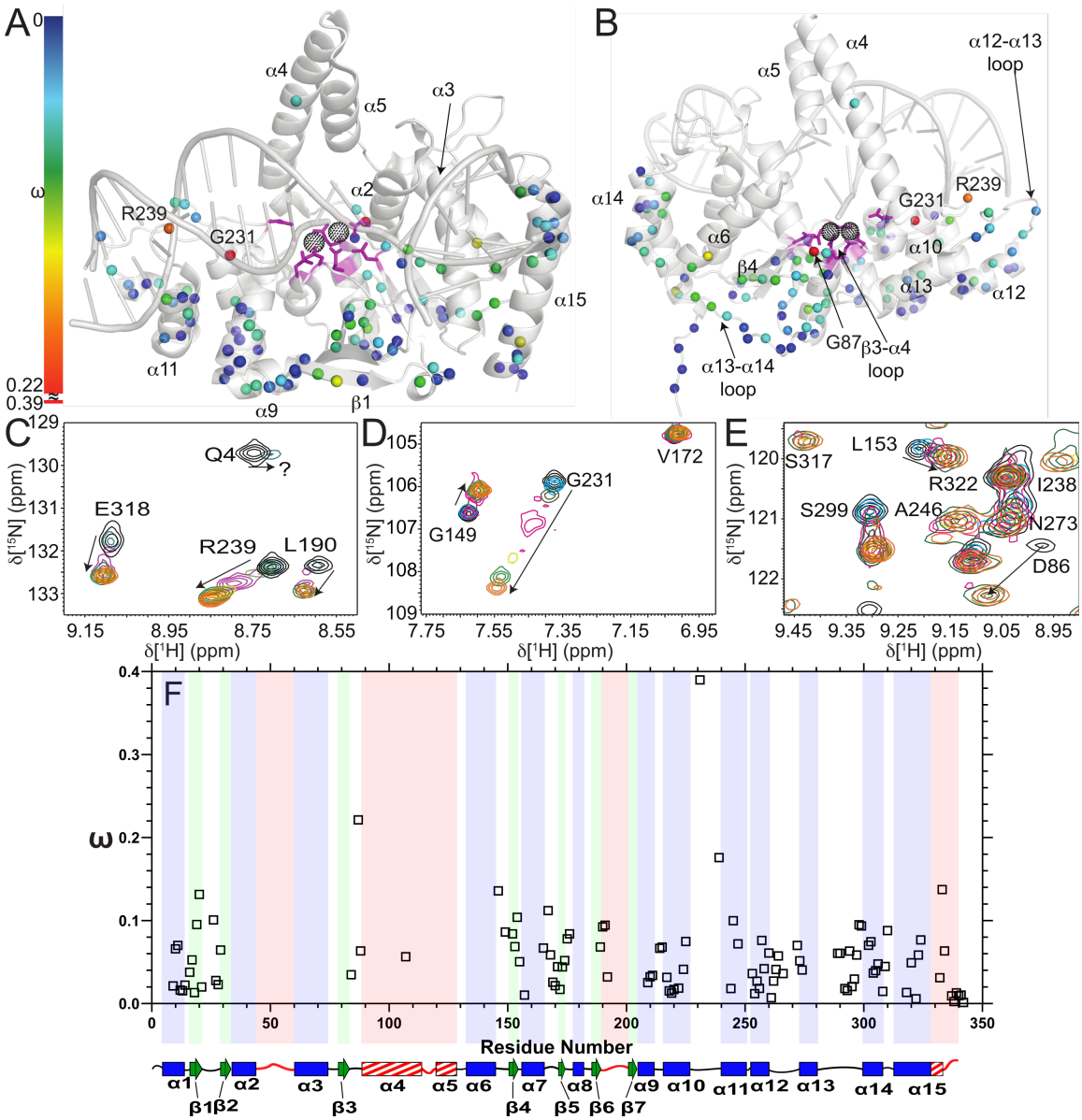

Figure 6.ix. Chemical shift differences between the hFEN1$_{K93A}$-DHPS1 complex in the absence and presence of Ca^{2+} ions. (A) Front and (B) rear views of chemical shift differences represented on the protein product structure 3Q8K. For every traceable peak pair there is a sphere represented on the amide residue. The magnitude of chemical shift difference is represented by color on each sphere as seen by the spectrum-bar. Secondary structure elements, and key residues are labeled. Active site carboxylates are represented in magenta and the positioning of the divalent cations in black mesh. (C-E) Show highlighted regions from **Fig.6.viii** with extra ^{1}H-^{15}N TROSY spectra of the Ca^{2+} titration of hFEN1$_{K93A}$-DHPS1. The Ca^{2+} concentrations are 0 mM (black), 0.5 mM (cyan), 1mM (olive), 2 mM (magenta), 4 mM (yellow), 6 mM (green) and 8 mM (orange). The panels represent disappearance through intermediate exchange

broadening of the N-terminal residue Q4 (C) and large chemical shift changes by fast exchange in R239 and R322 (E). L190 undergoes slow exchange (C), while G231 (D) and D86 (E) undergo intermediate to slow exchanges. Panel (E) shows the reappearance of S317 at higher Ca^{2+} concentrations. (F) A graph showing chemical shift perturbation (ω) measured in ppm plotted versus residue number with secondary structure elements shown below for clarity.

Overall the magnitude of chemical shift perturbation was greater in response to addition of Ca^{2+} in the presence of DNA than in the absence of DNA (**Fig 6.ix.a**,**b,f**). Chemical shift changes from Ca^{2+} ion addition in the β3-α4 loop where the 5' flap exits, the whole H2tH motif, β strands in the saddle region and α6, α14 and α15 are all larger in the $hFEN1_{K93A}$-DNA complex than $hFEN1_{K93A}$ alone. These observables point towards DNA binding regions of the protein being sensitive to local changes in the Ca^{2+} bound active site, even if they are removed in space to the active site.

6.4 Discussion

From **chapter 5**, it was shown that the arch region and α2-α3 loop were disordered in hFEN1 protein in the absence of ligand DNA. These regions are involved in crucial DNA selection steps in catalysis, the threading of the 5' flap (arch region) and 3' flap recognition (α2-α3 loop). The addition of substrate DNA to the protein changes the timescale of motion of the arch region and α2-α3 loop from around 10^{10} s^{-1} to 10 – 1000 s^{-1}. This reduction in rate implies that these regions are undergoing a conformational change at a different dynamic timescale. This might be the same conformational change that was responsible for the unexpectedly large influence of viscogen on rate constants. The single turnover rate of hFEN1 (15-30 s^{-1}) is within the estimate of the slower motion,[22] so it could be that the conformational change that occurs is linked to the rate-limiting step of reaction.

In the $hFEN1_{K93A}$-DNA complex, there are many chemical shift changes in residues distal from the active site upon addition of Ca^{2+}. The magnitude of the chemical shift changes was also greater than addition of Ca^{2+} to apo-$hFEN1_{K93A}$. The DNA on the protein is proposed to undergo conformational changes, mainly an untwisting of the 5' reacting strand into the active site.[29] Having metal ions present in the active site may distort the 5' flap and causes structural changes on the protein as far away as the H2tH motif or the upstream binding region (α14-α15). Furthermore residues in the N-terminus change dynamic regime when Ca^{2+} is titrated in the presence of DNA, but not in the absence. The G2 and the N-terminal amine have been proposed to aid in hydrolysis facilitating the formation of hydroxide. The change in dynamics of these residues strongly suggests that these residues are involved in the catalytic process, but only

when DNA is present. Thus preparation of the nucleophile may be controlled by access of DNA to the active site, linking reactivity to selectivity. Furthermore, the reappearance of S317 when Ca^{2+} is added indicates that the 3' flap duplex region is coordinating with the protein. Overall this shows that hFEN1 protein has a level of DNA-mediated allostery linking both 3' flap recognition and downstream duplex interactions to active site metals and catalysis.

Chapter 7: Conclusions

7.1 Summary

The work described in this thesis has used several approaches to study the reaction mechanism of human flap endonuclease. This structure-specific nuclease is an essential enzyme involved in lagging strand DNA replication and DNA repair mechanisms. Previous studies in the Grasby laboratory on hFEN1 highlighted the importance of a conformational change in the catalytic step. In chapter 3 the purpose was to further confirm the presence of a conformational change in the catalytic cycle of hFEN1. Dr. Blanka Séngerova in her PhD thesis postulated that for the T5 bacteriophage FEN some non-chemical step is likely to be involved in the conversion of the enzyme-substrate complex (ES) to product (EQ).[42] The work in this thesis attempts to extend these studies to the human endonuclease to determine if the same process is present. Indeed similar results for hFEN1 were found under k_{cat}/K_M conditions and single-turnover conditions. The rates were both insensitive to the pKa of the leaving group, consistent with a non-chemical step being involved with the conversion of enzyme-substrate complex to enzyme-product complex (i.e., $[ES] \rightarrow [EQ]$) for hFEN1.

However hFEN1 and T5 FEN differ in their responses to viscosity, with hFEN1 displaying an abnormally high viscosity dependence at relative viscosities higher than 2. This was likely attributable to the viscogen retarding two kinetic steps, firstly the bimolecular association rate and secondly, a conformational rearrangement. In order to separate the two processes, it would be necessary to perform these experiments under single-turnover conditions. This would be challenging under rapid-quench conditions due to the need of instantaneous mixing, which is slowed by viscous solutions. The only information at present is single-molecule FRET apparent single turnover 'rates' that show a dependence with the microviscogens glycerol and sucrose, but not with the macroviscogen polyethelyne glycol-8000.[40]

An attempt to quantify the rate of this conformational change was made by using NMR spectroscopy. In chapter 4 the challenges of studying a novel protein system are highlighted. Of particular concern was the stability of the sample, which directed much of the earlier work done by Dr. David Finger, and indeed was an ongoing problem throughout all experiments. The mutation of cysteine residues and the mutation of P188 failed to stop short term aggregation. Neither did the addition of various small hFEN1 ligands, such as mono- or di-nucleotides or known inhibitors aid in stability. While the addition of large excess of βME did aid stability in the short term, it should be noted that no one mutation or addition of small molecule stopped

aggregation of the protein over extended periods of time, a testament to instability of the protein. Nevertheless, samples were made that were amenable for assignment and relaxation experiments. Investigation of the peak heights showed the arch region and α2-α3 loop had much greater peak heights than would be expected for internal residues. These regions were key to the catalytic machinery, and displayed structural heterogeneity between DNA-bound[27,29] and non-DNA bound[24] hFEN1 X-ray crystal structures. Thus these two regions were likely candidates for sites of conformational change linked to catalysis.

In chapter 5 NMR relaxation rates were measured to quantify the motions the protein, and in particular the arch region and α2-α3 loop. Indeed these regions displayed much lower R_2/R_1 ratios and much more negative hNOE values than the other residues of the protein. Model free analysis further confirmed that the arch and α2-α3 loop were regions of intrinsic disorder within the protein. The fact that the arch is disordered in an extended conformation in solution in the absence of DNA means it can accommodate a wide variety of 5' flaps. It has been shown that 5' flaps with bulky streptavidin can pass through the arch, as well as even small regions of 5' DNA with secondary structure.[38] This lends weight to the arch undergoing a disorder-to-order transition to allow the reacting 5' flap strand to pass through or 'thread'.

Having disorder present in the protein could actually allow for a larger capture radius of the reacting flap and the 1nt 3' flap, aiding affinity. Furthermore, using chemical shift databases (ncIDP, TALOS-N) the arch displayed an interesting dual property, with the N-terminal portion sampling α-helical torsion angles while the C-terminal portion displaying fully extended conformations. Thus the proposed model is that the N-terminal portion of the arch is conformationally selected while the C-terminal portion and the α2-α3 loop are induced by presence of substrate. This makes sense from an energetic standpoint, as the so-called α4 is a large helix and for it to be induced would be a large energy barrier. Having the dual induced-fit and conformational selection of an ordered state allows for relatively low energy barriers of folding, while also allowing for a degree of specificity depending on what substrate is presented.

In chapter 6 the effect of adding DNA to hFEN1 was investigated. Adding sub-stoichiometric DNA was found to change the rate of motion of the arch and α2-α3 loop to an intermediate exchange regime. The intermediate exchange regime is usually found around 10-1000 s^{-1}, which places it close to the single turnover rate of hFEN1 catalysis (15-30 s^{-1}).[22] It is therefore a possibility that the movement of these regions when DNA is bound is the rate-limiting conformational change during catalysis. Furthermore, addition of Ca^{2+} ions to the hFEN1-DNA

complex causes widespread chemical shift perturbations in regions far away from the metal-binding active site. They also cause a change in dynamics of the N-terminus of the protein, which is facing into the active site. These widespread chemical shift changes and dynamic changes are not evident in hFEN1 in the absence of DNA. The regions of major perturbations, such as the H2tH motif and the 3' flap binding pocket are responsible for selectively binding a suitable substrate, hence aiding specificity. Thus the protein has a level of DNA-mediated allostery, which links DNA recognition regions to the active site.

7.2 Future Work

While the site of disorder-to-order in the protein has been located in a solution state system, the DNA dynamics are still relatively unknown. In chapter 1 it was proposed that two steps were required for DNA to position on hFEN1 in order for hydrolysis to occur. These were bending of the duplex, and twisting of the reacting strand into contact with the metals.[28] The ongoing question is whether the flapped substrate follows an induced fit,[40] or conformational selection model,[41] with respect to bending. In other words, whether the protein completely molds the DNA into a bent conformation, or whether the DNA exhibits the ability to bend in solution. The studies so far have focused on using single-molecule FRET techniques, which are limited in how fast detectors can measure photons, giving a millisecond limit on temporal resolution.

Using NMR spectroscopy it might be possible to study the ps-ns dynamics of sugar and base atoms by using model free analysis, similar to the technique applied for hFEN1. Special care is needed to decouple the tumbling of the molecule from the local librations of ^{15}N-H or ^{13}C-H bonds. A small construct can be made *in vitro* using ^{15}N, ^{13}C uniformly labeled nucleoside triphosphates.[127] Then an unlabeled RNA sequence is polymerised onto the end of the labeled domain of interest. This has shown to be a good way of analysing dynamics of small structural portions of RNA, such as regulatory RNAs in HIV-1.[128] Furthermore, atomic resolution relaxation dispersion can also be performed to see if any lowly populated excited states are present.[129] Measuring the dynamics in the presence and absence of hFEN1 should reveal any potential conformers in solution, highlighting any transiently sampled bent states. Thus information could be gained about a dynamic conformational selection by hFEN1 or whether the protein truly induces a bend in the DNA.

The lack of protein stability was a constant problem for hFEN1, however FENs are present in many other organisms. Of particular interest are the FENs from hyperthermophilic archaea, like

Pyrococcus furiosus[31] and *Archaeoglobus fulgidus*.[25] While the sequences of these enzymes are different, their activity and substrate specificity remains similar.[130] The interest of studying thermophilic species lies in the dynamics of the arch at temperatures that are lower than optimal for them. Firstly the stability of the protein is likely to be higher at lower temperatures. Secondly, by speeding up the molecular tumbling at higher temperatures, higher signal to noise ratios may be obtained. Furthermore, a link could be found between kinetic activities at different temperatures and the flexibility of arch and α2-α3 loop residues.

References

(1) Hershey, A. D. *J. Gen. Physiol.* **1952**, *36*, 39.

(2) McCarty, M. *J. Exp. Med.* **1946**, *83*, 89.

(3) Nakamura, J.; Swenberg, J. A. *Cancer Res.* **1999**, *59*, 2522.

(4) Fisher, L. M. *Nature* **1982**, *299*, 105.

(5) Gerlt, J. A.; Coderre, J. A.; Mehdi, S. *Adv. Enzymol. Relat. Areas Mol. Biol.* **1983**, *55*, 291.

(6) Williams, N. H.; Wyman, P. *Chem. Commun.* **2001**, 1268.

(7) Yang, W. *Q. Rev. Biophys.* **2011**, *44*, 1.

(8) Finger, L. D.; Atack, J. M.; Tsutakawa, S.; Classen, S.; Tainer, J.; Grasby, J.; Shen, B. *Eukaryot. Replisome A Guid. to Protein Struct. Funct.* **2012**, *62*, 301.

(9) Szankasi, P.; Smith, G. R. *Science* **1995**, *267*, 1166.

(10) Genschel, J.; Bazemore, L. R.; Modrich, P. *J. Biol. Chem.* **2002**, *277*, 13302.

(11) Sun, M.; Schwalb, B.; Pirkl, N.; Maier, K. C.; Schenk, A.; Failmezger, H.; Tresch, A.; Cramer, P. *Mol. Cell* **2013**, *52*, 52.

(12) Balakrishnan, L.; Bambara, R. A. *Annu. Rev. Biochem.* **2013**, *82*, 119.

(13) MacNeill, S. *Subcell. Biochem.* **2012**, *62*, 1.

(14) Blow, J. J.; Dutta, A. *Nat. Rev. Mol. Cell Biol.* **2005**, *6*, 476.

(15) Burgers, P. M. J. *Chromosoma* **1998**, *107*, 218.

(16) Pellegrini, L. *Eukaryot. Replisome a Guid. to Protein Struct. Funct.* **2012**, *62*, 135.

(17) Burgers, P. M. J. *J. Biol. Chem.* **2009**, *284*, 4041.

(18) Ceska, T. A.; Sayers, J. R.; Stier, G.; Suck, D. *Nature* **1996**, *382*, 90.

(19) Larsen, E.; Gran, C.; Saether, B. E.; Seeberg, E.; Klungland, A. *Mol Cell Biol* **2003**, *23*, 5346.

(20) Tishkoff, D. X.; Filosi, N.; Gaida, G. M.; Kolodner, R. D. *Cell* **1997**, *88*, 253.

(21) Zheng, L.; Jia, J.; Finger, L. D.; Guo, Z.; Zer, C.; Shen, B. *Nucleic Acids Res.* **2011**, *39*, 781.

(22) Finger, L. D.; Blanchard, M. S.; Theimer, C. A.; Sengerová, B.; Singh, P.; Chavez, V.; Liu, F.; Grasby, J. A.; Shen, B. *J. Biol. Chem.* **2009**, *284*, 22184.

(23) Yang, W.; Lee, J. Y.; Nowotny, M. *Mol. Cell* **2006**, *22*, 5.

(24) Sakurai, S.; Kitano, K.; Yamaguchi, H.; Hamada, K.; Okada, K.; Fukuda, K.; Uchida, M.; Ohtsuka, E.; Morioka, H.; Hakoshima, T. *EMBO J.* **2005**, *24*, 683.

(25) Chapados, B. R.; Hosfield, D. J.; Han, S.; Qiu, J.; Yelent, B.; Shen, B.; Tainer, J. A. *Cell* **2004**, *116*, 39.

(26) Exell, J. C.; Thompson, M. J.; Finger, L. D.; Shaw, S. J.; Debreczeni, J.; Ward, T. A.; McWhirter, C.; Siöberg, C. L. B.; Molina, D. M.; Abbott, W. M.; Jones, C. D.; Nissink, J. W. M.; Durant, S. T.; Grasby, J. A. *Nat. Chem. Biol.* **2016**, *12*, 815.

(27) Tsutakawa, S. E.; Classen, S.; Chapados, B. R.; Arvai, A. S.; Finger, L. D.; Guenther, G.; Tomlinson, C. G.; Thompson, P.; Sarker, A. H.; Shen, B.; Cooper, P. K.; Grasby, J. A.; Tainer, J. A. *Cell* **2011**, *145*, 198.

(28) Algasaier, S. I.; Exell, J. C.; Bennet, I. A.; Thompson, M. J.; Gotham, V. J. B.; Shaw, S. J.; Craggs, T. D.; Finger, L. D.; Grasby, J. A. *J. Biol. Chem.* **2016**, *291*, 8258.

(29) Tsutakawa, S. E.; Thompson, M. J.; Arvai, A. S.; Neil, A. J.; Shaw, S. J.; Algasaier, S. I.; Kim, J. C.; Finger, L. D.; Jardine, E.; Gotham, V. J. B.; Sarker, A. H.; Her, M. Z.; Rashid, F.; Hamdan, S. M.; Mirkin, S. M.; Grasby, J. A.; Tainer, J. A. *Nat. Commun.* **2017**, *8*, 15855.

(30) Mase, T.; Kubota, K.; Miyazono, K. I.; Kawarabayasi, Y.; Tanokura, M. *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.* **2011**, *67*, 209.

(31) Hosfield, D. J.; Mol, C. D.; Shen, B.; Tainer, J. A. *Cell* **1998**, *95*, 135.

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

(33) Matsui, E.; Musti, K. V.; Abe, J.; Yamasaki, K.; Matsui, I.; Harata, K. *J. Biol. Chem.* **2002**, *277*, 37840.

(34) Hwang, K. Y.; Baek, K.; Kim, H.-Y. Y.; Cho, Y. *Nat. Struct. Biol.* **1998**, *5*, 707.

(35) Shah, S.; Dunten, P.; Stiteler, A.; Park, C. K.; Horton, N. C. *Proteins Struct. Funct. Bioinforma.* **2015**, *83*, 188.

(36) Garforth, S. J.; Ceska, T. A.; Suck, D.; Sayers, J. R. *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *96*, 38.

(37) Patel, N.; Exell, J. C.; Jardine, E.; Ombler, B.; Finger, L. D.; Ciani, B.; Grasby, J. A. *J. Biol. Chem.* **2013**, *288*, 34239.

(38) Patel, N.; Atack, J. M.; Finger, L. D.; Exell, J. C.; Thompson, P.; Tsutakawa, S.; Tainer, J. A.; Williams, D. M.; Grasby, J. A. *Nucleic Acids Res.* **2012**, *40*, 4507.

(39) Fetrow, J. S. *FASEB J.* **1995**, *9*, 708.

(40) Rashid, F.; Harris, P. D.; Zaher, M. S.; Sobhy, M. A.; Joudeh, L. I.; Yan, C.; Piwonski, H.; Tsutakawa, S. E.; Ivanov, I.; Tainer, J. A.; Habuchi, S.; Hamdan, S. M. *Elife* **2017**, *6*.

(41) Craggs, T. D.; Hutton, R. D.; Brenlla, A.; White, M. F.; Penedo, J. C. *Nucleic Acids Res.* **2014**, *42*, 1857.

(42) Sengerová, B.; Tomlinson, C.; Atack, J. M.; Williams, R.; Sayers, J. R.; Williams, N. H.; Grasby, J. A. *Biochemistry* **2010**, *49*, 8085.

(43) Tom, S.; Henricksen, L. A.; Bambara, R. A. *J. Biol. Chem.* **2000**, *275*, 10498.

(44) Beattie, T. R.; Bell, S. D. *Biochem. Soc. Trans.* **2011**, *39*, 70.

(45) Loria, J. P.; Berlow, R. B.; Watt, E. D. *Acc. Chem. Res.* **2008**, *41*, 214.

(46) Schwartz, S. D.; Schramm, V. L. *Nat. Chem. Biol.* **2009**, *5*, 551.

(47) Smock, R. G.; Gierasch, L. M. *Science (80-.).* **2009**, *324*, 198.

(48) Kamerzell, T. J.; Middaugh, C. R. *J. Pharm. Sci.* **2008**, *97*, 3494.

(49) Neudecker, P.; Lundström, P.; Kay, L. E. *Biophys. J.* **2009**, *96*, 2045.

(50) Dyson, H. J.; Wright, P. E. *Methods Enzymol.* **2005**, *394*, 299.

(51) Kleckner, I. R.; Foster, M. P. *Biochim. Biophys. Acta - Proteins Proteomics* **2011**, *1814*, 942.

(52) Kempf, J. G.; Loria, J. P. *Cell Biochem. Biophys.* **2003**, *37*, 187.

(53) Palmer, A. G.; Grey, M. J.; Wang, C. *Methods Enzymol.* **2005**, *394*, 430.

(54) Massi, F.; Johnson, E.; Wang, C.; Rance, M.; Palmer, A. G. *J. Am. Chem. Soc.* **2004**, *126*, 2247.

(55) Pervushin, K.; Riek, R.; Wider, G.; Wuthrich, K. *Proc. Natl. Acad. Sci.* **1997**, *94*, 12366.

(56) Pervushin, K. D.; Riek, R.; Wider, G. D.; Wüthrich, K. P. D. Transverse relaxation-optimized NMR spectroscopy (TROSY), May 19, 1999.

(57) Reed, M. A. C.; Hounslow, A. M.; Sze, K. H.; Barsukov, I. G.; Hosszu, L. L. P.; Clarke, A. R.; Craven, C. J.; Waltho, J. P. *J. Mol. Biol.* **2003**, *330*, 1189.

(58) Raines, R. T. *Chem. Rev.* **1998**, *98*, 1045.

(59) Wlodawer, A.; Borkakoti, N.; Moss, D. S.; Howlin, B. *Acta Crystallogr. Sect. B* **1986**, *42*, 379.

(60) Hammes, G. G. *Biochemistry* **2002**, *41*, 8221.

(61) Rico, M.; Santoro, J.; González, C.; Bruix, M.; Neira, J. L.; Nieto, J. L.; Herranz, J. *J. Biomol. NMR* **1991**, *1*, 283.

(62) Toiron, C.; González, C.; Bruix, M.; Rico, M. *Protein Sci.* **1996**, *5*, 1633.

(63) Cole, R.; Loria, J. P. *Biochemistry* **2002**, *41*, 6072.

(64) Beach, H.; Cole, R.; Gill, M. L.; Loria, J. P. *J. Am. Chem. Soc.* **2005**, *127*, 9167.

(65) Zegers, I.; Maes, D.; Dao-Thi, M. -H; Wyns, L.; Poortmans, F.; Palmer, R. *Protein Sci.* **1994**, *3*, 2322.

(66) Kovrigin, E. L.; Loria, J. P. *J. Am. Chem. Soc.* **2006**, *128*, 7724.

(67) Watt, E. D.; Shimada, H.; Kovrigin, E. L.; Loria, J. P. *Proc. Natl. Acad. Sci.* **2007**, *104*, 11981.

(68) Kempf, J. G.; Jung, J. Y.; Sampson, N. S.; Loria, J. P. *J. Am. Chem. Soc.* **2003**, *125*, 12064.

(69) Igumenova, T. I.; Palmer, A. G. *J. Am. Chem. Soc.* **2006**, *128*, 8110.

(70) Lolis, E.; Alber, T.; Davenport, R. C.; Rose, D.; Hartman, F. C.; Petsko, G. A. *Biochemistry* **1990**, *29*, 6609.

(71) Lolis, E.; Petsko, G. A. *Biochemistry* **1990**, *29*, 6619.

(72) Guallar, V.; Jacobson, M.; McDermott, A.; Friesner, R. A. *J. Mol. Biol.* **2004**, *337*, 227.

(73) Jogl, G.; Rozovsky, S.; McDermott, A. E.; Tong, L. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100*,

50.

(74) Pompliano, D. L.; Peyman, A.; Knowles, J. R. *Biochemistry* **1990**, *29*, 3186.

(75) Kempf, J. G.; Jung, J. yeon; Ragain, C.; Sampson, N. S.; Loria, J. P. *J. Mol. Biol.* **2007**, *368*, 131.

(76) Sampson, N. S.; Knowles, J. R. *Biochemistry* **1992**, *31*, 8482.

(77) Berlow, R. B.; Igumenova, T. I.; Loria, J. P. *Biochemistry* **2007**, *46*, 6001.

(78) Studier, F. W. *Protein Expr. Purif.* **2005**, *41*, 207.

(79) Sivashanmugam, A.; Murray, V.; Cui, C.; Zhang, Y.; Wang, J.; Li, Q. *Protein Sci.* **2009**, *18*, 936.

(80) Gasteiger, E.; Gattiker, A.; Hoogland, C.; Ivanyi, I.; Appel, R. D.; Bairoch, A. *Nucleic Acids Res.* **2003**, *31*, 3784.

(81) Cavaluzzi, M. J. *Nucleic Acids Res.* **2004**, *32*, 13e.

(82) Åström, H.; Limén, E.; Strömberg, R. *J. Am. Chem. Soc.* **2004**, *126*, 14710.

(83) Williams, R.; Sengerová, B.; Osborne, S.; Syson, K.; Ault, S.; Kilgour, A.; Chapados, B. R.; Tainer, J. A.; Sayers, J. R.; Grasby, J. A. *J. Mol. Biol.* **2007**, *371*, 34.

(84) Exell, J. C. Spectroscopic evidence for catalytically-required FEN-1 mediated DNA conformational change; a novel strategy for FEN1 inhibition, University of Sheffield, 2015.

(85) Cheng, N. S. *Ind. Eng. Chem. Res.* **2008**, *47*, 3285.

(86) Brouwer, A. C.; Kirsch, J. F. *Biochemistry* **1982**, *21*, 1302.

(87) Cavanagh, J.; Fairbrother, W.; Palmer III, A.; Rance, M.; Skelton, N. *Protein NMR spectroscopy. Principles and Practice.*; 2007; Vol. 2nd.

(88) Vranken, W. F.; Boucher, W.; Stevens, T. J.; Fogh, R. H.; Pajon, A.; Llinas, M.; Ulrich, E. L.; Markley, J. L.; Ionides, J.; Laue, E. D. *Proteins Struct. Funct. Genet.* **2005**, *59*, 687.

(89) Shen, Y.; Bax, A. *J. Biomol. NMR* **2013**, *56*, 227.

(90) Lakomek, N. A.; Ying, J.; Bax, A. *J. Biomol. NMR* **2012**, *53*, 209.

(91) d'Auvergne, E. J.; Gooley, P. R. *J. Biomol. NMR* **2003**, *25*, 25.

(92) d'Auvergne, E. J.; Gooley, P. R. *J. Biomol. NMR* **2006**, *35*, 117.

(93) d'Auvergne, E. J.; Gooley, P. R. *Mol. Biosyst.* **2007**, *3*, 483.

(94) d'Auvergne, E. J.; Gooley, P. R. *J. Biomol. NMR* **2008**, *40*, 107.

(95) d'Auvergne, E. J.; Gooley, P. R. *J. Biomol. NMR* **2008**, *40*, 121.

(96) Williamson, M. P. *Prog. Nucl. Magn. Reson. Spectrosc.* **2013**, *73*, 1.

(97) Patel, D.; Tock, M. R.; Frary, E.; Feng, M.; Pickering, T. J.; Grasby, J. A.; Sayers, J. R. *J. Mol. Biol.* **2002**, *320*, 1025.

(98) Tock, M. R.; Frary, E.; Sayers, J. R.; Grasby, J. A. *EMBO J.* **2003**, *22*, 995.

(99) Briggs, G. E.; Haldane, J. B. S. *Biochem. J.* **1925**, *19*, 1037.

(100) Van Slyke, D. D.; Cullen, G. E. *J. Biol. Chem.* **1914**, *19*, 141.

(101) Berg, O. G.; von Hippel, P. H. *Annu. Rev. Biophys. Biophys. Chem.* **1985**, *14*, 131.

(102) Zalatan, J. G.; Herschlag, D. *J. Am. Chem. Soc.* **2006**, *128*, 1293.

(103) Mihai, C.; Kravchuk, A. V.; Tsai, M. D.; Bruzik, K. S. *J. Am. Chem. Soc.* **2003**, *125*, 3236.

(104) O'Brien, P. J.; Herschlag, D. *Biochemistry* **2002**, *41*, 3207.

(105) Zheng, L.; Dai, H.; Zhou, M.; Li, M.; Singh, P.; Qiu, J.; Tsark, W.; Huang, Q.; Kernstine, K.; Zhang, X.; Lin, D.; Shen, B. *Nat. Med.* **2007**, *13*, 812.

(106) Zheng, L.; Shen, B. *J. Mol. Cell Biol.* **2011**, *3*, 23.

(107) Rogers, J. M.; Steward, A.; Clarke, J. *J. Am. Chem. Soc.* **2013**, *135*, 1415.

(108) Sekhar, A.; Latham, M. P.; Vallurupalli, P.; Kay, L. E. *J. Phys. Chem. B* **2014**, *118*, 4546.

(109) Rauscher, A.; Derényi, I.; Gráf, L.; Málnási-Csizmadia, A. *IUBMB Life* **2013**, *65*, 35.

(110) Sakurai, S.; Kitano, K.; Yamaguchi, H.; Hamada, K.; Okada, K.; Fukuda, K.; Uchida, M.; Ohtsuka, E.; Morioka, H.; Hakoshima, T. *EMBO J.* **2005**, *24*, 683.

(111) Guo, Z.; Chavez, V.; Singh, P.; Finger, L. D.; Hang, H.; Hegde, M. L.; Shen, B. *J. Mol. Biol.* **2008**, *377*, 679.

(112) Conte, M. Lo; Carroll, K. S. In *Oxidative Stress and Redox Regulation*; Jakob, U.; Reichmann, D., Eds.; Springer Netherlands, 2013; Vol. 9789400757, pp. 1–42.

(113) Golovanov, A. P.; Hautbergue, G. M.; Wilson, S. A.; Lian, L. Y. *J. Am. Chem. Soc.* **2004**, *126*, 8933.

(114) Tumey, L. N.; Bom, D.; Huck, B.; Gleason, E.; Wang, J.; Silver, D.; Brunden, K.; Boozer, S.; Rundlett, S.; Sherf, B.; Murphy, S.; Dent, T.; Leventhal, C.; Bailey, A.; Harrington, J.; Bennani, Y. L. *Bioorganic Med. Chem. Lett.* **2005**, *15*, 277.

(115) Hunter, T. *Cell* **1998**, *92*, 141.

(116) Henneke, G.; Koundrioukoff, S.; Hübscher, U. *Oncogene* **2003**, *22*, 4301.

(117) Kay, L. E.; Torchia, D. A.; Bax, A. *Biochemistry* **1989**, *28*, 8972.

(118) Rossi, P.; Swapna, G. V. T.; Huang, Y. J.; Aramini, J. M.; Anklin, C.; Conover, K.; Hamilton, K.; Xiao, R.; Acton, T. B.; Ertekin, A.; Everett, J. K.; Montelione, G. T. *J. Biomol. NMR* **2010**, *46*, 11.

(119) Bieri, M.; d'Auvergne, E. J.; Gooley, P. R. *J. Biomol. NMR* **2011**, *50*, 147.

(120) Tamiola, K.; Acar, B.; Mulder, F. A. A. *J. Am. Chem. Soc.* **2010**, *132*, 18000.

(121) Lacroix, E.; Viguera, A. R.; Serrano, L. *J. Mol. Biol.* **1998**, *284*, 173.

(122) Das, R. K.; Ruff, K. M.; Pappu, R. V. *Curr. Opin. Struct. Biol.* **2015**, *32*, 102.

(123) Skrynnikov, N. R.; Dahlquist, F. W.; Kay, L. E. *J. Am. Chem. Soc.* **2002**, *124*, 12352.

(124) Lyamichev, V.; Brow, M.; Dahlberg, J. *Science.* **1993**, *260*, 778.

(125) Harrington, J. J.; Lieber, M. R. *EMBO J.* **1994**, *13*, 1235.

(126) Finger, L. D.; Patel, N.; Beddows, A.; Ma, L.; Exell, J. C.; Jardine, E.; Jones, A. C.; Grasby, J. A. *Nucleic Acids Res.* **2013**, *41*, 9839.

(127) Alvarado, L. J.; Longhini, A. P.; LeBlanc, R. M.; Chen, B.; Kreutz, C.; Dayie, T. K. *Methods Enzymol.* **2014**, *549*, 133.

(128) Zhang, Q.; Xiaoyan, S.; Watt, E. D.; Al-Hashimi, H. M. *Science.* **2006**, *311*, 653.

(129) Xue, Y.; Kellogg, D.; Kimsey, I. J.; Sathyamoorthy, B.; Stein, Z. W.; McBrairty, M.; Al-Hashimi, H. M. In *Methods in Enzymology*; 2015; Vol. 558, pp. 39–73.

(130) Kaiser, M. W.; Lyamicheva, N.; Ma, W.; Miller, C.; Neri, B.; Fors, L.; Lyamichev, V. I. *J. Biol. Chem.* **1999**, *274*, 21387.

Appendices

Residue Number	Residue Name	H	N	C	CA	CB
1	Met	-	-	-	-	-
2	Gly	-	-	170.41	43.74	-
3	Ile	7.99	122.22	176.64	57.16	35.46
4	Gln	8.75	129.11	176.58	57.09	27.2
5	Gly	8.73	110.27	176.12	45.73	-
6	Leu	7.15	121.99	176.82	57.54	39.74
7	Ala	8.5	121.5	180.25	56.41	16.48
8	Lys	7.95	119.33	177.66	58.18	31.24
9	Leu	7.76	119.97	179.18	57.59	41.19
10	Ile	8.37	117.53	177.11	64.07	35.92
11	Ala	7.4	120.09	178.29	54.87	17.02
12	Asp	7.74	115.76	178.47	56.29	41.18
13	Val	8.37	111.51	175.34	61.37	32.22
14	Ala	8.79	125.51	-	49.82	18.24
15	Pro	-	-	179.27	64.99	29.87
16	Ser	8.12	113.28	175.3	59.83	61.81
17	Ala	8.19	123.64	174.64	52.21	20.16
18	Ile	7.22	117.47	175.86	59.3	38.51
19	Arg	8.68	125.7	174.64	53.57	32.22
20	Glu	8.38	122.62	175.8	54.39	30.73
21	Asn	8.85	121.56	173.81	51.46	42.83
22	Asp	8.58	120.4	177.88	53.69	42.47
23	Ile	8.65	127.01	175.05	64.72	37.27
24	Lys	8.22	120.25	178.97	57.86	31.1
25	Ser	7.9	115	174.48	59.83	62.42
26	Tyr	7.46	119.23	172.94	58.31	36.99
27	Phe	6.96	120.21	177.18	60.26	38.32
28	Gly	8.17	117.07	174.3	44.61	-
29	Arg	7.88	119.04	175	54.03	29.57
30	Lys	8.35	121.64	177.38	-	34.08
31	Val	-	-	173.22	59.34	35.31
32	Ala	8.2	125.44	174.77	49.45	19.83
33	Ile	8.84	124.27	-	59.18	39.35
34	Asp	-	-	-	-	-
35	Ala	-	-	-	-	-
36	Ser	-	-	-	-	-
37	Met	-	-	-	-	-
38	Ser	-	-	-	-	-
39	Ile	-	-	-	-	-

40	Tyr	-	-	-	-	-
41	Gln	-	-	-	-	-
42	Phe	-	-	178.5	-	-
43	Leu	8.04	119.04	179.8	56.99	41.39
44	Ile	7.66	114.75	177.03	61.91	37
45	Ala	7.28	124.51	178.41	53.02	18.23
46	Val	7.5	117.1	176.56	62.41	31.41
47	Arg	8.09	123.73	-	55.79	29.55
48	Gln	-	-	-	-	-
49	Gly	-	-	-	-	-
50	Gly	-	-	173.95	44.92	-
51	Asp	8.19	120.86	176.3	53.96	40.67
52	Val	7.88	119.77	176.29	62.2	31.6
53	Leu	8.2	124.82	177.45	54.77	40.93
54	Gln	8.18	120.44	175.84	55.65	28.45
55	Asn	-	-	-	-	-
56	Glu	-	-	-	-	-
57	Glu	-	-	-	-	-
58	Gly	-	-	-	-	-
59	Glu	-	-	-	-	-
60	Thr	-	-	-	-	-
61	Thr	-	-	-	-	-
62	Ser	-	-	-	-	-
63	His	-	-	-	-	-
64	Leu	-	-	-	-	-
65	Met	-	-	178.1	-	-
66	Gly	7.83	104.6	174.9	46.08	-
67	Met	7.14	120.33	178.71	58.28	-
68	Phe	8.22	120.94	175.84	60.78	39.27
69	Tyr	8.12	115.29	180.24	62.9	36.76
70	Arg	8.87	118.41	-	58.32	27.79
71	Thr	-	-	176.26	-	-
72	Ile	7.81	122.85	176.52	65.85	36.99
73	Arg	7.48	118.65	179.05	58.68	28.26
74	Met	7.36	116.76	178.03	60.37	31.94
75	Met	-	-	181.27	57.88	31.43
76	Glu	9.14	121.46	177.02	58.17	28.47
77	Asn	7.17	117.08	173.44	53.84	40.16
78	Gly	7.75	105.75	174.34	46.04	-
79	Ile	7.76	124.84	174.23	60.51	37.77
80	Lys	7.92	126.92	-	51.92	31.38
81	Pro	-	-	176.84	61.37	32.69
82	Val	8.59	118.13	174.15	58.76	33.59
83	Tyr	7.94	127.43	174.64	58.36	40
84	Val	9.11	123.1	174.91	60.68	32.4

85	Phe	9.29	128.24	175.8	57.36	40.37
86	Asp	8.84	121.1	177.41	55.64	41.87
87	Gly	8.76	113.19	173.95	44.32	-
88	Lys	7.62	127.98	-	53.94	31.38
89	Pro	-	-	-	-	-
90	Pro	-	-	-	-	-
91	Gln	-	-	-	-	-
92	Leu	-	-	-	-	-
93	Lys	-	-	-	-	-
94	Ser	-	-	-	-	-
95	Gly	-	-	-	-	-
96	Glu	-	-	-	-	-
97	Leu	-	-	-	-	-
98	Ala	-	-	-	-	-
99	Lys	-	-	-	-	-
100	Arg	-	-	-	-	-
101	Ser	-	-	-	-	-
102	Glu	-	-	-	-	-
103	Arg	-	-	-	-	-
104	Arg	-	-	178.8	56.56	-
105	Ala	7.89	123.14	179.39	53.42	17.72
106	Glu	8.13	119.61	177.74	57.52	28.75
107	Ala	7.97	123.32	179.4	53.42	17.78
108	Glu	8.11	119	177.86	57.51	28.8
109	Lys	7.86	120.96	177.77	57.4	31.49
110	Gln	7.99	119.51	177.18	56.47	27.83
111	Leu	7.93	122.14	178.07	55.81	41.02
112	Gln	8.07	120	176.98	56.39	28.04
113	Gln	8.12	120.48	176.42	56.15	28.33
114	Ala	8.09	124.37	178.21	52.67	18.09
115	Gln	8.13	119.04	176.32	55.66	28.3
116	Ala	8.09	124.92	177.7	52.25	18.21
117	Ala	8.1	123.04	178.41	52.23	18.28
118	Gly	8.22	108.81	174.25	45.05	-
119	Ala	8	123.97	178.05	52.17	18.35
120	Glu	8.42	120.23	176.91	56.79	28.99
121	Gln	8.23	120.63	176.37	55.76	28.49
122	Glu	8.25	122.3	176.91	56.49	29.05
123	Val	8	120.9	176.84	62.75	31.5
124	Glu	8.28	123.74	177.32	56.82	28.96
125	Lys	-	-	-	-	-
126	Phe	-	-	-	-	-
127	Thr	-	-	-	-	-
128	Lys	-	-	-	-	-
129	Arg	-	-	-	-	-

130	Leu	-	-	-	-	-
131	Val	-	-	-	-	-
132	Lys	-	-	-	-	-
133	Val	-	-	-	-	-
134	Thr	-	-	-	-	-
135	Lys	-	-	177.59	59.37	31.54
136	Gln	7.8	117.79	178.35	59.01	26.7
137	His	7.4	118.06	178.99	60.41	31.24
138	Asn	7.7	117.37	177.09	56.57	38.31
139	Asp	9.04	120.94	180.26	57.21	39.09
140	Glu	8.63	120.63	178.76	59.25	29.18
141	Cys	8.1	117.2	176.35	65.08	26.95
142	Lys	8.31	120.48	177.62	60.26	31.44
143	His	7.63	119.39	177.41	58.52	29.45
144	Leu	7.94	118.48	178.22	58.08	41.07
145	Leu	8.46	118.51	179.6	58.01	39.47
146	Ser	8.33	115.16	178.5	61.31	62.27
147	Leu	7.79	122.75	178.09	57.07	41.3
148	Met	8.15	116.38	173.98	57.42	34.99
149	Gly	7.54	106.17	173.98	45.36	-
150	Ile	7.98	122.17	-	54.16	37.4
151	Pro	-	-	175.38	62.18	31.49
152	Tyr	7.34	112.84	173.91	54.56	41.62
153	Leu	9.09	119.36	175.19	52.46	46.02
154	Asp	8.74	123	174.94	53.62	41.07
155	Ala	8.88	127.12	175.7	49.55	16.98
156	Pro	-	-	178.24	63.31	29.98
157	Ser	6.39	113.52	173.76	55.81	61.6
158	Glu	8.61	124.06	179.96	56.77	28.15
159	Ala	10.52	134.14	181.02	55.85	17.75
160	Glu	-	-	-	-	-
161	Ala	-	-	180.16	54.52	17.22
162	Ser	7.01	115.74	175.8	64.47	63.35
163	Cys	8.76	120.92	175.93	64.54	25.65
164	Ala	7.4	119.3	178.65	54.86	17.86
165	Ala	7.71	120.25	180.26	54.83	16.5
166	Leu	7.54	117.73	178.33	57.9	41.65
167	Val	7.34	121.77	180.99	64.95	30.75
168	Lys	8.63	123.49	178.65	59.42	32.2
169	Ala	7.41	116.56	177.5	51.42	19.31
170	Gly	8.06	107.61	175.13	45.48	-
171	Lys	8.04	116.38	176.57	55.47	31.73
172	Val	6.94	104.55	173.8	57.79	33.34
173	Tyr	8.62	120.63	173.53	58.72	39.36
174	Ala	6.74	110.3	174.6	50.18	24.04

175	Ala	6.95	121.68	175.34	49.21	21.14
176	Ala	8.99	126.37	174.91	50.38	20.7
177	Thr	7.55	116.5	-	58.81	67.7
178	Glu	-	-	-	-	-
179	Asp	-	-	-	-	-
180	Met	-	-	-	-	-
181	Asp	-	-	-	-	-
182	Cys	-	-	-	-	-
183	Leu	-	-	-	-	-
184	Thr	-	-	176	-	-
185	Phe	7.76	114.9	175.04	59.72	39.76
186	Gly	7.76	106.6	174.84	44.87	-
187	Ser	8.26	116.16	174.57	58.57	64.09
188	Ala	7.74	127.81	178.43,174.57	54.61	18.77
189	Val	6.59	111.86	175.25	60.06	34.44
190	Leu	8.44	131.63	175.02	52.75	46.02
191	Met	9.34	125.23	174.36	54.01	36.25
192	Arg	9.38	123.98	175.51	55.16	31.59
193	His	9.2	112.73	177.51	55.96	26.71
194	Leu	9.11	124.58	-	57.02	41.6
195	Thr	-	-	173.86	59.93	67.67
196	Ala	6.69	124.42	-	51.56	18.87
197	Ser	-	-	-	-	-
198	Glu	-	-	-	-	-
199	Ala	-	-	178.77	53.47	17.7
200	Lys	7.27	115.12	175.87	56.21	31.45
201	Lys	7.54	115.32	175.87	56.23	28.88
202	Leu	8.44	121.81	-	52.03	40.73
203	Pro	-	-	175.42	62	30.86
204	Ile	8.4	120.95	175.99	59.43	34.7
205	Gln	7.79	126.12	174.17	53.82	31.53
206	Glu	9.3	126.41	174.85	53.75	31.97
207	Phe	9.37	122.12	175.64	56.14	42.05
208	His	8.55	121.63	176.04	54.82	31.64
209	Leu	8.05	128.35	177.87	57.9	41.93
210	Ser	9.53	112.69	177.28	61.14	-
211	Arg	6.8	120.47	177.8	56.64	28.6
212	Ile	7.49	121.2	177	65.94	36.98
213	Leu	7.88	116.03	179.93	58.02	37.98
214	Gln	7.77	117.23	179.43	58.44	28.1
215	Glu	8.25	118.65	179.07	58.84	28.77
216	Leu	7.79	114.51	176.72	54.67	41.73
217	Gly	7.65	110.75	174.71	46.47	-
218	Leu	7.85	117.53	176.15	52.52	44.88

219	Asn	7.49	116.1	176.24	50.13	38.59
220	Gln	8.54	119.1	177.94	59.03	26.71
221	Glu	8.52	118.59	178.53	60.97	28.25
222	Gln	8.2	118.68	178.41	57.77	28.54
223	Phe	8.45	120.91	177.47	60.38	38.67
224	Val	8.62	122.92	177.22	67.06	29.85
225	Asp	7.33	120.15	178.75	57.14	40.3
226	Leu	8.41	120.93	177.58	57.78	39.71
227	Cys	7.62	116.15	176.26	63.68	25.63
228	Ile	7.84	120.74	178.73	64.61	36.65
229	Leu	7.85	118.99	179.75	56.23	40.9
230	Leu	8.12	121.55	177.49	57	39.19
231	Gly	7.24	106.33	171.5	46.05	-
232	Ser	9.93	128.82	173.9	57.25	67.79
233	Asp	9.16	117.06	177.98	54.8	40.16
234	Tyr	7.31	113.1	174.77	60.28	39.67
235	Cys	7.17	108.57	-	55.74	29
236	Glu	-	-	175.36	57.16	30.06
237	Ser	8.03	115.23	172.91	56.21	63.68
238	Ile	8.3	119.05	176.83	60.71	37.14
239	Arg	8.61	131.02	177.1	57.48	28.82
240	Gly	8.59	110.33	174.18	44.87	-
241	Ile	7.24	120.87	175.3	59.47	35.63
242	Gly	8.29	116.1	171.66	44.27	-
243	Pro	-	-	177.84	65.75	32
244	Lys	7.95	114.53	179.56	59.11	30.43
245	Arg	7.59	121.82	177.56	57.22	28.12
246	Ala	8.51	121.92	178.66	55.38	16.99
247	Val	7.39	115.69	176.75	66.45	30.56
248	Asp	7.27	119.31	178.64	57.29	41.14
249	Leu	8.5	117.1	179.4	57.28	-
250	Ile	8.19	121.11	178.82	60.7	34.74
251	Gln	8.52	120.89	178.67	58.97	27.44
252	Lys	7.69	116.23	178.39	58.44	32.23
253	His	8.22	114.84	177.64	56.05	31.06
254	Lys	8.05	113.91	175.69	59.79	29.3
255	Ser	8.01	112.82	174.59	55.56	66.12
256	Ile	8.37	122.26	176.82	66.04	36.45
257	Glu	8.92	119.57	178.46	59.69	28.91
258	Glu	7.33	117.86	178.48	57.68	28.46
259	Ile	7.79	120.96	178.38	65.8	37.07
260	Val	8.51	116.62	178.06	66.19	30.59
261	Arg	7.05	117.18	177.63	58.27	29.77
262	Arg	7.33	116.39	176.29	54.82	29.91
263	Leu	7.31	121.44	176.99	53.99	42.12

264	Asp	8.41	124.84	175.69	51.37	40.97
265	Pro	-	-	177.52	63.76	31.07
266	Asn	8.26	115.75	175.86	54.31	38.2
267	Lys	7.34	118.86	175.54	56.47	33.91
268	Tyr	7.75	116.94	-	54.11	38.53
269	Pro	-	-	-	-	-
270	Val	-	-	-	-	-
271	Pro	-	-	175.91	62.58	30.82
272	Glu	8.38	122.67	177.14	56.63	29.12
273	Asn	8.95	121.09	-	53.24	-
274	Trp	7.81	118.55	-	-	28.99
275	Leu	-	-	-	-	-
276	His	-	-	-	-	-
277	Lys	-	-	179.43	58.32	-
278	Glu	8.81	123.03	179.48	59.82	26.91
279	Ala	7.81	123.19	177.15	54.43	18.39
280	His	8.31	117.95	176.47	59.04	28.15
281	Gln	7.6	116.57	177.53	58.25	27.64
282	Leu	7.48	121.51	-	57.76	-
283	Phe	-	-	-	-	-
284	Leu	-	-	-	-	-
285	Glu	-	-	-	-	-
286	Pro	-	-	176.66	62.44	31.05
287	Glu	8.51	120.65	176.94	56.24	28.6
288	Val	8.59	120.98	175.02	58.95	36.89
289	Leu	6.94	118.99	174.31	53.6	40.8
290	Asp	8.28	119.85	176.75	50.35	40.13
291	Pro	-	-	176.95	63.89	32.31
292	Glu	7.94	113.93	177.73	56.45	27.96
293	Ser	7.55	112.81	174.08	58.25	63.91
294	Val	6.77	118.36	174.47	60.79	32.17
295	Glu	8.42	127.39	175.12	54.53	29.68
296	Leu	8.46	126.22	174.97	53.76	41.22
297	Lys	7.17	122.73	173.36	54.84	34.46
298	Trp	8.73	126.63	176.25	58.89	30.04
299	Ser	9.2	120.58	172.81	56.62	64.57
300	Glu	8.41	121.25	174.76	54.48	28.44
301	Pro	-	-	175.64	62.08	31.56
302	Asn	8.52	121.23	174.74	50.82	36.56
303	Glu	8.3	124.1	177.38	61.01	29.2
304	Glu	8.74	116.27	180.42	59.28	28
305	Glu	7.52	117.92	179.11	57.79	28.66
306	Leu	8.5	122.89	178.78	58.73	40.95
307	Ile	8.06	120.36	178.64	63.62	35.78
308	Lys	7.75	121.8	178.53	60.04	31.35

309	Phe	8.04	117.07	176.36	59.12	39.59
310	Met	8.73	114.47	176.34	56.59	32.07
311	Cys	8.36	116.42	177.13	58.81	27.5
312	Gly	7.63	109.5	175.16	47.35	-
313	Glu	7.56	118.89	173.76	57.97	28.69
314	Lys	7.2	113.81	175.62	53.66	29.9
315	Gln	6.77	109.62	175.87	57.21	24.73
316	Phe	8.58	118.92	176.8	58.47	38.5
317	Ser	8.72	116.68	176.09	57.47	62.94
318	Glu	8.98	130.68	178.37	59.96	28.51
319	Glu	8.51	118.92	179.81	59.31	28.25
320	Arg	7.29	118.14	179.87	58.2	29.37
321	Ile	7.93	120.42	178.73	61.08	33.88
322	Arg	9.07	120.18	179.54	60.68	28.51
323	Ser	7.86	113.93	177.52	61.43	-
324	Gly	7.91	109.88	174.66	47.36	-
325	Val	8.59	121.41	178.19	66.44	30.25
326	Lys	7.98	120.86	179.51	58.81	30.78
327	Arg	7.41	118.72	178.99	58.81	29.98
328	Leu	7.82	119.29	178.8	56.51	40.85
329	Ser	8.31	115.43	176.5	60.32	62.53
330	Lys	8.05	122.05	178.1	57.6	31.3
331	Ser	7.85	115.04	174.97	59.66	62.77
332	Arg	7.67	121.94	176.79	56.26	29.39
333	Gln	-	-	-	-	-
334	Gly	-	-	-	-	-
335	Ser	-	-	-	-	-
336	Thr	-	-	174.45	61.66	69.2
337	Leu	8.05	124.21	177.2	54.96	41.21
338	Glu	8.17	122.05	176.27	56.22	29.33
339	Val	7.92	121.54	175.75	61.87	31.63
340	Leu	8.04	125.56	176.8	54.55	41.44
341	Phe	8.02	121.01	174.69	57.14	38.59
342	Gln	7.68	126.4	180.28	56.88	29.66

A1. ^{1}H, ^{15}N and ^{13}C chemical shifts from the assignment of hFEN1$_{P188A}$. Also found in the BMRB under accession code 27403.

Residue Number	Residue Name	H	N	C	CA	CB
8	Lys	8.07	120.5	-	58.35	31.07
9	Leu	7.87	120.64	-	57.87	41.36
10	Ile	8.54	118.52	-	64.62	36.77
11	Ala	7.64	120.37	178.34	54.85	17.17
12	Asp	7.86	116.42	-	56.34	41.55
13	Val	8.5	112.34	-	61.54	-
14	Ala	8.83	125.49	-	49.76	18.75
16	Ser	8.13	113.24	-	59.8	61.98
17	Ala	8.34	124.24	-	52.31	20.44
18	Ile	7.31	117.73	-	59.39	38.64
19	Arg	8.85	126.23	174.64	53.51	32.16
20	Glu	8.36	123.14	175.83	54.37	30.26
21	Asn	8.8	122.13	173.9	51.37	43.14
22	Asp	8.49	120	-	53.07	42.89
25	Ser	7.81	115.27	-	60.34	62.65
26	Tyr	7.67	119.16	-	58.28	37.32
27	Phe	7.04	120.97	177.18	60.54	38.33
28	Gly	8.41	117.88	174.31	44.61	-
29	Arg	7.93	119.23	175.02	54.06	29.7
30	Lys	8.47	121.6	-	54.19	34.19
86	Asp	9.04	122.23	-	55.25	42.26
87	Gly	8.6	112.72	173.73	44.18	-
88	Lys	7.64	128.13	174.29	53.94	31.57
107	Ala	7.95	123.07	-	54.66	17.18
108	Glu	8.31	119.11	-	-	-
146	Ser	8.3	114.89	-	61.14	62.42
149	Gly	7.56	105.98	173.88	45.17	-
152	Tyr	7.37	113.47	173.61	54.62	41.78
153	Leu	9.12	119.85	-	52.35	-
154	Asp	8.77	123.75	-	53.5	41.07
155	Ala	8.95	127.58	-	49.6	17.42
157	Ser	6.51	114.06	-	55.68	61.7
158	Glu	8.19	124.36	-	56.02	-
164	Ala	7.42	119.89	178.78	54.85	17.69
165	Ala	7.49	120.02	180.19	54.79	16.95
166	Leu	7.74	118.1	-	57.9	-
167	Val	7.54	120.74	-	64.35	-
168	Lys	8.61	122.69	-	59.93	32.27
169	Ala	7.4	116.19	-	51.47	19.48
170	Gly	8	107.54	175.08	45.35	-
171	Lys	8.08	116.64	-	55.41	31.92
172	Val	6.97	104.68	-	57.84	33.38
173	Tyr	8.76	120.94	-	59.54	39.35

174	Ala	6.84	109.9	-	50.82	-
175	Ala	6.8	122.32	-	49.01	21.87
176	Ala	9.35	127.1	-	50.21	19.87
189	Val	6.56	110.92	-	60.62	34.73
190	Leu	8.59	132.82	-	52.68	46.16
191	Met	9.54	126.38	-	54.13	35.89
192	Arg	9.42	123.42	-	54.93	31.32
208	His	8.71	121.9	-	54.66	31.4
209	Leu	8.07	128.77	-	57.83	42.06
210	Ser	9.52	113.04	-	61.22	-
211	Arg	6.92	120.9	-	56.69	28.96
214	Gln	7.81	117.56	-	58.4	28.31
215	Glu	8.24	118.77	-	58.67	28.61
216	Leu	7.85	115.48	-	54.88	-
217	Gly	7.75	110.83	-	46.37	-
218	Leu	7.92	118	176.17	52.55	45.16
219	Asn	7.57	116.43	-	50.19	38.57
220	Gln	8.6	119.45	-	58.98	26.61
221	Glu	8.56	119.08	-	60.81	-
222	Gln	8.2	119.09	-	57.73	28.37
223	Phe	8.59	121.22	-	60.43	38.89
224	Val	8.74	123.47	-	67.32	29.92
225	Asp	7.41	120.87	-	57.35	40.52
226	Leu	8.44	121.6	-	57.94	39.84
227	Cys	7.82	116.57	-	63.56	-
237	Ser	8.35	116.78	-	56.94	63.65
238	Ile	8.91	119.94	-	60.7	37.39
239	Arg	8.81	132.98	-	57.82	28.55
244	Lys	7.9	120.85	174.38	59.07	30.7
245	Arg	7.63	116.59	-	58.69	32.26
246	Ala	9.11	120.91	-	55.39	18.59
247	Val	7.43	115.73	-	66.3	31.19
248	Asp	6.86	119.03	-	57.32	41.49
249	Leu	8.45	117.14	-	57.37	41.13
253	His	8.24	115.11	-	56.28	31.75
254	Lys	8.13	113.99	-	59.75	29.46
255	Ser	8.02	112.89	174.57	55.73	66.21
256	Ile	8.5	122.72	-	65.85	36.61
257	Glu	8.85	118.63	-	59.99	28.59
258	Glu	7.44	118.38	-	57.6	28.58
259	Ile	7.99	121.39	-	65.95	37.41
260	Val	8.59	116.64	-	66.11	30.51
261	Arg	7.08	117.39	-	58.17	29.8
262	Arg	7.4	116.85	-	54.68	30.07
263	Leu	7.43	121.84	-	54.24	42.29

264	Asp	8.62	125.61	175.65	51.31	41.09
272	Glu	8.44	122.89	177.19	56.6	29.31
273	Asn	9	120.89	-	53.38	36.99
274	Trp	7.84	118.69	-	55.53	-
287	Glu	8.38	120.65	-	-	28.48
288	Val	8.57	121.2	174.95	58.76	36.94
289	Leu	7.03	119.59	174.48	53.57	40.96
290	Asp	8.33	120.31	176.77	50.42	40.67
292	Glu	8	114.21	-	56.58	28.09
293	Ser	7.6	113.05	174.07	58.2	64.03
294	Val	6.82	118.87	174.53	60.97	32.33
295	Glu	8.48	127.82	175.1	54.5	29.93
296	Leu	8.55	126.32	-	53.75	41.35
297	Lys	7.33	123.11	173.3	54.89	34.53
298	Trp	8.82	127.51	-	58.91	29.63
299	Ser	9.26	121.42	172.62	56.81	64.72
300	Glu	8.49	121.52	174.63	54.54	28.58
302	Asn	8.53	121.14	-	50.77	36.61
303	Glu	8.29	124.86	-	61.01	29.44
304	Glu	8.81	116.65	-	59.42	28.17
305	Glu	7.55	117.88	-	57.78	-
306	Leu	8.57	123.04	-	58.52	41.14
307	Ile	8.2	120.72	-	63.76	35.78
308	Lys	7.75	121.97	-	60.11	31.52
309	Phe	7.99	116.64	-	59.14	40.05
317	Ser	9.39	119.57	176.42	57.36	62.96
318	Glu	9.07	132.47	-	60.11	28.43
319	Glu	8.54	119.03	179.92	59.25	28.45
320	Arg	7.34	118.88	-	58.84	29.27
321	Ile	8.21	121.09	-	61.19	33.78
322	Arg	9.01	120.15	179.81	60.83	28.48
323	Ser	8	115.22	-	61.33	-
324	Gly	8.31	110.51	174.78	47.28	-
325	Val	8.85	122.43	-	66.88	30.34
332	Arg	7.52	121.77	176.96	56.57	29.44
333	Gln	7.87	120.02	176.72	56.13	-
334	Gly	8.25	109.64	-	45.17	-
336	Thr	8.13	116.42	-	61.74	69.27
337	Leu	8.1	124.45	177.21	54.96	41.28
338	Glu	8.22	122.34	176.25	56.2	29.42
339	Val	7.98	121.94	175.73	61.85	31.74
340	Leu	8.1	125.97	176.79	54.54	41.55
341	Phe	8.09	121.41	174.69	57.15	38.68
342	Gln	7.74	126.75	180.27	56.89	29.85

A2: ^{1}H, ^{15}N and ^{13}C chemical shifts from the hFEN1$_{K93A}$-DHPS1 protein complex. Also found in the BMRB under accession code 27404.