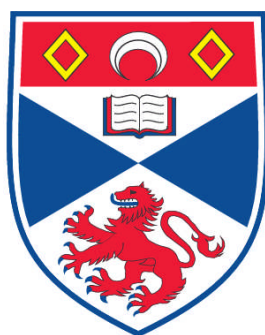# SPLITTING, JOINING AND CUTTING : MECHANISTIC STUDIES OF ENZYMES THAT MANIPULATE DNA

Anne-Marie Maclean McRobbie

A Thesis Submitted for the Degree of PhD
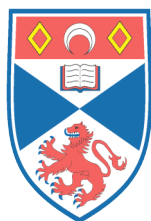## at the
University of St. Andrews

2010

SPLITTING, JOINING AND CUTTING

MECHANISTIC STUDIES OF ENZYMES THAT MANIPULATE DNA

ANNE-MARIE MACLEAN MCROBBIE

A THESIS SUBMITTED TO THE UNIVERSITY OF ST ANDREWS
FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

JUNE 2010

University
of
St Andrews

for Douglas and the bump…

TABLE OF CONTENTS

3 MOLECULAR BASIS OF XPD-RELATED DISEASES   78

CONTENTS

CD-ROM Includes movie entitled 'Hel308 translocation'

FIGURES AND TABLES

FIGURES AND TABLES

ABBREVIATIONS

6-4 PPs	6-4 Photoproducts
3'	3 prime DNA end
5'	5 prime DNA end
[γ-^{32}P] ATP	Adenosine triphosphate with a 32-phosphate radioactive isotope in the gamma phosphate position
A_{600}	Absorbance at 600 nm
ADP	Adenosine 5'-diphosphate
Afu	*Archaeoglobus fulgidus*
AMP-PNP	Adenosine 5'-(β, γ-imido) triphosphate (non-hydrolysable analog of ATP)
Ape	*Aeropyrum pernix*
AP site	Apurinic / apyrimidine site
APS	Ammonium persulphate
ATP	Adenosine 5'-triphosphate
BA	Bathophenanthroline
bp	Base pair
BSA	Bovine serum albumin
CAK	Cdk-activating kinase
CD	Circular dichroism
Cdk	Cyclin-dependent kinase
CPD	Cyclobutane pyrimidine dimer
CS	Cockayne Syndrome
CV	Column volume
DinG	damage inducible helicase G
Dme	*Drosophila melanogaster*
DMSO	Dimethyl sulfoxide
ds / ssDNA	Double / single-stranded deoxyribonucleic acid
DRD	Damage recognition domain
D / SSB	Double / single strand break
DTT	1,4 – dithiothreitol
EDTA	Ethylenediaminetetraacid acid

EMSA	Electrophoretic mobility shift assay
EPR	Electron paramagnetic resonance
ERCC1	Excision repair cross complementing 1
ESI-TOF MS	Electrospray ionisation-time of flight mass spectrometry
FancJ	Fanconi anemia complementation group J
Fe^{2+}	Ferrous iron
Fe^{3+}	Ferric iron
FeS	Iron-sulphur
FRET	Fluorescence resonance energy transfer
GGR	Global genome repair
H_2O	Ultrapure water
H_2O_2	Hydrogen peroxide
HAT	Histone acetyltransferase
HCl	Hydrochloric acid
HD	Helicase domain
HGT	Horizontal gene transfer
HhH	Helix-hairpin-helix
HJ	Holliday junction
HR	Homologous recombination
ICLs	Interstrand crosslinks
IPTG	Isopropyl-beta-D-thiogalactopyranoside
K_D	Dissociation constant
(k)Da	(Kilo)Dalton
KOH	Potassium hydroxide
LB	Luria Bertani
LGT	Lateral gene transfer
MALDI-TOF	Matrix-assisted laser desorption/ ionisation-time of flight
MCM	Minichromosome maintenance
Mfd	Mutation frequency decline
MS	Mass spectrometry
MTSL	(1-oxyl-2,2,5,5-tetramethylpyrroline-3-methyl)-methanethio-sulfonate spin label
Mth	*M. thermautotrophicus*
MW	Molecular weight

NER	Nucleotide excision repair
NHEJ	Non-homologous end joining
nt	Nucleotide
NTP	Nucleotide triphosphate
OD	Optical density
PBL 2025	Paul Bloom 2025 strain of *Sulfolobus solfataricus*
PBS	Phosphate buffered saline
PCA	Perchloric acid
PCNA	Proliferating cell nuclear antigen
Pfu	*Pyrococcus furiosus*
(RT) PCR	(Reverse transcriptase) polymerase chain reaction
PEG	Polyethylene glycol
PIP box	PCNA interacting protein box
PK	Protein kinase
PNK	Polynucleotide kinase
RFC	Replication factor C
RNA	Ribonucleic acid
RNAP	RNA polymerase
RPA	Replication protein A
Rpm	Revolutions per minute
RT-PCR	Reverse transcriptase-polymerase chain reaction
RTS	Rothmund Thomson Syndrome
Sac	*Sulfolobus acidocaldarius*
Sce	*Saccharomyces cerevisiae*
SDS	Sodium dodecyl sulphate
SDSL	Site-directed spin-labelling
SDS-PAGE	SDS-polyacrylamide gel electrophoresis
SF1/2/3/4/5/6	Superfamily 1/2/3/4/5/6 (helicase)
sm	Single-molecule
SSB	Single-stranded DNA binding protein
Sso	*Sulfolobus solfataricus*
Sto	*Sulfolobus tokodaii*
T7g4p	Bacteriophage T7 gene 4 protein
Tac	*Thermoplasma acidophilum*

TCR	Transcription-coupled repair
TBE	Tris-borate EDTA
TEMED	Tetramethylethylenediamine
TEV	Tobacco Etch virus
TFIIH	Transcription factor II H
TLS	Translesion synthesis
TTD	Trichothiodystrophy
UV	Ultraviolet
WS	Werner's Syndrome
WT	Wild type
XP	*Xeroderma pigmentosum*
XPA-G	*Xeroderma pigmentosum* complementation group A-G

DECLARATION

I, Anne-Marie Maclean McRobbie, hereby certify that this thesis, which is approximately 65,0000 words in length, has been written by me, that it is the record of work carried out by me and that it has not been submitted in any previous application for a higher degree.

Date …………………… Signature of candidate …………………………….

I was admitted as a research student in October 2006 and as a candidate for the degree of PhD in October 2007; the higher study for which this is a record was carried out in the University of St Andrews between 2006 and 2009.

Date …………………… Signature of candidate …………………………….

I hereby certify that the candidate has fulfilled the conditions of the Resolution and Regulations appropriate for the degree of PhD in the University of St Andrews and that the candidate is qualified to submit this thesis in application for that degree.

Date …………………… Signature of supervisor …………………………….

In submitting this thesis to the University of St Andrews we understand that we are giving permission for it to be made available for use in accordance with the regulations of the University Library for the time being in force, subject to any copyright vested in the work not being affected thereby. We also understand that the title and the abstract will be published, and that a copy of the work may be made and supplied to any bona fide library or research worker, that my thesis will be electronically accessible for personal or research use unless exempt by award of an embargo as requested below, and that the library has the right to migrate my thesis into new electronic forms as required to ensure continued access to the thesis. We have obtained any third-party copyright permissions that may be required in order to allow such access and migration, or have requested the appropriate embargo below.

Access to Printed copy and electronic publication of thesis through the University of St Andrews.

Date …………………… Signature of supervisor …………………………….

ABSTRACT

DNA is a reactive and dynamic molecule that is continually damaged by both exogenous and endogenous agents. Various DNA repair pathways have evolved to ensure the faithful replication of the genome. One such pathway, nucleotide excision repair (NER), involves the concerted action of several proteins to repair helix-distorting lesions that arise following exposure to UV light. Mutation of NER proteins is associated with several genetic diseases, including *xeroderma pigmentosum* that can arise upon mutation of the DNA helicase, XPD. The consequences of introducing human mutations into the gene encoding XPD from *Sulfolobus acidocaldarius* (SacXPD) were investigated to shed light on the molecular basis of XPD-related diseases.

XPD is a 5'-3' DNA helicase that requires an iron-sulphur (FeS) cluster for activity (Rudolf et al., 2006). Several proteins related to SacXPD, including human XPD, human FancJ and *E. coli* DinG, also rely on an FeS cluster for DNA unwinding (Rudolf et al., 2006; Pugh et al., 2008; Ren et al., 2009). Sequence analysis of the homologous protein, DinG, from *Staphylococcus aureus* (SarDinG) suggests that this protein does not encode a FeS cluster. In addition, SarDinG comprises an N-terminal extension with homology to the epsilon domain of polymerase III from *E. coli*. This thesis describes the purification and characterisation of SarDinG.

During replication, DNA lesions or other 'roadblocks', such as DNA-bound proteins, can lead to replication fork stalling or collapse. To maintain genomic integrity, the fork must be restored and replication restarted. In archaea, the DNA helicase Hel308 is thought to play a role in this process by removing the lagging strands of stalled forks, thereby promoting fork repair by homologous recombination. Potential roles of Hel308 during replication fork repair are discussed in this thesis. The mechanism by which Hel308 moves along and unwinds DNA was also investigated using a combined structural and biophysical approach.

The exchange of DNA between homologous strands, catalysed by a RecA family protein (RecA in bacteria, RAD51 in eukaryotes, and RadA in archaea), defines homologous recombination. While bacteria encode a single RecA protein, both eukaryotes and archaea encode multiple paralogues that have implications in the regulation of RAD51 and RadA activity, respectively. This thesis describes the purification and characterisation of one of the RadA paralogues (Sso2452) in archaea.

ACKNOWLEDGEMENTS

First and foremost, thank-you to my supervisor Professor Malcolm White for his support, enthusiasm and guidance. You have inspired me and made me confident in my own abilities, something that will last much longer than the three years I have been in your lab. Dr Carlos Penedo-Esteiro, thank-you very much for your collaboration on the single-molecule studies of Hel308. It was a real joy to work with you. You have been a fantastic teacher and sacrificed so much time to help me – I am extremely grateful. Thanks also to Dr Olav Schiemann for his collaboration on the EPR study of Hel308.

My PhD was funded by the BBSRC and I thank them for giving me the opportunity to learn and 'do' science, as well as providing financial means to travel both within and outside the UK to attend scientific meetings.

I would like to thank members of the Scottish Structural Proteomics Facility (SSPF) for their contributions to my work, as well as the mass spectrometry team (Dr Catherine Botting, Dr Sally Shirran and Alex Houston) who have analysed endless samples for me and tolerated my relentless questioning. Many thanks to Biljana Petrovic-Stojanovska for her help generating protein mutants – you are definitely the green-fingered molecular biologist of the lab! Thanks also extend to Margaret Wilson and Jean Johnston for whom I think we would all be lost without.

Of course, the extent to which you enjoy any job depends on the people you work with everyday. I have been lucky enough to share the lab space with a fantastic group of people, so thanks to all members (past and present) of the Malcolm White, Carlos Penedo-Esteiro and Peter Coote lab.

Thanks to my parents, who never seem to doubt my ability to do anything. I am sure you are as pleased as I am that this thesis is now complete! And to my parents-in-law, for all your support and encouragement. Finally, my warmest thanks go to my husband and our newborn son, Euan – you are the reasons for everything I have and ever will do.

CHAPTER 1

INTRODUCTION

1.1 ARCHAEA

Classic taxonomic differentiation of species, relying on cell morphology, physiology and pathogenicity, revealed a bipartite organisation of life divided into prokaryotes and eukaryotes. In 1977, Woese and Fox performed a comparative analysis of small-subunit ribosomal RNA sequences from the entire spectrum of living organisms (Woese and Fox, 1977), which revealed that the prokaryotic domain was more diverse than previously thought and necessitated sub-division to create the domains of Bacteria and Archaebacteria (later renamed Archaea). The new archaeal grouping was recognised as an ancient branch of life with a unique metabolism well suited to the conditions of early earth. Consequently, the archaea are often regarded as the closest living relatives to the last universal common ancestor (LUCA) of life on earth (White, 2003). It should be noted that the tree of life, presented in figure 1.1, represents a very simplified account of evolutionary relationships and does not highlight the massive impact of lateral gene transfer between various phyla during prokaryotic evolution (Koonin et al., 2001; Boucher et al., 2003).

The elucidation of several archaeal sequences led to further differentiation of the primary archaeal grouping into the diverse Euryarchaeota, consisting of methanogenic and halophilic archaea, as well as some thermophilic and psychrophilic organisms, and the Crenarchaeota, which predominantly include hyperthermophilic species (Barns et al., 1996; Allers and Mevarech, 2005). It has been suggested that two additional phyla of the archaeal domain exist, namely the Korarchaeota (Barns et al., 1996) and the Nanoarchaeota (Huber et al., 2002; Brochier et al., 2005; Forterre et al.,

2009). However, these groupings remain controversial and may represent an artefact of phylogenetic analysis (Brochier et al., 2005; Forterre et al., 2009).
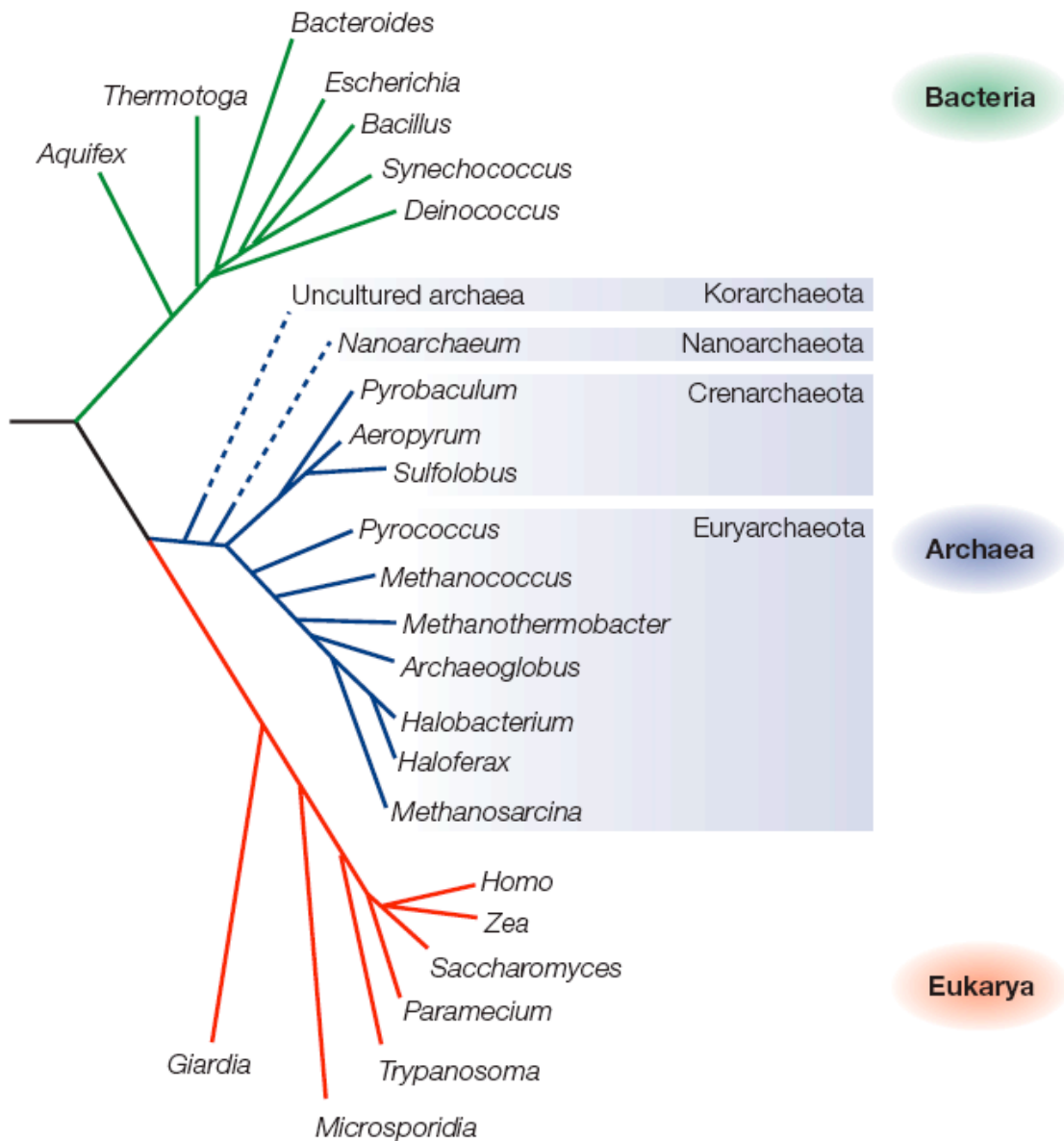
Figure 1.1: The Tree of Life

In the 1970s, analysis of small-subunit ribosomal RNA (rRNA) revealed that life could be divided into three taxonomic groups: the eukarya, archaea and bacteria. Four archaeal groups have been identified, namely the Euryarchaeota, Crenarchaeota, Korarchaeota and Nanoarchaeota. Adapted from (Allers and Mevarech, 2005).

Although the archaea resemble bacteria in a morphological and metabolical sense, it appears that their informational processing pathways (including DNA replication, recombination, transcription, translation and repair) are most similar to that of eukarya (Bell and Jackson, 1998; White, 2003). The archaeal system, however,

represents a stream-lined system utilising only a subset of factors employed in more complex eukaryal processes (Bell and Jackson, 1998). Together with the inherent stability of archaeal proteins, this makes the archaea a very attractive system to unravel the complexities of eukaryotic biology.

The work presented in this thesis outlines the characterisation of several conserved archaeal proteins involved in DNA informational pathways. These proteins were derived from two crenarchaeal organisms, *Sulfolobus acidocaldarius* (strain DSM639) and *Sulfolobus solfataricus* (strains PBL2025 and P2), which are considered aerobic thermoacidophiles with optimal growth occurring at 80 ºC in sulphur-rich environments of pH 2 - 4 (She et al., 2001; Chen et al., 2005). The genomes of both species have been fully sequenced (She et al., 2001; Chen et al., 2005), enabling cloning and heterologous expression of recombinant *Sulfolobus* proteins.

1.2   DNA DAMAGE AND REPAIR

Faithful replication of the genome is central to cell survival. The adoption of DNA as the genetic molecule suggests that it must be inherently stable and chemically inert. However, DNA is both reactive and dynamic and, as a result, is vulnerable to damage from both exogenous and endogenous sources (figure 1.2).

In the former case, UV and ionising radiation and genotoxic chemicals, including anti-tumour agents such as cis-platin, are potentially mutagenic (Hoeijmakers, 2001). For instance, UV-radiation is associated with the formation of dipyrimidine photoproducts, including cyclobutane pyrimidine dimers (CPD) and (6-4) photoproducts (Lindahl and Wood, 1999). These agents introduce bulky, helix-distorting lesions to DNA that influence the structural integrity of the molecule.

Endogenous sources of DNA damage arise from the by-products of normal cellular metabolism and include reactive oxygen species (ROS), such as superoxide anions, hydroxyl radicals and hydrogen peroxide (Truglio et al., 2006a). ROS are created during aerobic respiration and lipid peroxidation and can lead to the oxidation of bases (including 8-oxo guanine and thymine glycol) that inhibit DNA replication and transcription (Hoeijmakers, 2001). Further endogenous sources of DNA damage result from deregulation of the metabolic cofactor, S-adenosylmethionine, which acts as a methyl donor in the cell. Uncontrolled methylation of bases is associated with increased cancer risk (Lindahl and Wood, 1999).

Finally, spontaneous DNA damage represents a major threat to genomic integrity. Hydrolysis of the glycosidic bond linking a purine to its deoxyribose sugar is a prominent source of damage in the cell, leading to the inclusion of abasic sites in the DNA structure that can, in turn, lead to the formation of single-strand DNA breaks (Hoeijmakers, 2001). Base deamination also exists as an inherent problem that leads to base mismatches, for instance deamination of cytosine generates uracil, which will base pair with adenine. Subsequent rounds of replication will alter the genetic sequence from a C-G to an A-T pair.
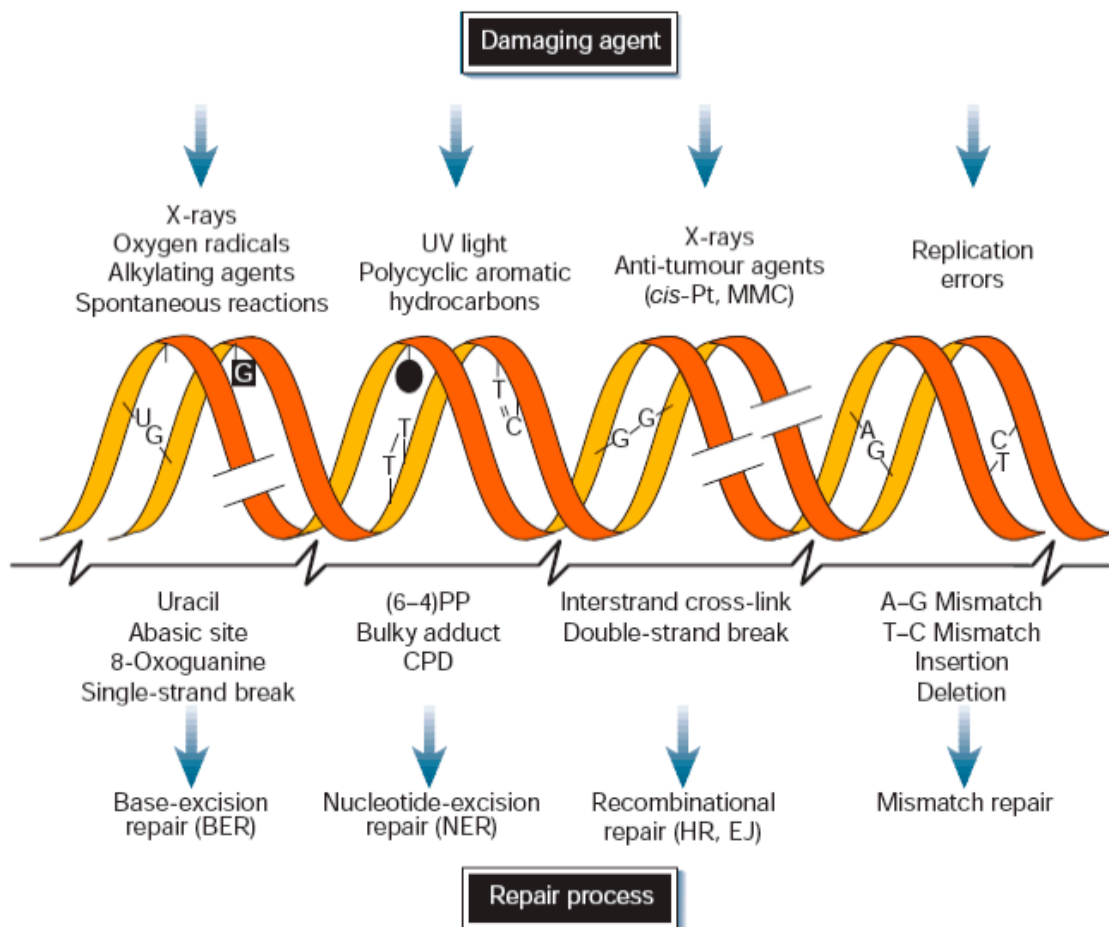
Figure 1.2: DNA damage and repair

A variety of endogenous and exogenous agents (top) generate a myriad of DNA lesions (middle). To overcome the mutagenic impact of these lesions, cells have evolved a set of repair pathways (bottom) to restore genomic integrity. Adapted from (Hoeijmakers, 2001).

To reflect the repertoire of DNA lesions, a labyrinth of biochemical strategies exists to protect the informational and structural integrity of DNA (summarised in figure 1.2). This project focuses on the characterisation of proteins involved in the nucleotide excision repair (NER) and homologous recombination (HR)

CHAPTER 1 | INTRODUCTION

pathways. The NER pathway is responsible for repairing UV-induced helix-distorting lesions, including CPD and (6-4) photoproducts. The HR pathway is the principal pathway involved in processing double strand breaks (DSBs).

1.2.1 NUCLEOTIDE EXCISION REPAIR

Although the process of NER has been conserved during evolution, the factors involved in the pathway differ between the three domains of life (Garfinkel and Bailis, 2002). The process of NER can be summarised as the coordinated recognition of DNA damage, unwinding of the duplex around the lesion and excision of the lesion prior to resynthesis and ligation of the correct DNA sequence (de Laat et al., 1999). In bacteria, this process involves four proteins including a DNA helicase, UvrD, and the UvrABC endonuclease complex (Van Houten, 1990; Sancar, 1996; Van Houten et al., 2005). By contrast, eukaryotes orchestrate the coordinated activity of approximately 30 proteins to achieve the same result. Initial studies with *Methanothermobacter thermoautotrophicum* revealed clear UvrABC homologues leading to proposals that the archaea employ a bacterial-like NER system (Ogrunc et al., 1998). However, it has become apparent that while some euryarchaeal organisms do use a bacterial-like system, the majority possess eukaryotic homologues (White, 2003).

NER can be subdivided into two distinct pathways, namely global genome repair (GGR), which repairs lesions throughout the genome, and transcription-coupled repair (TCR), which is responsible for the repair of certain types of DNA damage that arise on the actively-transcribed strand during transcription elongation.

1.2.1.1 BACTERIAL NER

Bacterial GGR is initiated upon formation of either a heterotrimeric (Theis et al., 2000) or heterotetrameric (Verhoeven et al., 2002) tracking complex, comprising two UvrA and one or two UvrB molecules ($UvrA_2B$ or $UvrA_2B_2$, respectively) that scan the genome for sites of damage. UvrA scans DNA for structural perturbations and prompts localised unwinding of the duplex to enable DNA binding by UvrB and formation of the stable pre-incision complex (PIC) (Truglio et al., 2006b). UvrB serves to verify the presence of DNA damage by inserting a β-hairpin loop between the two strands of the duplex, thus causing the nucleotide directly behind the loop to flip out of helical alignment and into a hydrophobic binding pocket (figure 1.3) (Truglio et al.,

2006b; Waters et al., 2006; Malta et al., 2008). This pocket is small and cannot accommodate bulky adducts, such as UV-induced CPDs, due to steric hindrance (Verhoeven et al., 2002; Malta et al., 2006; Truglio et al., 2006b; Waters et al., 2006). ATP-dependent 3'-5' translocation of UvrB along the DNA strand allows damage verification by sequential base flipping; upon meeting a bulky adduct, base flipping into the pocket is prevented and UvrB translocation stalls causing kinking of the DNA molecule by 130 º (Verhoeven et al., 2002).
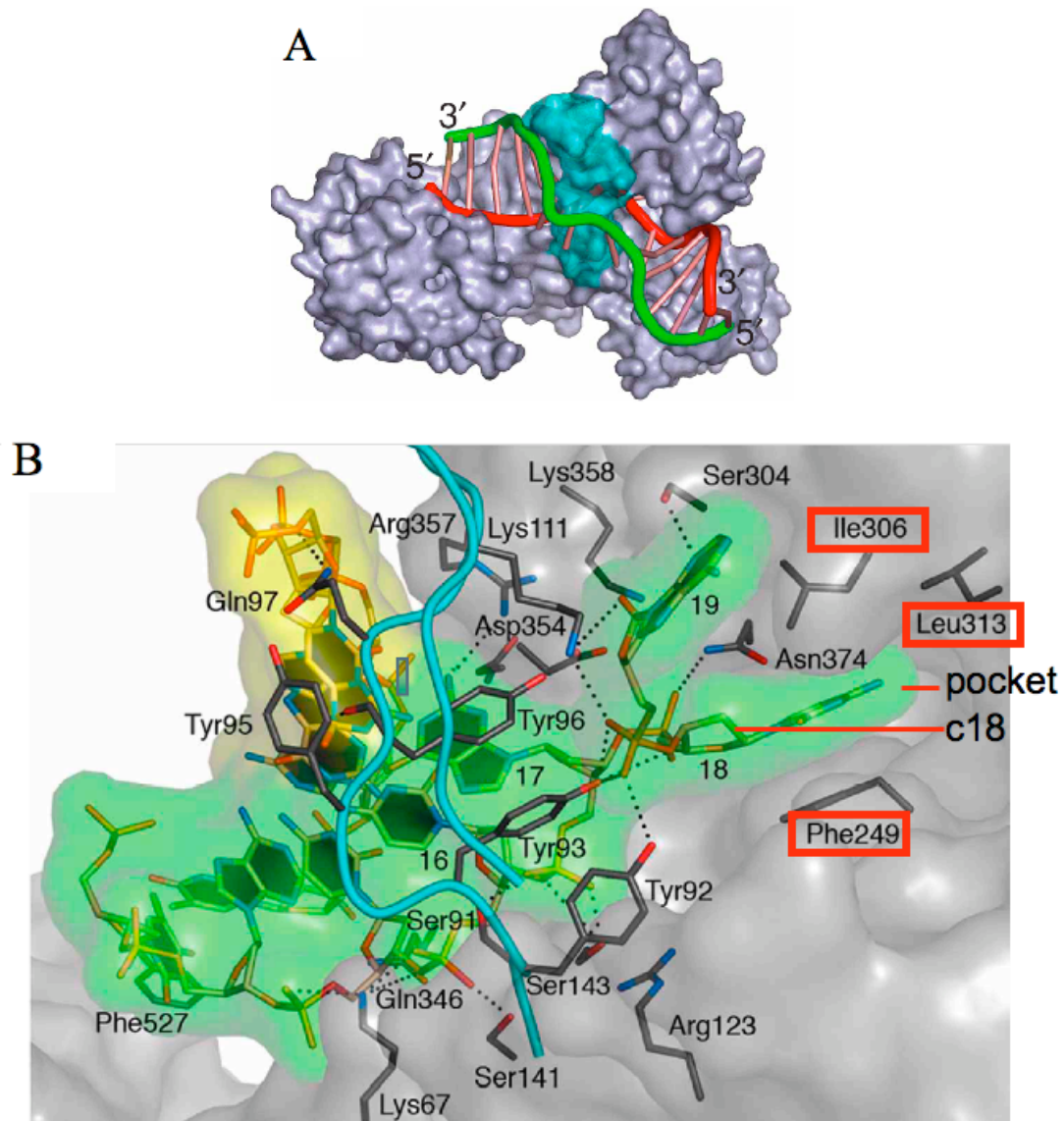
Figure 1.3: Structural basis for DNA damage verification by UvrB

(A) UvrB (grey) binds DNA such that the inner strand (red) passes behind the β-hairpin (cyan) and the outer strand (green) sits in front of the β-hairpin. (B) Nucleotides of the inner strand of DNA (green) pass beneath the β-hairpin (cyan) and rotate to enter a hydrophobic binding pocket that only provides space for planar molecules. The pocket is lined with conserved residues, including Phe249, Leu313 and Ile306, which interact with the base (C18 shown). Adapted from (Truglio et al., 2006b).

This triggers the ATP-dependent dissociation of UvrA (Oh and Grossman, 1986) and subsequent recruitment of UvrC to form the incision complex. UvrC catalyses dual incisions around the site of damage: the first is located 3' of the lesion, between the fourth and fifth phosphodiester bond, while the 5' incision is made at the 8th phosphodiester bond (Sancar and Rupp, 1983; Lin et al., 1992; Lin and Sancar, 1992). The 3'-5' helicase activity of UvrD displaces the excised stretch of 12-13 nucleotides, together with UvrC, prior to the recruitment of DNA polymerase I, which serves to resynthesise the correct sequence of DNA using the undamaged strand as a template. UvrB stimulates the helicase activity of UvrD and only dissociates from DNA upon completion of this stage (Orren et al., 1992; Atkinson et al., 2009). Bacterial GGR is concluded upon the ligation of the newly synthesised strand to the original DNA by DNA ligase I.

The TCR subpathway of bacterial NER is involved in the rapid repair of bulky lesions that occur on actively transcribed genes (Svejstrup, 2002; Reardon and Sancar, 2005). Such lesions block the progression of the elongating RNAP, which subsequently undergoes reverse-translocation (or 'backtracking') such that the 3' end of the RNA transcript is extruded from the transcription bubble (Park et al., 2002; Mellon, 2005; Smith and Savery, 2008). A transcription-repair coupling factor, known as Mfd (or *mutation frequency decline*), is implicated in coupling transcription to the NER pathway and comprises three structural regions: an N-terminal, middle, and C-terminal region. The middle section of Mfd serves to recognise and bind the β-subunit of stalled RNAP and docks onto the DNA strand immediately upstream of the transcription complex (Park et al., 2002). In an ATP-dependent manner, the C-terminal region of Mfd promotes DNA translocation such that RNAP moves forwards relative to the DNA. At sites of bulky lesions, where transcription cannot reinitiate, Mfd translocation is expected to rewind the DNA upstream of the transcription bubble, unwind the DNA-RNA hybrid, and promote RNAP displacement (Selby and Sancar, 1993; Park and Roberts, 2006; Smith and Savery, 2008). Alternatively, Mfd can promote transcription reactivation by enabling lesion bypass by RNAP (Park et al., 2002). This has shown to occur in the presence of non-bulky lesions, including single-strand breaks and abasic sites (Smith and Savery, 2008). After RNAP dissociation, Mfd binds DNA near the downstream end of the transcription bubble. The crystal structure of the N-terminus of Mfd reveals three domains that bear architectural similarity to UvrB (figure 1.4). Of particular note is the similarity between domain 2 of Mfd and the UvrA-interacting

domain of UvrB. Structural studies have shown that the N-terminus of Mfd serves to recruit UvrA to the site of damage after the dissociation of RNAP (Assenmacher et al., 2006).
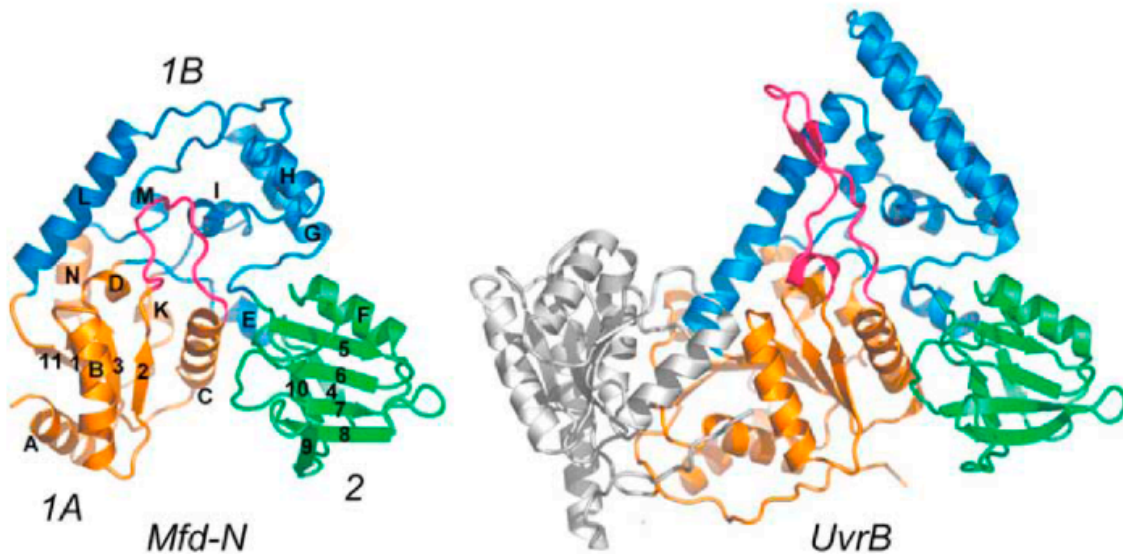
Figure 1.4: Bacterial Mfd resembles UvrB

Crystal structures of the N-terminus of Mfd (Mfd-N) from *E. coli* (left) and UvrB from *Bacillus caldotenax* (right). The three domains of Mfd-N (domain 1a, orange; domain 1b, blue (β-hairpin loop, magenta); domain 2, green) are structurally reminiscent of the three N-terminal domains of UvrB. Adapted from (Assenmacher et al., 2006).

Subsequent recruitment of UvrB to form the pre-incision complex triggers the binding of UvrC, and the TCR pathway proceeds according to the GGR pathway (figure 1.5). The TCR pathway promotes rapid repair of transcription-blocking lesions by directing the assembly of NER factors at the site of damage. This is in contrast to the slow repair of lesions via the GGR pathway as a result of the low cellular concentration of UvrA and UvrB (Smith and Savery, 2008).
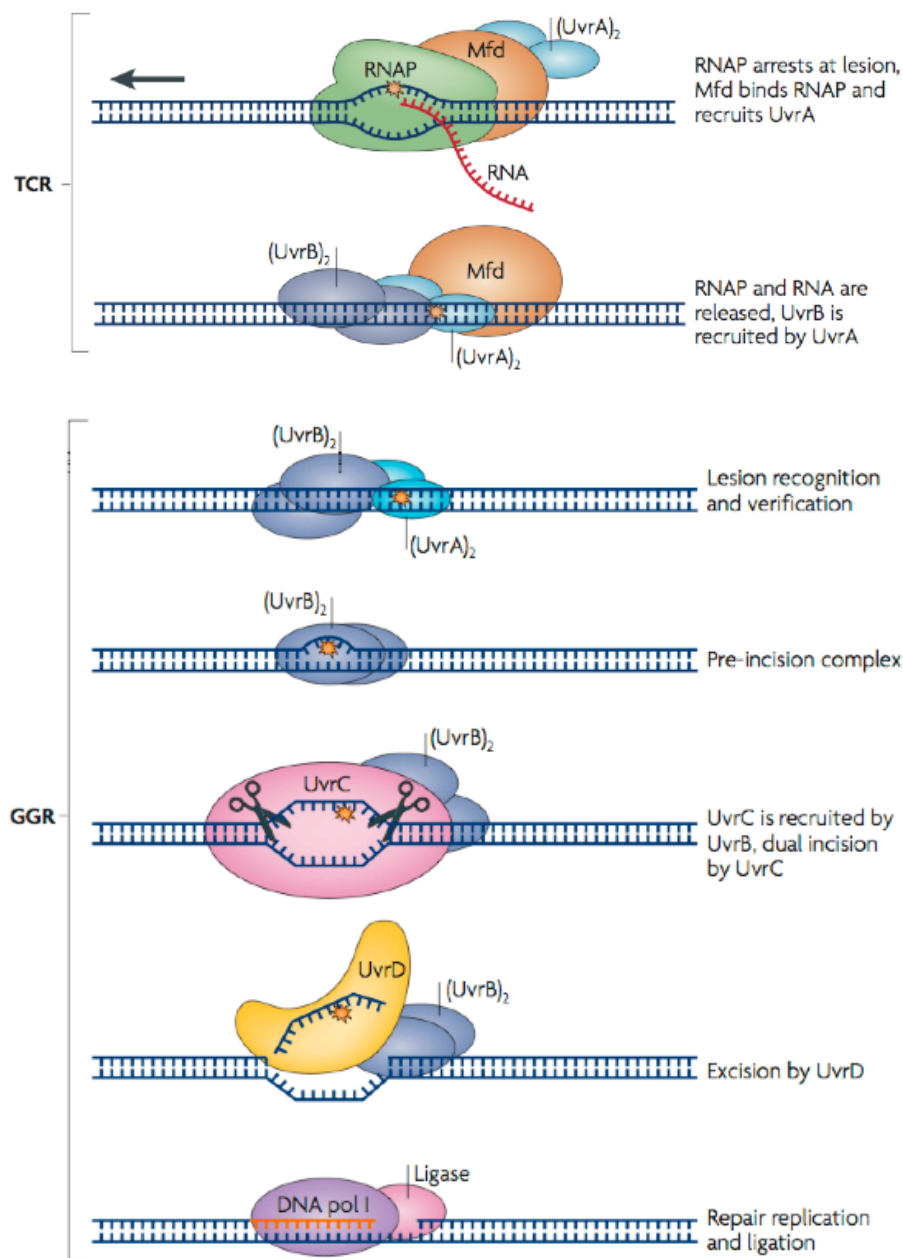
Figure 1.5: Bacterial global-genome and transcription-coupled NER pathway

Bacterial TCR initiates upon stalling of the RNA polymerase (RNAP) at a lesion on the transcribed strand. This prompts the recruitment of Mfd and, in turn, UvrA. UvrA forms a homodimer and recruits UvrB, which verifies the presence of the lesion and promotes Mfd dissociation. The TCR pathway proceeds according to the GGR pathway, in which the heterotrimeric or heterotetrameric assembly of UvrA and UvrB (UvrA$_2$B or UvrA$_2$B$_2$, respectively) recruit UvrC, which makes incisions on either side of the lesion, 12 nucleotides apart. The DNA helicase, UvrD, binds and displaces the damage-containing oligonucleotide and DNA polymerase I (DNA pol I) performs repair synthesis. The repair patch is subsequently sealed by DNA ligase. Modified from (Hanawalt and Spivak, 2008).

1.2.1.2   EUKARYOTIC NER

Unlike bacterial GGR, the eukaryotic pathway involves multiple factors that recognise and bind UV-induced photoproducts and helix-distorting adducts. One such factor is XPC, which, in complex with RAD23 and centrin-2, recognises a wide variety of lesions throughout the genome (Riedl et al., 2003; Nishi et al., 2009). The XPC homolog, Rad4, from *S. cerevisiae* is composed of an N-terminal transglutaminase-like domain (TGD) and three consecutive β-hairpin domains (BHD1-3) (figure 1.6), which scan the genome for disrupted Watson-Crick base pairing. Damage detection results in the insertion of the β-hairpin of BHD3 into the duplex via the major groove, serving to bend the duplex and flip the damaged base out of helical alignment (figure 1.6) (Min and Pavletich, 2007; Sugasawa and Hanaoka, 2007).
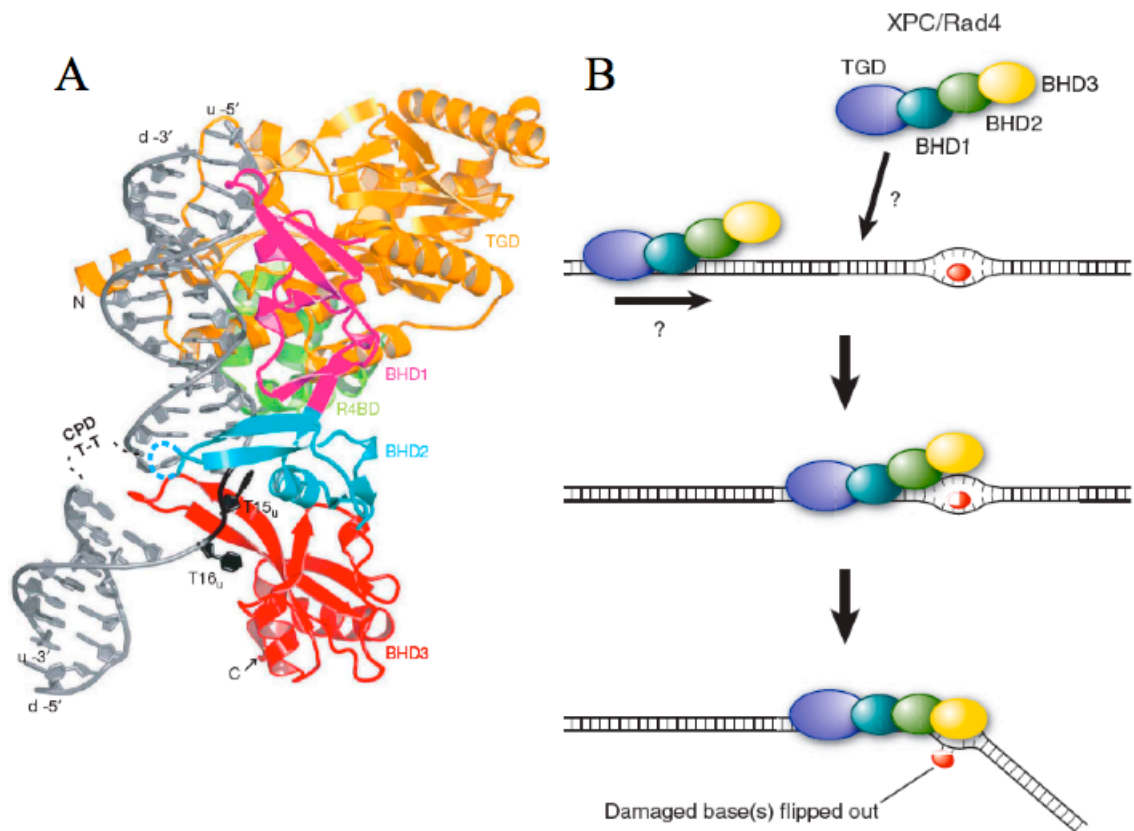
Figure 1.6: DNA damage recognition by XPC/ Rad4

(A) Structure of *S. cerevisiae* Rad4-Rad23 in complex with DNA containing a CPD lesion. TGD, gold; BHD1, magenta; BHD1, cyan; BHD3, red; R4BD domain of Rad23, green; DNA, grey. CPD lesion (residues T15 and T16) in black. (B) RAD4 may employ a random or logical search strategy to identify DNA damage. Domains BHD2 and BHD3 recognise regions of the duplex in which Watson-Crick base pairing have been lost or destabilised, and the β-hairpin of BHD3 inserts into the DNA at the site of damage. This causes the duplex to kink and the damaged bases are flipping out of the helix. Adapted from (Min and Pavletich, 2007; Sugasawa and Hanaoka, 2007).

A second damage-recognition complex, known as the UV-damaged DNA binding (UV-DDB) protein, plays a central role in eukaryotic GGR and exhibits high affinity and specificity for UV-induced lesions. This protein consists of two subunits, DDB1 and DDB2 (figure 1.7), the former involved in mediating multiple protein-protein interactions, including an interaction with a cullin 4A-based ubiquitin E3 ligase, and the latter operating as the damage-recognition component (Keeney et al., 1993). Damage-detection is achieved by DDB2 through the insertion of a highly conserved β-hairpin loop into the minor groove of DNA, which, upon encounter with a photoproduct, flips the affected bases into a binding pocket and induces a 40 ° kink of the duplex (figure 1.7 B-C) (Scrima et al., 2008). This kink serves to partially open the helix such that XPC-mediated binding of TFIIH can proceed.
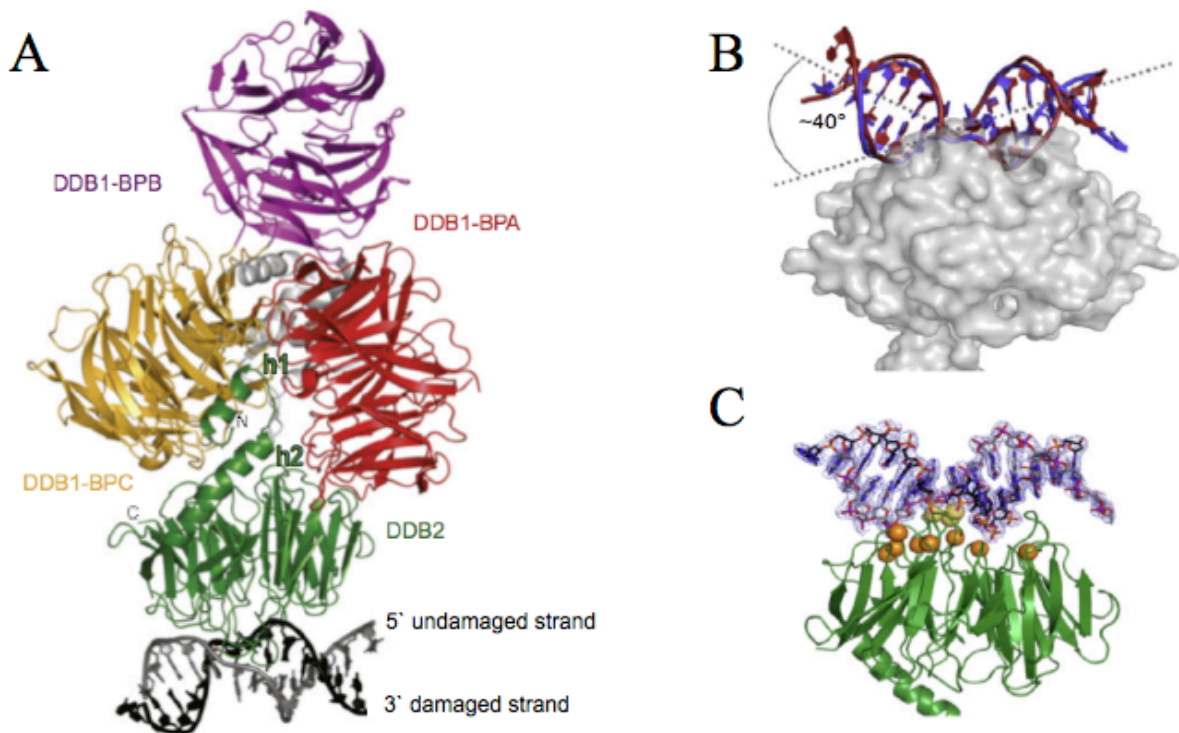
Figure 1.7: Structural basis of DNA damage recognition by UV-DDB complex

(A) Structure of UV-DDB in complex with DNA containing a 6-4PP lesion. The DDB1 subunit is composed of three WD40 β-propeller domains (BPA, BPB and BPC) and a C-terminal helical domain (grey). The DDB2 subunit comprises an N-terminal helix-loop-helix segment and a 7-bladed WD40 β propeller domain. (B) Binding of damaged DNA by DDB2 bends the duplex by 40 °. (C) Residues of the DDB2 β-hairpin (yellow spheres) contact the DNA at the site of damage. Additional residues that interact with the DNA backbone are shown as orange spheres. Adapted from (Scrima et al., 2008).

TFIIH is a complex comprising two DNA helicases, XPB and XPD, which stimulate further unwinding around the site of damage with 3'-5' and 5'-3' polarity, respectively (figure 1.8). This generates an open complex, referred to as the preincision complex (PIC) that is stabilised by RPA (eukaryotic homolog of bacterial SSB) and XPA (Evans et al., 1997; de Laat et al., 1999; Riedl et al., 2003). To excise the lesion, dual incisions are performed by structure-specific endonucleases, ERCC1-XPF and XPG. However, this process must be coordinated with repair synthesis by DNA polymerase δ/ε since dual excision prior to synthesis would result in the formation of a ssDNA gap, which itself would be the target of repair (Staresincic et al., 2009). Consequently, ERCC1-XPF cleaves 5' of the lesion at the $20^{th} \pm 5$ phosphodiester bond, generating a 3'-OH group that serves as the substrate for repair synthesis. DNA synthesis by DNA polymerase is stimulated by the processivity factor, proliferating cell nuclear antigen (PCNA), which is loaded onto the DNA by replication factor C (RFC) (Hanawalt and Spivak, 2008). XPG then cleaves the DNA 3' of the lesion at the $6^{th} \pm 3$ phosphodiester bond prior to release of the damage-containing oligonucleotide and repair synthesis (Sijbers et al., 1996; Mu et al., 1997; Staresincic et al., 2009). The repair patch is sealed into the contiguous DNA strand by DNA ligase 1 (Costa et al., 2003).

Eukaryotic transcription-coupled NER begins following RNA polymerase II (RNAP II) arrest upon meeting a DNA lesion on the actively-transcribed strand. In contrast to global-genome NER, the damage-recognition factors XPC/ RAD23/ centrin-2 and UV-DDB are dispensable for TCR (Mellon, 2005; Reardon and Sancar, 2005). RNAP II stalling presents a fundamental problem: the footprint of the polymerase extends 10 nucleotides ahead and 25 nucleotides behind the lesion, which prohibits continuation of transcription but also limits access to the site of damage by NER factors (Hanawalt and Spivak, 2008). It is therefore anticipated that RNAP II arrest will precede the recruitment of a transcription-repair coupling factor, similar to bacterial Mfd. The tight association between the stalled transcription complex and CSB suggests that this protein may occupy this role. Like Mfd, CSB is a member of the SWI/ SNF2 family of DNA-dependent ATPases (Troelstra et al., 1992) and exhibits DNA translocase activity (Reardon and Sancar, 2005), which it exploits to promote the reverse translocation of the stalled transcription complex (Fousteri et al., 2006). This activity would expose the transcription bubble and site of damage. RPA would be

required to stabilise and protect the open complex, facilitating entry of NER factors, including TFIIH, which would enable lesion repair according to the GGR pathway (Hanawalt and Spivak, 2008).

Based on a direct association with the stalled transcription complex, a number of accessory proteins are likely to participate in eukaryotic TCR. One such protein is CSA, which forms a complex with a E3-ubiquitin ligase (Groisman et al., 2003) similar to the DDB1 subunit of the damage-recognition factor UV-DDB in global-genome NER. As for DDB1, CSA binds and recruits several proteins to the site of damage including XPA-binding protein-2 (XAB2), the high-mobility group binding domain-containing protein 1 (HMGN1), p300 and TFIIS (Fousteri et al., 2006). The association of XAB2 with XPA, a protein involved in both GGR and TCR, suggests that XAB2 may by essential for the convergence of the two pathways (Hanawalt and Spivak, 2008). A fundamental issue in eukaryotic repair is the compaction, and therefore inaccessibility, of DNA into chromatin. HMGN1 decreases chromatin compaction (Hanawalt and Spivak, 2008) and, therefore, its interaction with XPA may be essential to facilitate entry of NER factors to the DNA. In addition, HMGN1 has been implicated in nucleosome displacement (Hanawalt and Spivak, 2008), which may facilitate uninterrupted reverse-translocation by RNAP II. The size of mammalian genes in comparison to those of bacteria renders it crucial that, in the majority of cases, transcription proceeds, rather than aborts, following polymerase stalling. The recruitment of TFIIS to the site of damage by XPA may favour this outcome as it has been shown to stimulate RNAP II nuclease activity, a function that may be important in cleaving the 3' end of the RNA transcript following its retreat from the transcription bubble during RNAP II back-tracking (Hanawalt and Spivak, 2008)
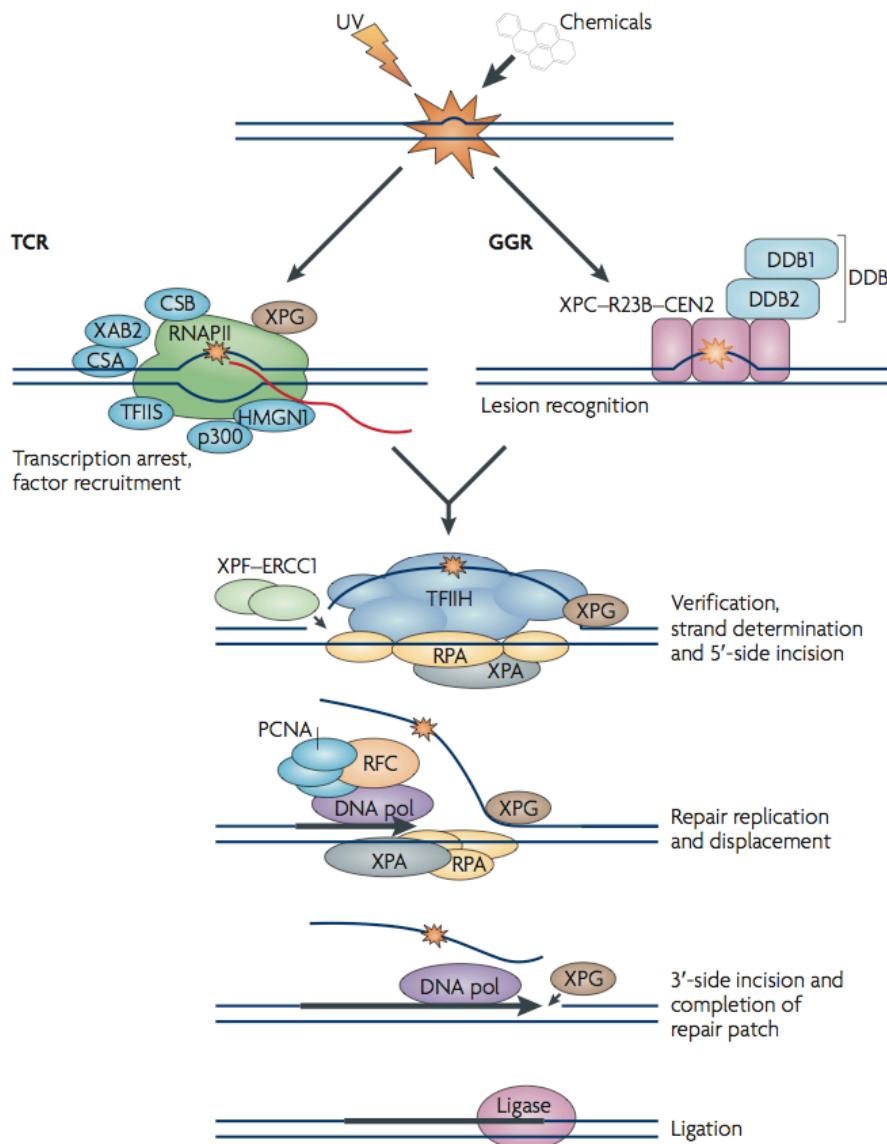
Figure 1.8: Global-genome and transcription-coupled NER in eukaryotes

In TCR, RNAP II stalls upon encountering a lesion, which triggers the recruitment of several factors, including CSA and CSB (see text for details). In GGR, lesions are initially recognized by XPC, in complex with RAD23 (R23B) and centrin-2 (CEN2), and/ or the heterodimer UV-DDB complex comprising DDB1 and DDB2. Following damage-recognition, the TCR and GGR pathways converge and repair lesions accordingly: TFIIH, comprising DNA helicases XPB and XPD, is recruited and catalyses duplex opening. XPA and RPA bind to the open helix to protect the single-stranded region of DNA. ERCC1-XPF cuts the DNA 5' to the lesion and DNA pol initiates repair synthesis. The activity of DNA pol is stimulated by the processivity factor, PCNA, which is loaded by replication factor (RFC). The 3' incision is then made by XPG followed by further repair synthesis and patch repair by DNA ligase. Adapted from (Hanawalt and Spivak, 2008).

1.2.1.3   ARCHAEAL NER

The molecular details of the archaeal NER pathway (figure 1.11) remain poorly understood. Previously, it was assumed that the archaeal NER pathway resembled that of bacteria owing to the inclusion of UvrABC homologues in the genome of *Methanothermobacter thermoautotrophicum* (Smith et al., 1997; Ogrunc et al., 1998). However, this likely represents the product of lateral gene transfer since subsequent studies have showed that the majority of archaeal species possess homologues of eukaryotic RPA (archaeal single-stranded DNA binding protein (SSB)), XPB, XPD, XPF and Fen1 (homologue of eukaryotic XPG), and therefore are likely to adopt an eukaryal-like system (White, 2003). Importantly, archaea do not appear to encode homologues of XPA, XPC/ RAD23/ centrin-2, UV-DDB, CSA nor CSB that are implicated in DNA damage detection in eukaryotic NER (Costa et al., 2003). However, many of these factors are also absent from higher plants, suggesting that they are a relatively recent addition to the NER pathway (Kelman and White, 2005). The ability of SSB to bind and melt mismatched DNA templates has led to the hypothesis that it may represent the damage recognition component of archaeal NER (Cubeddu and White, 2005). Alternatively, the archaeal homologue of eukaryotic XPB may also play a role in damage recognition. The XPB protein from *Archaeoglobus fulgidus* comprises five structural domains, including two RecA-like motor domains (HD1 and HD2), a DNA damage recognition domain (DRD), a RED motif implicated in helicase activity, and a flexible thumb motif (ThM) (Fan et al., 2006). Although the protein preferentially unwinds DNA with a 3' single-stranded tail, unwinding activity has been observed in the presence of fully base-paired duplex DNA that contains a damaged site (Fan et al., 2006). Structural studies revealed that binding of damaged DNA by XPB induces rotation of HD2 relative to HD1 to produce an active, closed conformation that enveloped the DNA molecule (figure 1.9). This re-positions the RED motif to facilitate duplex unwinding. In eukaryotes, XPB is part of the multi-subunit TFIIH complex where it functions during both transcription initiation and NER. The ability of XPB to recognise DNA damage may determine TFIIH function, switching it between transcription and NER depending on the presence of DNA damage (Fan et al., 2006). The helicase activity of XPB opposes that of archaeal XPD, which has been shown to possess ATP-dependent 5'-3' helicase activity (Rudolf et al., 2006), an activity that relies on the presence of an iron-sulphur cluster. The structure of XPD from *S.*

acidocaldarius revealed that this cluster is ideally positioned to physically separate the two strands of a DNA duplex (Fan et al., 2008) (discussed in detail in chapter 3).
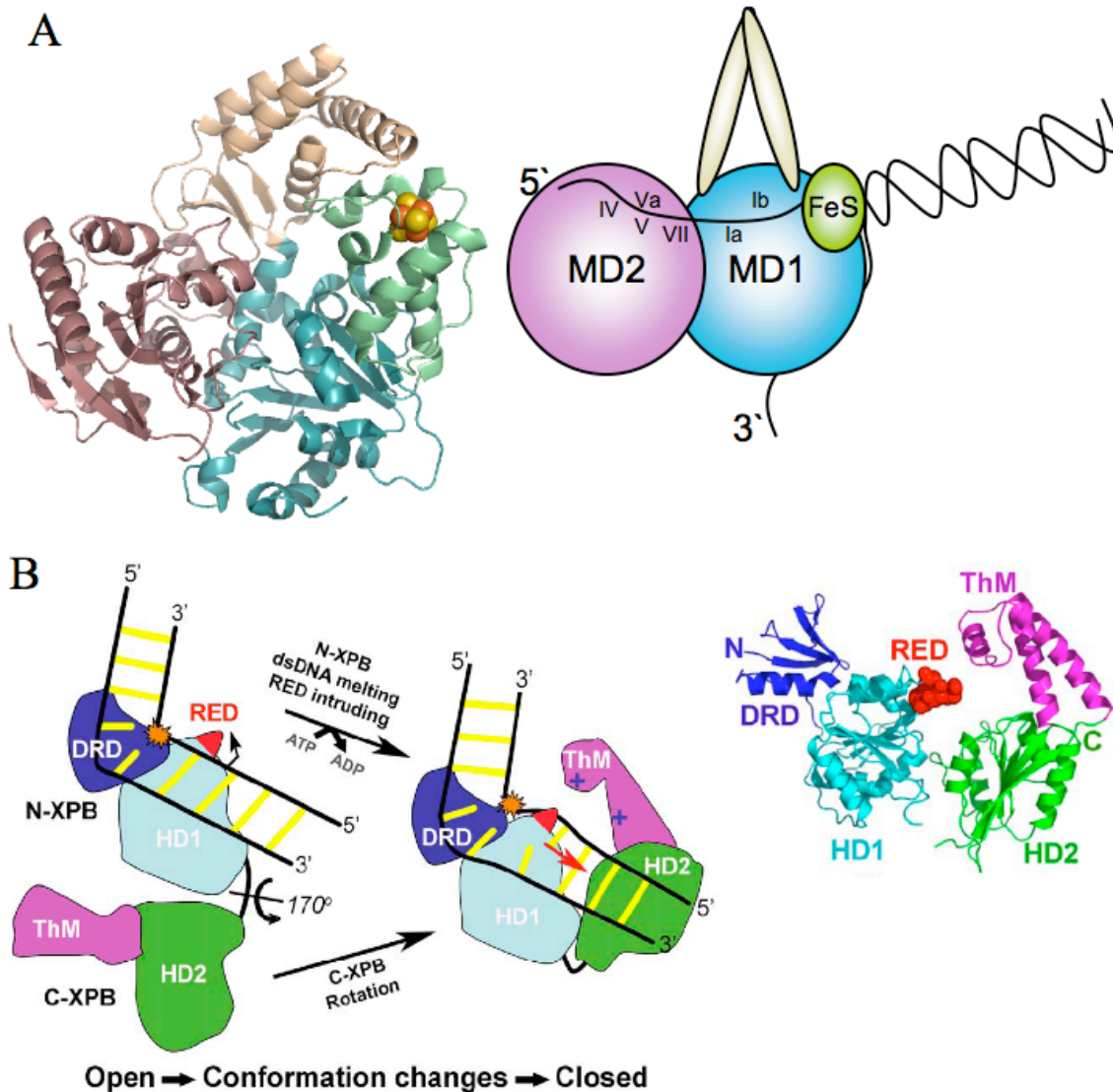
Figure 1.9: Structures of archaeal XPD and XPB and proposed mechanism of action

(A) Left: Structure of XPD from *S. acidocaldarius* (left) with domains coloured as follows: motor domain 1 (MD1), cyan; motor domain 2 (MD2), pink; FeS domain, green (FeS cluster shown as space-fill); arch domain, wheat. Right: DNA enters XPD via MD2, where the FeS cluster, Arch domain and MD1 form a tunnel through which ssDNA is expected to pass. The FeS cluster is ideally positioned to physically separate the duplex. (B) Right: Structure of XPB from *A. fulgidus* with structural domains coloured as follows: HD1 and HD2 (RecA-like motor domains), cyan and green, respectively; damage recognition domain (DRD), blue; RED motif, red; flexible thumb motif (ThM), magenta. Left: Binding of XPB to dsDNA (yellow) containing a lesion (orange) in the open conformation. Verification of damaged site by the DRD induces rotation of XPB between HD1 and HD2 to form the closed complex. This re-positions the RED motif such that DNA unwinding can take place (as indicated by the red arrow). (B) adapted from (Fan et al., 2006).

Archaeal XPF, in complex with the heterotrimeric proliferating cell nuclear antigen (PCNA), is a structure-specific nuclease that cleaves asymmetric, branched substrates including nicked Holliday junctions, D-loops and 3' flap substrates (Roberts et al., 2003; Roberts and White, 2005). XPF from *Aeropyrum pernix* (ApeXPF) is composed of a catalytic domain linked, via a flexible loop region, to a DNA binding domain consisting of two helix-hairpin-helix motifs ((HhH)$_2$) (Newman et al., 2005). Catalytic activity requires dimerisation of XPF, which is facilitated by the linker region. Binding of DNA to one of the DNA binding domains of the dimer causes the linker region to flex such that the associated nuclease domain of the same monomer is brought into close proximity to form a compact and active conformation (Newman et al., 2005) (figure 1.10). The ability to independently activate one nuclease protomer ensures that the DNA is cleaved asymmetrically. PCNA increases the catalytic rate of XPF (Hutton et al., 2008), perhaps by promoting the structural rearrangement of XPF upon DNA binding (Newman et al., 2005).
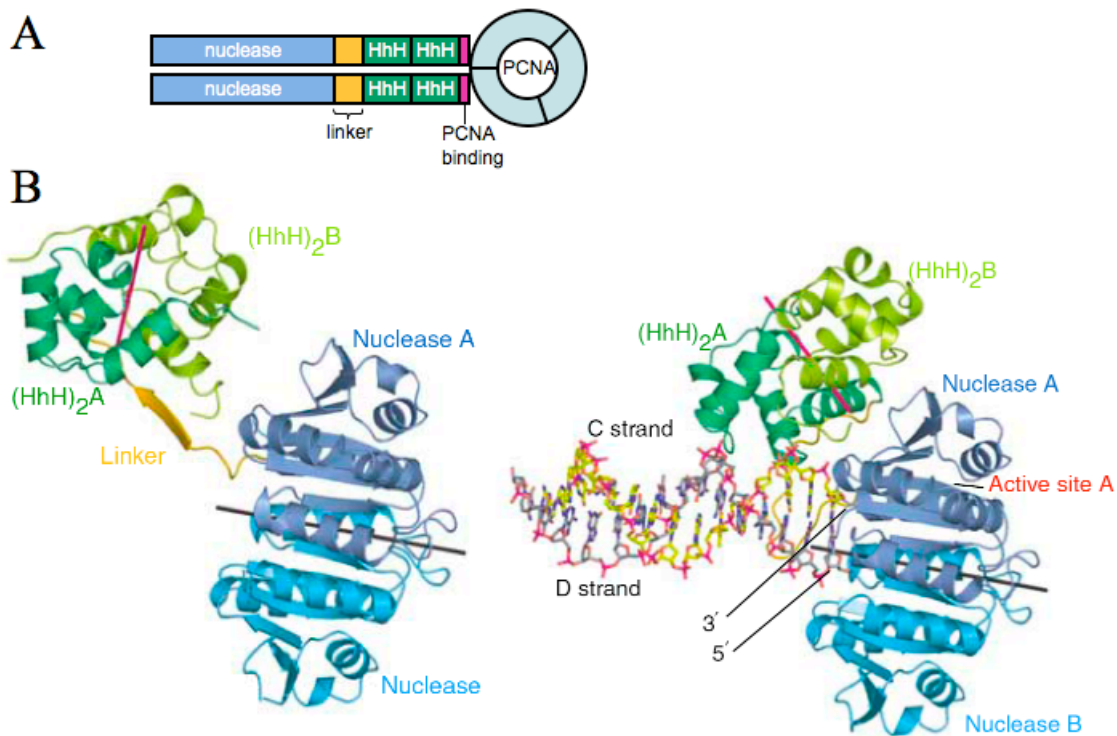
Figure 1.10: Structural basis for asymmetric cleavage by archaeal XPF

(A) Domain organisation of crenarchaeal XPF (dimeric) with associated PCNA. HhH, helix-hairpin-helix motif. (B) Structure of the apo (left) and DNA-bound (right) conformations of dimeric XPF from *Aeropyrum pernix*. Domains coloured as indicated in (A). DNA binding in one orientation flexes the linker to alter the relative position of the (HhH)$_2$ and nuclease domain of monomer A to promote the 3' cleavage of DNA. Adapted from (Newman et al., 2005).

Archaea also possess a homologue of eukaryotic Fen-1, and studies have shown that the archaeal protein is a 5' flap endonuclease. Binding of Fen-1 to PCNA increases its affinity for DNA, leading to enhanced DNA cleavage (Hutton et al., 2008). The absence of proteins homologous to bacterial Mfd or eukaryotic CS proteins suggests that the archaea do not perform transcription-coupled NER.
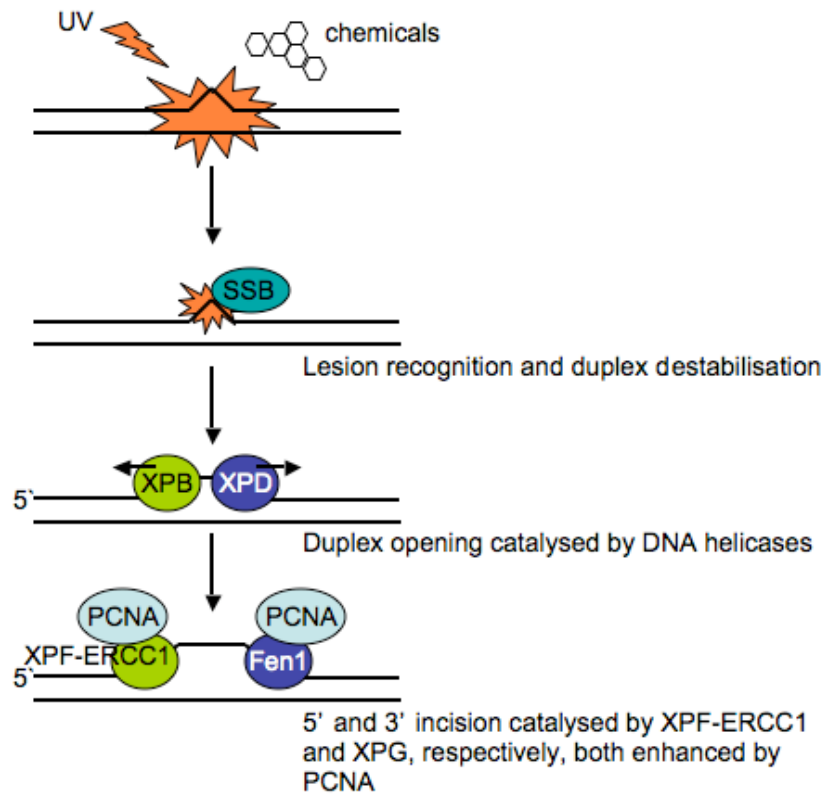
Figure 1.11: Archaeal NER pathway

The single-stranded DNA binding protein (SSB) has been implicated in the initial DNA damage detection event of archaeal NER. SSB can bind and destabilise damaged DNA (Cubeddu and White, 2005), which would permit entry of the DNA helicases XPB and XPD that unwind duplex DNA with 3'-5' and 5'-3' polarity, respectively. The resulting open complex would be bound by structure-specific endonucleases, ERCC1-XPF and Fen-1, to excise the damage-containing stretch of DNA. The activity of both endonucleases is enhanced by processivity factor PCNA. The pathway would be anticipated to conclude via patch repair DNA synthesis followed by ligation of the newly synthesised strand into the original duplex.

1.2.2   HOMOLOGOUS RECOMBINATION

Homologous recombination (HR) is a high-fidelity, template-dependent repair process that comprises a group of inter-related pathways that function in the repair of DNA single-strand breaks (SSBs), double-strand breaks (DSBs), interstrand

crosslinks (ICLs) and the recovery of stalled or broken replication forks (Li and Heyer, 2008). DSBs are among the most deleterious lesions that occur in the genome since they affect both strands of the DNA helix. This is reflected in the fact that HR is complemented by two alternative repair pathways, namely non-homologous end joining (NHEJ) and translesion synthesis (TLS), for DSB repair (Dudas and Chovanec, 2004). Both exogenous agents, such as ionising radiation, and endogenous agents, including free radicals generated during aerobic metabolism, can lead to the generation of DSBs (Shrivastav et al., 2008).

The HR pathway has been conserved throughout evolution and can be divided into three distinguishable stages: pre-synapsis, synapsis and post-synapasis, as summarised in figure 1.12. During the pre-synapsis stage, the ends of the DSB are resected to produce 3' ssDNA tails, which are subsequently bound by a RecA-type DNA recombinase during synapsis. Nucleation of the recombinase along the ssDNA tail generates a nucleoprotein filament that searches a donor (homologous chromosome or sister chromatid) for a region of sequence homology, prior to strand invasion to create a joint molecule known as a displacement loop (D-loop). During post-synapsis, the D-loop is processed to regenerate the original duplex free from errors. The mechanism by which D-loop processing occurs varies among the three domains of life.

Bacteria | Eukaryotes | Archaea (table within figure)

	Bacteria	**Eukaryotes**	**Archaea**
End resection / Presynapsis	RecBCD	MRN/ MRX complex	MR complex
Strand invasion DNA synthesis / Synapsis	RecA	Rad51	RadA
Second end capture Branch migration Gap synthesis / Post-synapsis	Branch migration: UvrD, RuvAB, RecG	Branch migration: RAD54, BLM	Branch migration: Hel308?, HerA?
HJ resolution	RuvABC	Gen1/ Yen1	Hjc + PCNA

Figure 1.12: Homologous recombination/ double strand break repair in the three domains of life

Summary of the principal proteins involved in each stage of the double strand break repair pathway of homologous recombination in the three domains of life.

1.2.2.1 BACTERIAL HOMOLOGOUS RECOMBINATION

In bacteria, recombination-dependent repair of DSBs is initiated by the RecBCD/ AddAB family of enzymes. The RecBCD complex is a 330 kDa assembly comprising 3'-5' helicase and multifunctional nuclease (RecB), Chi (χ)-recognition (RecC) and 5'-3' helicase (RecD) activities to unwind and degrade linear dsDNA (Taylor and Smith, 1980; Muskavitch and Linn, 1982; Singleton et al., 2004; Dillingham and Kowalczykowski, 2008). The complex binds blunt DNA ends and the RecC subunit physically splits the complementary strands of the duplex apart via a protruding 'pin' region (figure 1.13). The displaced 3' and 5' tails of the duplex are fed to the RecB and RecD subunit, which act as ssDNA motors to pull the complementary strands of the duplex through RecC. The 3' tail eventually docks within the active site of the nuclease domain of RecB, where it is digested until the RecC domain encounters a recombination hotspot, otherwise called a Chi (χ)-site. At this point, RecC tightly grips the 3' tail enabling the 5' strand, emerging from RecD, to access the nuclease

active site. As the enzyme moves along the DNA, the tightly bound 3' strand forms a ssDNA loop that is subsequently bound by the single-stranded DNA binding protein (SSB) (figure 1.13) (Singleton et al., 2004).
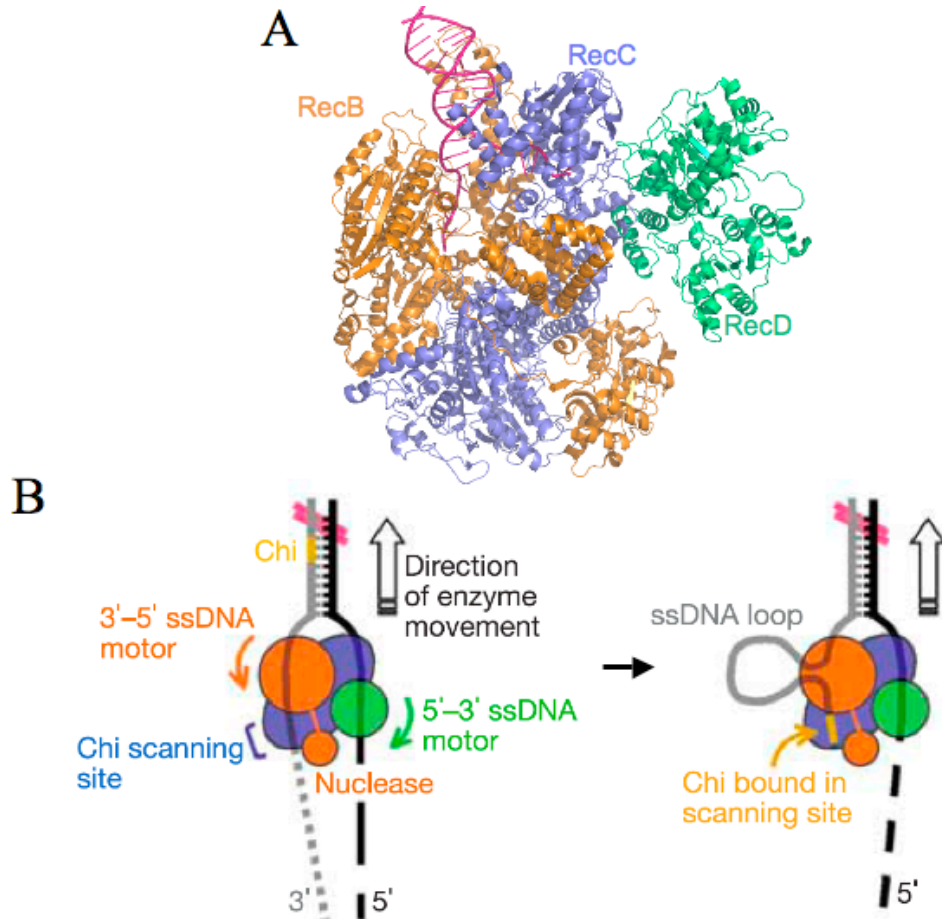
Figure 1.13: Double strand break processing by RecBCD

(A) Structure of RecBCD: DNA complex. (B) dsDNA is bound by RecC (blue circle) and split apart by a protruding pin, causing the displaced 3' and 5' strands to bind the RecB (large orange circle, helicase domain; small orange circle, exonuclease domain) and RecD (green circle) subunits, respectively. As the enzyme moves along the DNA the 3' tail is digested, as indicated by the orange arrow. Upon encountering a Chi-site, however, RecC tightly grips the 3' tail and the 5' strand is digested (indicated by the green arrow). As the enzyme continues to move along DNA, the 3' tail forms a ssDNA loop that is bound by SSB (and later, RecA). Adapted from (Singleton et al., 2004).

A RecA-loading heterotrimeric RecFOR complex is employed to create a nucleation site for RecA to bind. Once bound, RecA polymerises on ssDNA 3' tails, one molecule at a time, displacing SSB at a rate comparable to filament extension (Joo et al., 2006). The resulting nucleoprotein filament (figure 1.14) invades a homologous duplex (homologous chromosome or sister chromatid) to form a joint molecule. The

synapsis stage of bacterial HR is complete when the strand displaced from the donor pairs with the second ssDNA tail of the linear dsDNA, forming a double Holliday junction (Kowalczykowski et al., 1994).

Post-synapsis HR begins with branch migration or extension of the two Holliday junctions to increase the region of heteroduplex. In bacteria, the heterodimeric complex of RuvA and DNA helicase RuvB performs highly processive, unidirectional branch migration at a rate of approximately 20 bp/ s (Dawid et al., 2004). Alternatively, RecA and/ or the DNA helicases UvrD or RecG may catalyse this process (Ponticelli et al., 1985; Kodadek and Alberts, 1987; Kodadek, 1991; Tsaneva et al., 1992; West and Connolly, 1992; Morel et al., 1993). RuvAB subsequently recruits the junction-specific endonuclease, RuvC, which cleaves the Holliday junction at two sites; cleavage can occur in a symmetric (cleavage of both junctions in opposite orientations) or asymmetric (cleavage of paired junctions in the same orientation) manner to produce spliced/ crossover or patched/ non-crossover progeny, respectively (Connolly and West, 1990; Connolly et al., 1991). Spliced molecules possess exchanged flanking markers and contain heteroduplex DNA, while patched molecules have simply swapped ssDNA strands to produce heteroduplex DNA. HR is completed upon ligation of generated fragments by DNA ligase (Kowalczykowski et al., 1994; Sung and Klein, 2006).

1.2.2.2   EUKARYOTIC HOMOLOGOUS RECOMBINATION

In eukaryotes, processing of DSB processing to create 3' ssDNA tails relies on the cooperative activities of MRE11, RAD50 and NBS1 (or Mre11/Rad50/Xrs2 in yeast), collectively referred to as the MRN (or MRX in yeast) complex (Shrivastav et al., 2008). MRE11 is crucial for formation of the trimeric MRN complex and represents the catalytic subunit, possessing four N-terminal phosphoesterase motifs responsible for 3'-5' exonuclease and ssDNA endonuclease functions (Hopkins and Paull, 2008). RAD50, a member of the structure maintenance of chromosomes (SMC) protein family, encodes Walker A and B motifs that endow the protein with ATP binding and hydrolysis capabilities. In addition, RAD50 possesses chromatin-remodelling and adenylate kinase activities, the latter suggesting probable roles in DNA tethering (Bhaskara et al., 2007; Hopkins and Paull, 2008). Unlike bacterial HR, resection of DSBs in eukaryotes does not rely on the presence of χ-sequences, and instead DSBs are likely detected by an alteration to chromatin structure due to the combined action of

RAD50 and MRE11 (Symington, 2002). The helicase and nuclease functions of the MRN complex are presumed to be regulated by the NBS1 module, which also plays a role in nuclear localisation following DNA damage, signal transduction and catalytic activation (Dudas and Chovanec, 2004). Specifically, the NBS1 subunit binds to the ATM checkpoint kinase that ultimately leads to cell cycle arrest (Williams et al., 2008). DSBs may also be resected by the combined activities of DNA helicases, such as the RecQ family human BLM (yeast Sgs1) helicase, and the 5'-3' nuclease activity of EXO1 or DNA2.

The bacterial homologue of SSB, RPA, binds to the 3' tails generated by the MRN complex to protect them from nucleolytic degradation and to prevent the formation of secondary structure. However, RPA binding also impedes nucleation of ssDNA by RAD51 (homologue of bacterial RecA; 'RAD51', mammalian; 'Rad51', yeast). A number of RAD51 paralogues, collectively referred to as 'mediator proteins' are encoded by eukaryotic cells to facilitate RPA displacement (Morimatsu and Kowalczykowski, 2003). Budding yeast have two such proteins, namely Rad55 and Rad57, while mammalian cells encode seven RAD51-like genes (RAD51 A/B/C/D, XRCC2/3 and DMC1) to fulfil this function (Heyer et al., 2006; Sung and Klein, 2006). Although the exact mechanism of these mediator proteins is not known, they are all required for RAD51 filament formation (Li and Heyer, 2008) (figure 1.14). Further regulation of RAD51, in terms of its nucleation along ssDNA and its subsequent catalytic activity, is provided by members of the RAD52 epistasis group, which, in budding yeast, includes Rad50, Rad51, Rad52, Rad54, Rdh54, Rad55, Rad57, Rad59, Mre11, and Xrs2 genes (Robertson et al., 2009). Rad52 is a pro-recombinogenic protein, implicated in Rad51 recruitment to damage foci and RPA displacement (Plate et al., 2008). Rad52 forms a multimeric, ring-shaped structure that binds ssDNA on the outer rim of each subunit. The structural architecture of the protein promotes the cooperative binding of ssDNA such that it competitively displaces RPA from DNA and facilitates Rad51 nucleation via a direct interaction (Plate et al., 2008). Rad52 has also been shown to anneal homologous ssDNA strands coated with RPA (Sugiyama et al., 1998), which may implicate it during second end capture (Sugiyama et al., 2006). The Swi2/Snf family member, Rad54, has been described as the Swiss-army knife of eukaryotic HR since it modulates the pathway at various stages (Heyer et al., 2006). Rad54, an ATP-dependent dsDNA translocase, is structurally reminiscent of SF2 DNA helicases and thus may employ a helicase-like mechanism to translocate along either

strand of a duplex. The protein is highly processive, moving along stretches of DNA approaching 15000 bp without dissociation, at a rate of ~ 301 bp/ s (Thoma et al., 2005), which has been proposed to enhance the ability of Rad51 nucleoprotein filaments to invade regions of homology (Heyer et al., 2006). Rad54 also stimulates Rad51 dissociation from DNA; the absence of a bacterial homologue of Rad54 may reflect the higher DNA-dependent ATPase activity, and therefore turnover, of RecA (Heyer et al., 2006). Excessive homologous recombination can prove deleterious to the cell and therefore a number of proteins have been found to display anti-recombinogenic activities, such as the DNA helicase Srs2 from yeast that acts to disrupt Rad51 nucleoprotein filaments (Veaute et al., 2005).

The post-synaptic stage of eukaryotic HR can be subdivided into three pathways: double-strand break repair (DSBR), synthesis-dependent strand annealing (SDSA) and break-induced repair (BIR). DSBR is the classical post-synaptic process (figure 1.12) whereby the second 3' ssDNA tail of the linear dsDNA is captured by the donor molecule, either via simple strand annealing or a second invasion step, to prime a second round of DNA synthesis. This creates an intermediate that can be asymmetrically cleaved by the heterodimer MUS81-EME1 (Mus81-Mms4 in yeast) to yield crossover products, or can progress via branch migration to produce a double Holliday junction that can be resolved by junction-specific resolvases (Mimitou and Symington, 2009). The recombination mediator protein, RAD54 (Rad54 in yeast), is thought to play a role in branch migration and is stimulated by RAD51, an activity that is thought to drive the passage of joint molecules through the DSBR, as opposed to the SDSA, pathway (Haseltine and Kowalczykowski, 2009). Recently, a Rad2/ XPG family nuclease, called GEN1 (Yen1 in yeast) was identified in humans (Ip et al., 2008). GEN1 is a structure-specific nuclease that cleaves Holliday junctions symmetrically, analogous to that described for bacterial RuvC, to produce crossover and non-crossover products with equal efficiency (Ip et al., 2008). Holliday junctions can be alternatively processed by dissolution by the RecQ DNA helicase BLM (Sgs1 in yeast), in combination with topoisomerase IIIα and RMI1, to produce non-crossover products (Mimitou and Symington, 2009). Gaps in the DNA strands are repaired and sealed by DNA polymerase and DNA ligase to complete the DSBR process.
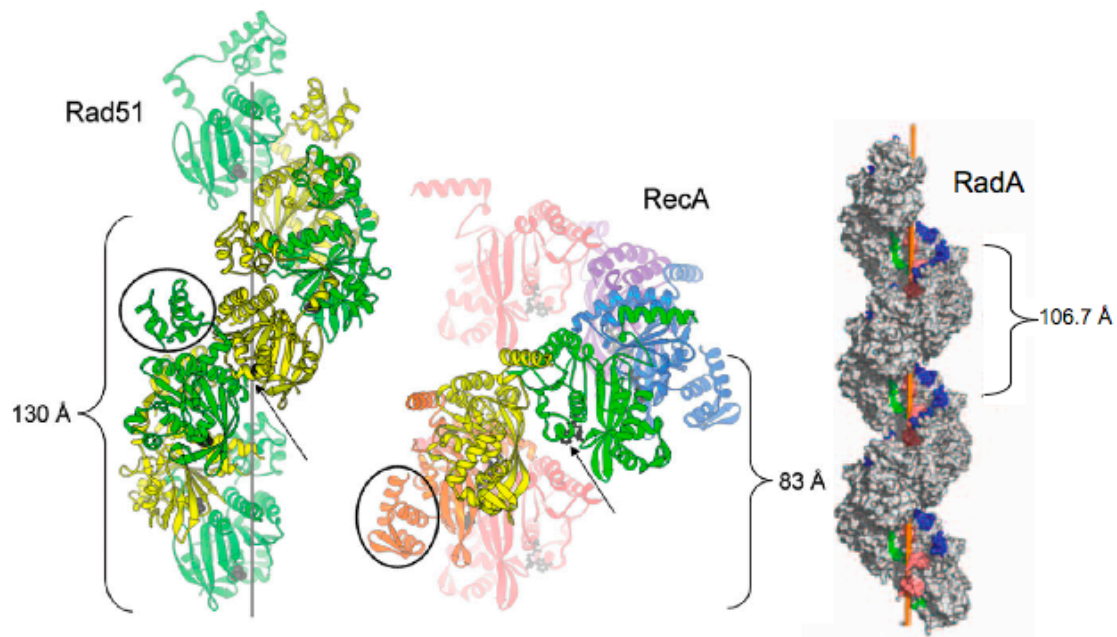
Figure 1.14: Nucleoprotein filaments of Rad51, RecA and RadA

The extended, right-handed nucleoprotein filaments of Rad51, RecA and RadA are shown with helical pitches of 130 Å, 83 Å and 106.7 Å, respectively. In the Rad51 nd RecA structure, the circle regions correspond to the N-terminal and C-terminal domains of the protein, respectively, while the black area highlights the position of ATP/ ADP binding. Modified from (Conway et al., 2004; Chen et al., 2007a).

1.2.2.3   ARCHAEAL HOMOLOGOUS RECOMBINATION

The repair of DSBs in the archaea is expected to occur upon damage detection by the SSB protein, which serves to recruit homologues of the eukaryotic Mre11 and Rad50 proteins (forming the MR complex). The presence of such homologues renders it likely that DSB processing in the archaea proceeds as described for eukaryotes, whereby the endonucleases activity of Mre11 disrupts dsDNA ends at the break site and Rad50 serves to recruit downstream processing proteins. The *mre11* and *rad50* genes are located on the same operon as *herA*, an ATP-dependent, dsDNA helicase, and *nurA, a* 5'-3' exonuclease (Constantinesco et al., 2004), the gene products of which have been shown to functionally interact with the MR complex and RadA (RecA family protein) to process DSBs and promote D-loop formation (Hopkins and Paull, 2008; Quaiser et al., 2008) (figure 1.15). HerA and NurA homologues are absent from the eukarya, but likely represent the functional counterparts of DNA helicase Sgs1, and exonucleases Exo1 and Dna2, respectively (Hopkins and Paull, 2008).
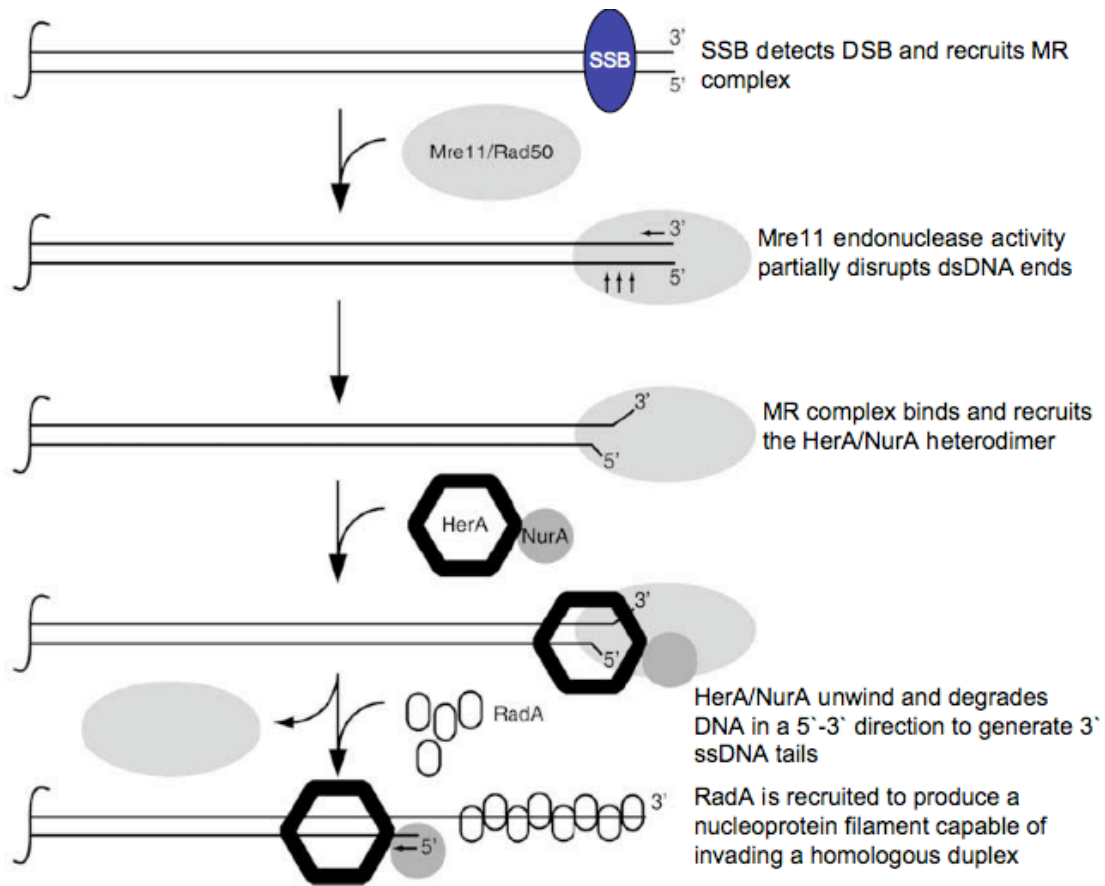
Figure 1.15: DSB processing in the archaea

Detection of a DSB may be initially performed by SSB, which binds to Rad50 to recruit the MR complex. The MR complex subsequently processes the dsDNA ends of the break via the endonuclease activity of Mre11, which permits binding of the HerA/ NurA heterodimer. HerA unwinds the duplex to produce a 5' tail that is subsequently digested by NurA. RadA polymerises along the 3' tail to produce a nucleoprotein filament capable of invading a homologous duplex. Modified from (Hopkins and Paull, 2008).

Similar to eukaryotes, archaea encode a number of RadA paralogues (at least one, and up to four) that are presumed to act as pro- or anti-recombinogenic factors, regulating the nucleation of RadA along ssDNA. As described above, eukaryotic Rad54 was identified as a key player in promoting the Rad51-mediated strand exchange reaction. Crenarchaeal *S. solfataricus* encodes a functional homologue of Rad54, referred to as SsoRad54. This protein functions as a dsDNA-dependent ATPase that physically interacts with RadA to stimulate strand exchange (Haseltine and Kowalczykowski, 2009). RadB, a RadA paralogue confined to euryarchaea, has been shown to negatively influence the strand-exchange activities of RadA, via a direct interaction, by interfering with nucleoprotein filament formation (Komori et al., 2000a). A group of RadA paralogues, structurally and functionally distinct from RadA and

RadB, have been identified in several crenarchaeal genomes and are collectively referred to as the aRadC family (Haldenby et al., 2009). *S. solfataricus* encodes three such proteins, namely Sso0777, Sso1861, and Sso2452, the latter of which inhibits the recombinase activity of RadA (McRobbie et al., 2009). In this light, RadB and Sso2452 (and perhaps other aRadC family proteins) play a similar role as the eukaryotic anti-recombinogenic DNA helicases Srs2 (budding yeast) and BLM (human) (discussed further in chapter 7).

Following D-loop formation, the second 3' ssDNA tail is captured to form a Holliday junction that subsequently undergoes branch migration. In bacteria, this activity is primarily catalysed by the RuvAB complex, while the RecQ DNA helicases, including BLM, are implicated in eukaryotic branch migration. While homologues of these proteins do not exist in the archaea, several archaeal helicases have been shown to unwind or regress Holliday junctions, including HerA (Zhang et al., 2008) and Hel308 (or Hjm) (Li et al., 2008). Hel308 is a SF2 DNA helicase (Chapter 5) with homology to proteins of the Mus308 family, including human and *Drosophila* Hel308, human Pol Θ and *Drosophila* Mus308. A further potential candidate for migration of Holliday junctions is *S. solfataricus* Hel112, an ATP-dependent 3'-5' DNA helicase that exhibits similar substrate specificities as the eukaryotic RecQ-helicases, including branched or forked DNA (De Felice et al., 2007).

The archaeal HR pathway is completed following the symmetrical cleavage of Holliday junctions by the structure-specific endonuclease, Hjc (Komori et al., 1999; Komori et al., 2000a; Komori et al., 2000b), which is conserved in all archaea and exhibits structural and functional similarities to bacterial RuvC (figure 1.16). Hjc has a typical α/ β structure and DNA cleavage relies on Hjc dimerisation and its interaction with the minor groove (via two basic residues on a loop between β1 and β2) and major groove (via two basic residues between β3 and β4) of DNA (figure 1.15). The physical interaction of Hjc with PCNA stimulates the cleavage reaction (Dorazi et al., 2006) and is further regulated by RadA and RadB in an ATP-dependent manner (Komori et al., 2000a). As for bacterial and eukaryotic recombination, the final step of archaeal HR involves the ligation of newly synthesised segments of DNA into the original strand by DNA ligase.
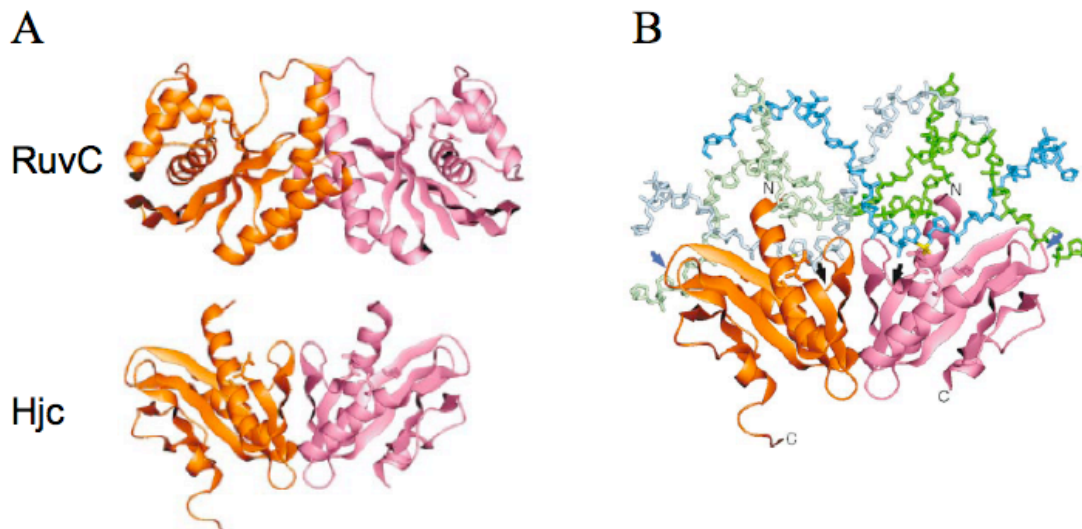
Figure 1.16: Structural comparison of bacterial RuvC and archaeal Hjc

(A) Structure of bacterial RuvC and archaeal Hjc (both dimers). (B) A Holliday junction (each strand coloured differently) has been modelled onto the Hjc structure. Black and blue arrows indicate the regions of Hjc that interact with the minor (β1 and β2) and major groove (β3 and β4) of DNA, respectively. Adapted from (Nishino et al., 2001).

1.2.2.4   STALLED REPLICATION FORK RESTART

Replication forks can stall or collapse upon meeting DNA 'road-blocks', such as bulky secondary structures, DNA-bound proteins, and DNA lesions (Ward, 2007). To maintain genome integrity, forks must be repaired and replication restarted at sites remote from replication origins.

Fork stalling typically leads to nascent strand regression (Marians, 2004), which involves fork reversal to generate a 4-way 'chicken foot' structure that permits DNA synthesis past the lesion by template switching. The chicken foot is subsequently regressed to recover the fork and the lesion is repaired at a later stage (Singleton et al., 2001). Fork regression appears to be an activity conferred by DNA helicases, including RecG in bacteria, the RecQ family of helicases in eukaryotes (including BLM and WRN helicases) and the Hjm/ Hel308 helicase in archaea (Marians, 2004). Bacterial RecG comprises three structural domains: an N-terminal domain involved in DNA binding (domain 1) and two C-terminal RecA-like motor domains (domains 2 and 3) that encode the canonical motifs representative of SF2 helicases (Singleton et al., 2001). Domain 1 interacts at the replication fork junction such that one side of the junction (e.g. the lagging arm) interacts with a 'wedge' domain and the other (e.g. the leading arm) interacts with a β-hairpin structure (figure 1.17). As the protein moves along the

29

fork in an ATP-dependent manner, the 'wedge' domain and β-hairpin serve to split apart the duplex arms, leaving the template strands to pass along narrow grooves either side of the 'wedge' domain (figure 1.17 A). The displaced strands of the leading and lagging duplexes then rebind to form the fourth arm of the 'chicken foot' (figure 1.17 B and C). Once DNA synthesis past the lesion has taken place, the 'chicken foot' can be reversed to reform the replication fork prior to the assembly of the replisome to restart replication (Marians, 2004).
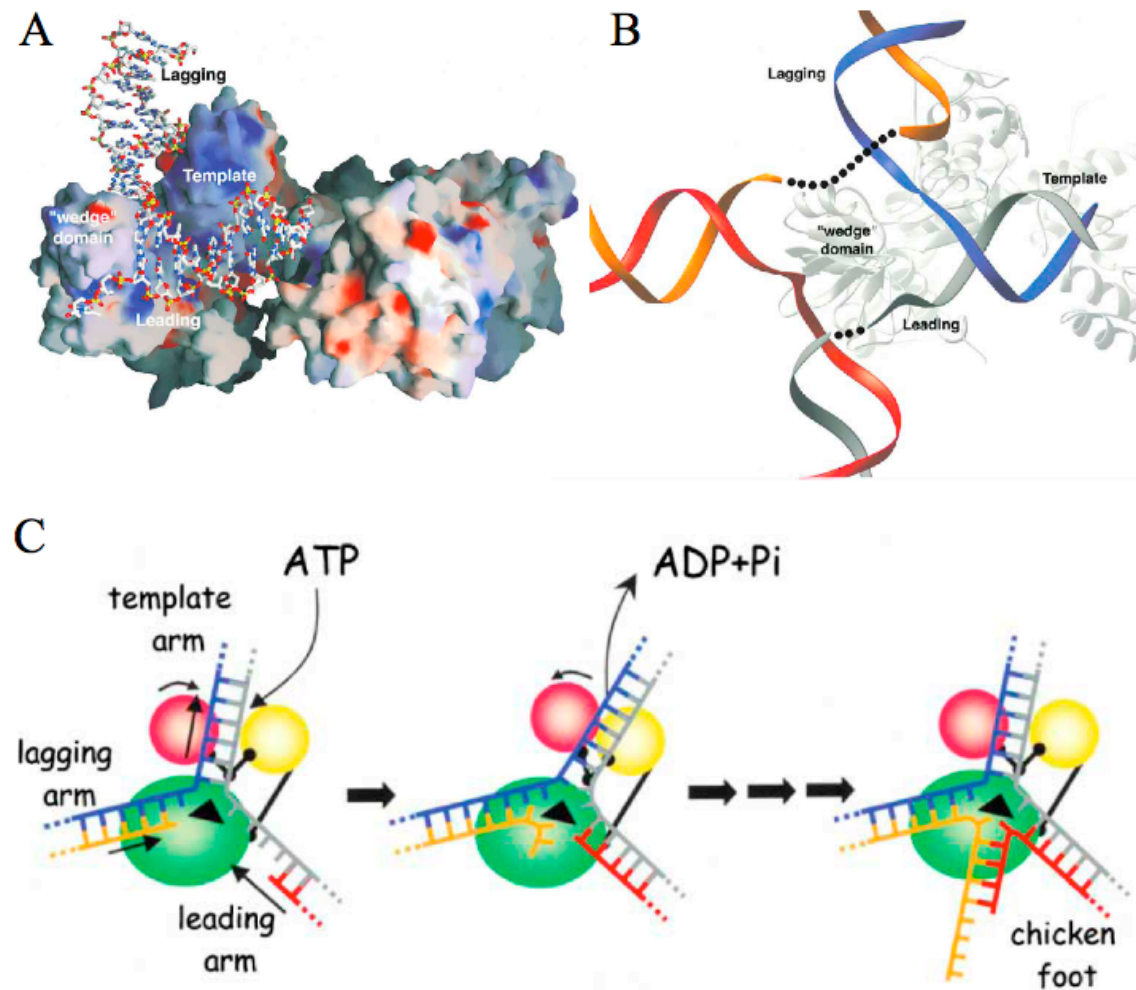
Figure 1.17: Fork reversal by bacterial RecG

(A) Surface representation of RecG bound to a replication fork DNA structure. Positive and negative surface potential is coloured blue and red, respectively. The structure highlights how the fork is split across the wedge domain. (B) Model depicting Holliday junction formation during catalysis by RecG. (C) RecG binds the replication fork at the junction. Upon ATP binding, domains 2 (yellow) and 3 (red) undergo conformational changes that pull the template arm such that lagging and leading arms are dragged across and, thus separated, by the 'wedge' domain (black triangle). ATP hydrolysis allows domains 2 and 3 to return to their original positions. The displaced nascent strands associate to form a chicken foot structure. Adapted from (Singleton et al., 2001).

Fork-blocking lesions can also be repaired, rather than bypassed, since regression generates a substrate for excision repair by re-positioning the site of damage opposite the undamaged lagging strand (Marians, 2004). Alternatively, the Holliday junction can be cleaved by a junction-specific resolvase (such as bacterial RuvC, archaeal Hjc, or eukaryotic GEN1), which is subsequently processed via a D-loop substrate to promote replication restart (figure 1.18).
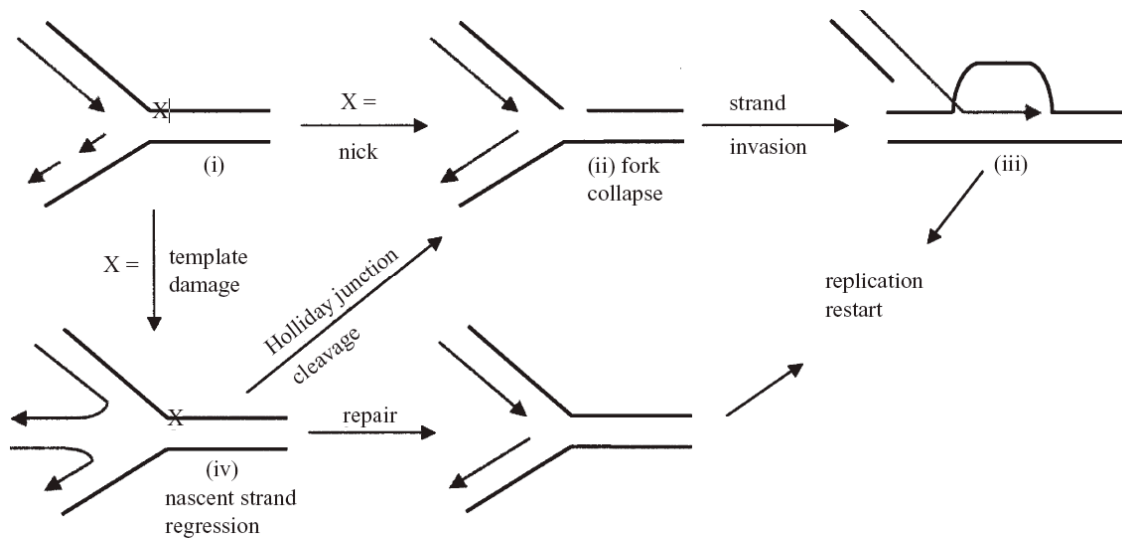
Figure 1.18: Repair of stalled or collapsed replication forks

Replication forks stall or collapse in the presence of template damage and strand-specific nicks, respectively. Stalled forks undergo fork regression (iv) to produce a Holliday junction. This regenerates the template opposite the lesion to permit excision repair. Reversal of the regression restores the fork. Alternatively, the Holliday junction can be cleaved by a junction-specific resolvase (ii) as occurs upon fork collapse. Fork restoration requires a strand-invasion reaction to produce a D loop (iii), which is a substrate for replication fork restart (Marians, 2004).

Particular lesions including single-strand breaks, cause replication forks to collapse and must be repaired by the HR pathway (McGlynn and Lloyd, 2002), a process known as recombination-dependent replication. In bacteria, this involves the combined helicase and nuclease activities of the RecBCD complex with the strand-exchange RecA enzyme to produce a D-loop (Marians, 2004). This joint molecule can be subsequently processed by either branch migration, catalysed by RecG or RuvAB, to eventually produce a Holliday junction that can be resolved by junction-specific nucleases, or can be used as a substrate to prime replication restart (Marians, 2004). In bacteria, it is thought that PriA, a DNA helicase known to recognise D-loop structures, binds to the joint molecule and triggers replisome assembly (Sandler and Marians, 2000).

1.3   DNA HELICASES

Helicases are molecular motors that couple the energy of nucleoside triphosphate (NTP) hydrolysis to the separation of nucleic acid (NA) duplexes. This reaction is central to all DNA metabolism, including DNA replication, transcription and translation, RNA processing, DNA recombination and repair. Although there are exceptions to the rule, the majority of DNA helicases bind to ssDNA or ss/ dsDNA junctions and translocate unidirectionally (either 3'-5' or 5'-3'), along the DNA to disrupt hydrogen bonds between paired strands (Singleton and Wigley, 2002) (figure 1.19).

Helicases belong to one of six superfamilies (SF1-6), determined largely on the inclusion of up to 11 conserved 'helicase motifs' (figure 1.19). These motifs are confined to the 'core domains' that comprise 2 RecA-like folds responsible for coupling the energy of NTP binding and hydrolysis to conformational changes that, in turn, drive DNA translocation and unwinding (Patel and Picha, 2000). In all cases, motifs I (Walker A motif), II (Walker B motif) and VI (contains residues that form an 'arginine finger') are absolutely conserved and are involved in energy coupling (Singleton et al., 2007). The majority of helicases belong to SF1 and SF2 and are typically monomeric, while those of SF3-6 form hexameric (or double-hexameric) rings (Singleton et al., 2007). The regions between the helicase motifs are considered highly divergent in sequence but conserved in length, and are thought to bestow functional characteristics specific to each individual protein (Tuteja and Tuteja, 2004). Although referred to as 'helicase motifs', it has become increasingly clear that the presence of these motifs does not necessarily dictate that a protein will unwind DNA; rather, these motifs are representative of NTP-dependent translocases, i.e. proteins that hydrolyse NTP to move along NA (Singleton and Wigley, 2002).
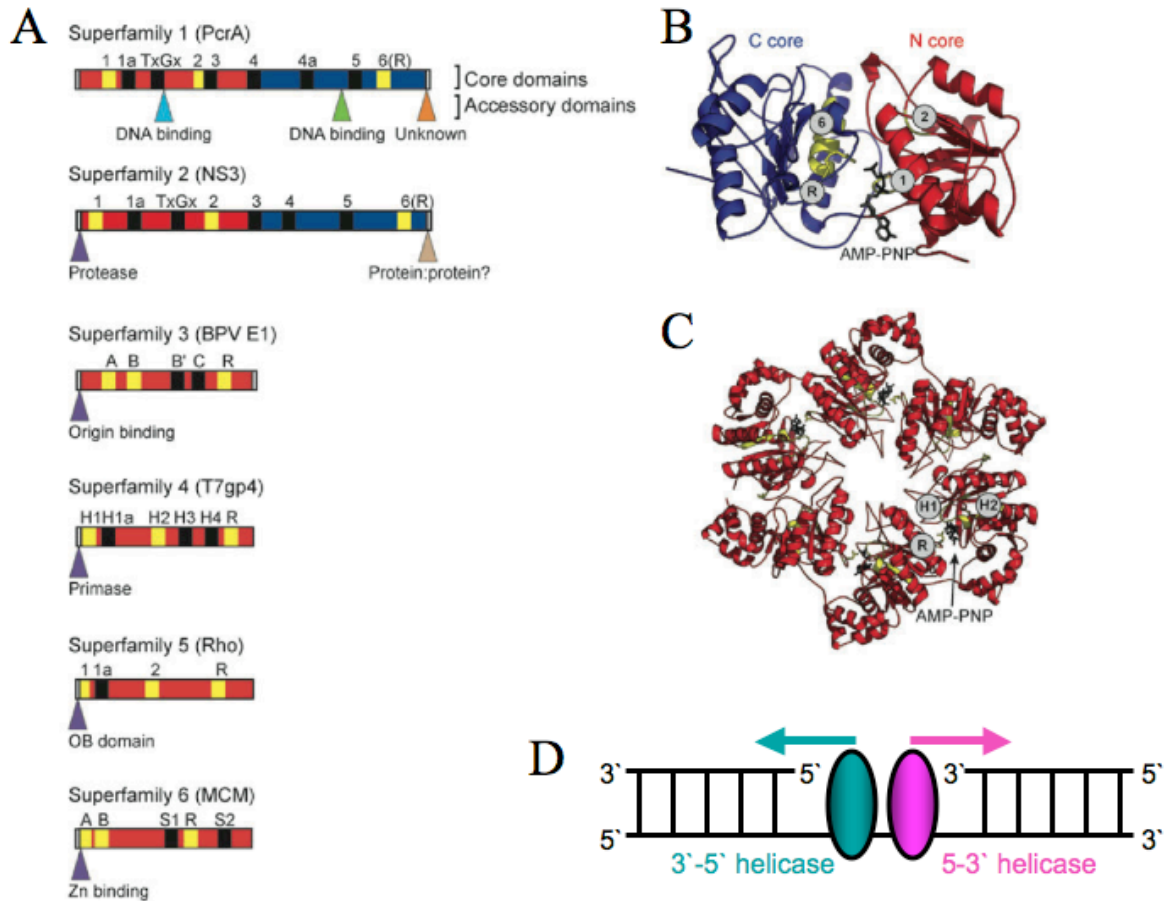
Figure 1.19: Classification of helicases

(A) Helicases are divided between six superfamilies. The 'core domains', the helicase motifs (universally conserved motifs, yellow; family-specific motifs, black) and the positions and functions of accessory domains (specific to the protein shown) are highlighted. Representative core structures of the SF1-2 (B; the core domains of PcrA helicase from SF1 is shown) and SF3-6 (C; T7 gene 4 protein from SF4 is shown) are shown. Universal structural elements involved in NTP binding and hydrolysis are shown in yellow. (B) The SF1 and SF2 helicases comprise a monomeric core formed from two RecA-like folds (N core and C core). The N core contains motifs 1 and 2 (as indicated), while the C core contains the arginine (R) finger of motif 6. AMP-PNP, an NTP analog (black), binds the protein at the interface between the N and C core. (C) SF3-6 helicases comprise a ring-shaped core of six RecA-like folds (red). NTP (black) binds at the interface of each domain. Motifs H1 and H2, containing the Walker A and B boxes respectively, sit on the opposite side of the cleft to the arginine (R) finger. (D) Helicase directionality is defined by the strand upon which the helicase (coloured oval) translocates. Modified from (Singleton et al., 2007).

1.3.1 MECHANISM OF HELICASE ACTION

Helicases unwind DNA or RNA duplexes passively, resulting from thermal fraying at the ss-/ds-DNA junction, or actively whereby the protein mechanically separates the NA duplex using the energy from ATP binding and hydrolysis. Active mechanisms of NA unwinding by helicases are described by two principal models, the

Inchworm Model and the Active Rolling Model, with a third model proposed solely for hexameric helicases. The Inchworm and Active Rolling models differ with regards to the oligomeric state of the protein, the mode of NA binding, and the effect of NTP binding on NA affinity (Patel and Picha, 2000).

1.3.1.1 THE INCHWORM MODEL

The inchworm model describes the mechanism of strand displacement by monomeric helicases and requires two DNA binding sites per monomer (Lohman and Bjornson, 1996). The crystal structures of PcrA in complex with DNA and either a non-hydrolysable analog of ATP (representing the 'substrate' configuration) or a sulphate ion (representing the 'product' configuration) have been used to proposal a general inchworm mechanism (figure 1.20). The process can be subdivided into DNA translocation and duplex destabilisation, the coupling of which results in helicase activity. Prior to ATP binding, PcrA binds the single-stranded tail of DNA across its 'core domains' (domains 1A and 2A). Binding of ATP at the cleft between the 'core domains' produces conformational changes that result in cleft closure and a weakening of the interaction between ssDNA and domain 1A. At the same time, domain 2A tightens its grip on the ssDNA tail such that translocation across the surface of the 'core domains' occurs. These conformational changes also alter the surface of domains 1B and 2B such that its affinity for duplex DNA is increased. Domains 1B and 2B subsequently bind the duplex adjacent to the ss-/ ds-DNA junction, causing the duplex to bend and become destabilised as bases are flipped out of helical alignment. This has been proposed to resemble a 'Mexican wave' (figure 1.19). Hydrolysis of ATP re-opens the ATP-binding cleft and domain 1A tightens its grip on the ssDNA tail, relative to domain 2A, allowing the protein to slide along the DNA towards the ss-/ds-DNA junction in preparation for the next cycle of ATP-dependent DNA translocation and unwinding. The mechanism of DNA unwinding by PcrA is revisited in chapter 6.
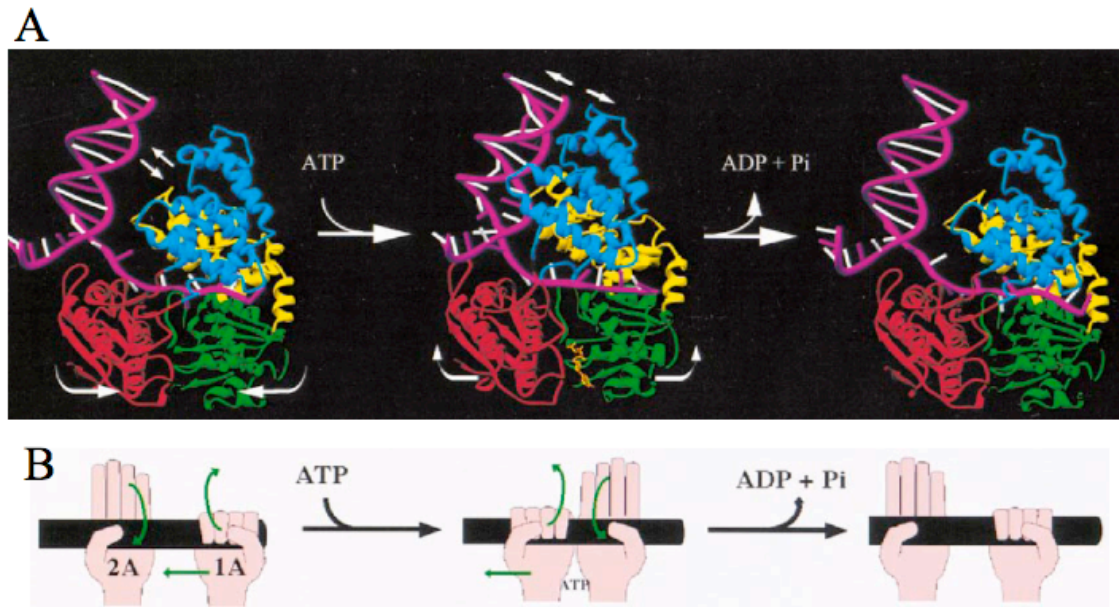
Figure 1.20: Inchworm model of DNA unwinding by PcrA

(A) Prior to ATP binding, PcrA binds the ssDNA tail. ATP binding at the interface between the core domains leads to a conformational change such that domains 1B and 2B bind the duplex and unwind several base pairs at the junction. ATP hydrolysis allows the protein to return to its original conformation, the duplex is released and the protein translocates one base pair along the ssDNA tail towards the junction. (B) Cartoon representing the relative affinity of domains 1A and 2A (the 'core domains') for ssDNA during DNA translocation. An open or closed hand represents a relatively loose or tight grip on DNA, respectively. Modified from (Velankar et al., 1999).

The 'cooperative inchworm model' represents a refinement of the general inchworm model to explain the cooperative activity of multiple bacteriophage T4 Dda helicase monomers during DNA unwinding (Byrd and Raney, 2006; Mackintosh and Raney, 2006). It was proposed that such cooperativity might enable proteins to gather sufficient momentum to overcome the inhibitory effect of a NA 'roadblock', such as bound protein. Indeed, such NA 'roadblocks' may provide an explanation for the small step sizes exhibited by many proteins: the free energy released following the hydrolysis of one ATP molecule should be sufficient to unzip multiple base pairs; the fact that many helicases, including *E. coli* PcrA, unwind only one base pair per ATP hydrolysed makes it likely that the energy is used by the protein for other tasks, such as displacement of NA-bound proteins.

Finally, the 'quantum inchworm' model was proposed to explain the mechanism of DNA translocation and unwinding by the bacterial RecBC helicase (Bianco and Kowalczykowski, 2000). This protein binds dsDNA ends and is involved in the resection of DNA to initiate homologous recombination. Although the protein

required more than one ATP molecule to unwind one base pair of DNA, translocation occurred in step sizes of 23 nucleotides. To explain this mode of unwinding, the RecBC helicase was proposed to have two non-equivalent binding domains, a 'leading' and 'trailing' domain. Upon binding a dsDNA end, the leading domain swings forward to bind and anchor the protein 23 nucleotides from the duplex end. The trailing domain, considered as the helicase domain, binds and hydrolyses multiple ATP molecules to unwind DNA in small step sizes. As it advances towards the anchored leading domain, the trailing domain emits a signal causing the leading domain to dissociate and rebind the DNA a further 23 nucleotides from the duplex end. This cycle of protein expansion and contraction then continues until the duplex has been completely unwound (Bianco and Kowalczykowski, 2000).

1.3.1.2   THE ACTIVE ROLLING MODEL

The active rolling model requires an oligomeric helicase (at least dimeric), with each monomer possessing identical NA binding sites that can bind both ssDNA and dsDNA (Soultanas and Wigley, 2001). The model dictates that at least one subunit is bound to ssDNA ahead of the ss-/ds-DNA junction at any one time, while the second binding site can either bind to the same single-strand region or to dsDNA ahead of the fork. ATP binding and hydrolysis modulates the relative affinity of each subunit for NA such that each subunit binds ssDNA or dsDNA alternately. The alternate binding of individual subunits to DNA necessitates that the step size of translocation must equal (at least) the footprint of the individual binding sites of each monomer.

Biochemical and structural studies of the *E. coli* Rep DNA helicase support an active, rolling model for DNA translocation and unwinding that relies on the functional asymmetry and cooperativity of the two subunits of the Rep dimer. In an ATP-dependent manner, the two Rep subunits alternate their relative affinity for DNA such that DNA translocation proceeds by a subunit-switching, or 'hand-over-hand', mechanism in which multiple base pairs are unwound by each Rep subunit during a single binding event (figure 1.21 B) (Bjornson et al., 1996). Release of ssDNA from one subunit presumably allows it to rebind closer to the ss-/ ds-DNA junction. Structural data revealed that binding of ssDNA across the 'core domains' of one Rep subunit triggered a large 130 ° rotation of domain 2B, which is implicated in Rep dimerisation (Korolev et al., 1997). It follows that binding of DNA would be coupled to large

relative movements (of approximately 75 Å) of the two DNA binding sites (i.e. the two monomers) of the Rep dimer (Korolev et al., 1997), providing a structural explanation for the subunit-switching mechanism employed by the protein (figure 1.21 A). The DNA translocation and unwinding mechanism of the Rep helicase is further discussed in chapter 6.
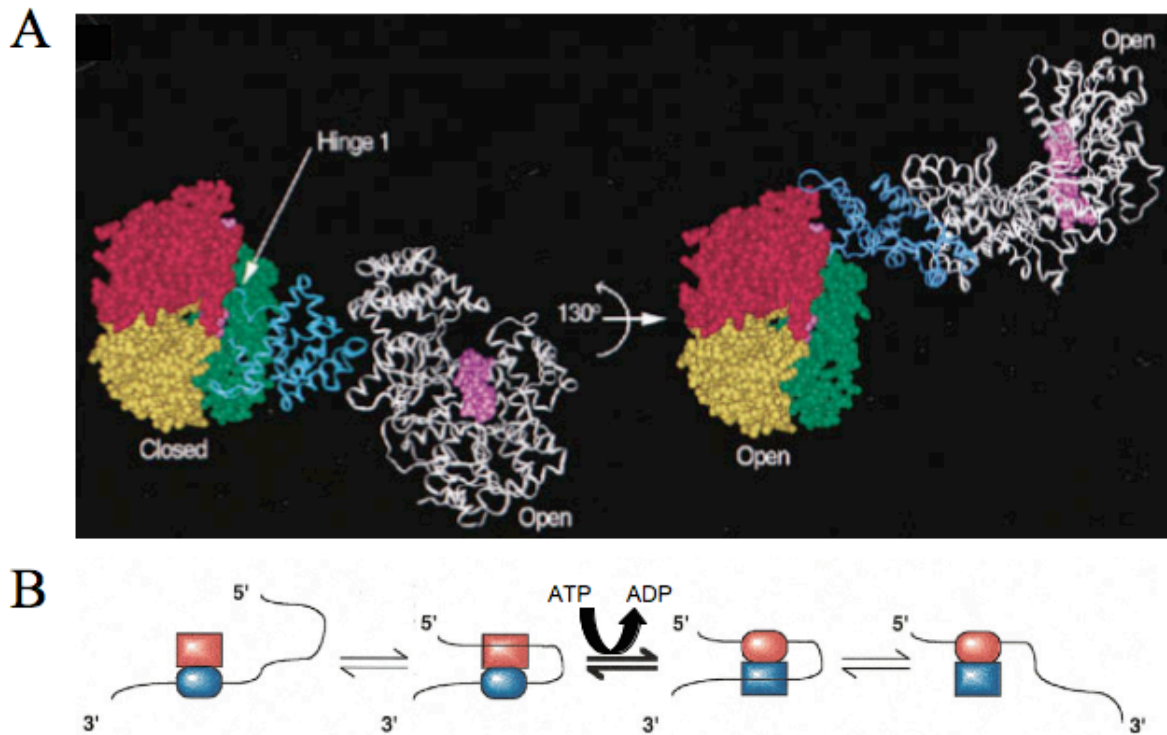
Figure 1.21: Active-rolling model of DNA unwinding by Rep

(A) Binding of DNA (pink space fill) to one Rep subunit (left) causes domain 2B (blue ribbon) to rotate by 130 º such that the second subunit (grey) would shift 75 Å relative to the first subunit. (B) ATP-dependent translocation of Rep by a subunit-switching or 'hand-over hand' mechanism. In an ATP-dependent manner, binding of ssDNA to the second Rep subunit (red) results in the release of the originally bound (blue) subunit via a transient intermediate whereby both subunits are bound to ssDNA. Modified from (Bjornson et al., 1996; Korolev et al., 1997).

1.3.1.3   HEXAMERIC HELICASES

All domains of life encode helicases that can assemble into hexameric, ring-shaped structures to carry out essential strand-displacement functions during DNA metabolism. The most common topological configuration for ring helicases was shown by the electron microscopy (EM)-based reconstruction of the bacteriophage T7 gene *4* helicase/ primase (T7gp4) protein (Egelman et al., 1995), which revealed the passage of a single-strand of a NA duplex through the central pore of the ring, whilst excluding the

complementary strand (figure 1.22). The T7 helicase loads the 5' strand within the central pore via a ring-opening mechanism, whereby initial docking of DNA to the outside of the protein triggers a conformational change, opening the central pore to allow entry of a single DNA strand before subsequent closure of the pore (Ahnert et al., 2000). To prevent reannealing, the T7 helicase binds the 3' strand to a site outside the central pore. Such an arrangement allows the protein to maintain a tight grip on the NA, facilitating unidirectional translocation and unwinding of the duplex with enhanced processivity (Patel and Picha, 2000).

The structure of the Papillomavirus SF3 E1 hexameric helicase shed light on the mechanism employed by ring-shaped helicases to translocate and unwind NA (Enemark and Joshua-Tor, 2006). As for the T7 helicase, one strand of DNA passes through the hexamer pore and NTP binds at the subunit interface (figure 1.22 B). Each subunit of the hexamer interacts with DNA bases via a protruding hairpin loop. For each round of ATP hydrolysis, the hairpin of each subunit remains in contact with the same base. The position of these DNA-binding hairpins along the vertical axis of the NA strand correlates to the nucleotide-bound state of each subunit, such that ATP binding raises the position of the loop to encourage binding of DNA close to the entrance of the pore, and subsequent hydrolysis to generate the ADP-bound and then the free state gradually lowers the position of the hairpin loop towards the pore exit. Therefore, during each cycle of ATP hydrolysis, the protein migrates with the oligonucleotide such that one nucleotide of ssDNA passes through the channel (figure 1.22 B-C). The cycle begins again following ATP binding to the first subunit, allowing engagement with the next available ssDNA nucleotide. The process visually resembles six hands pulling on a rope, in a hand-over-hand fashion (Enemark and Joshua-Tor, 2008).
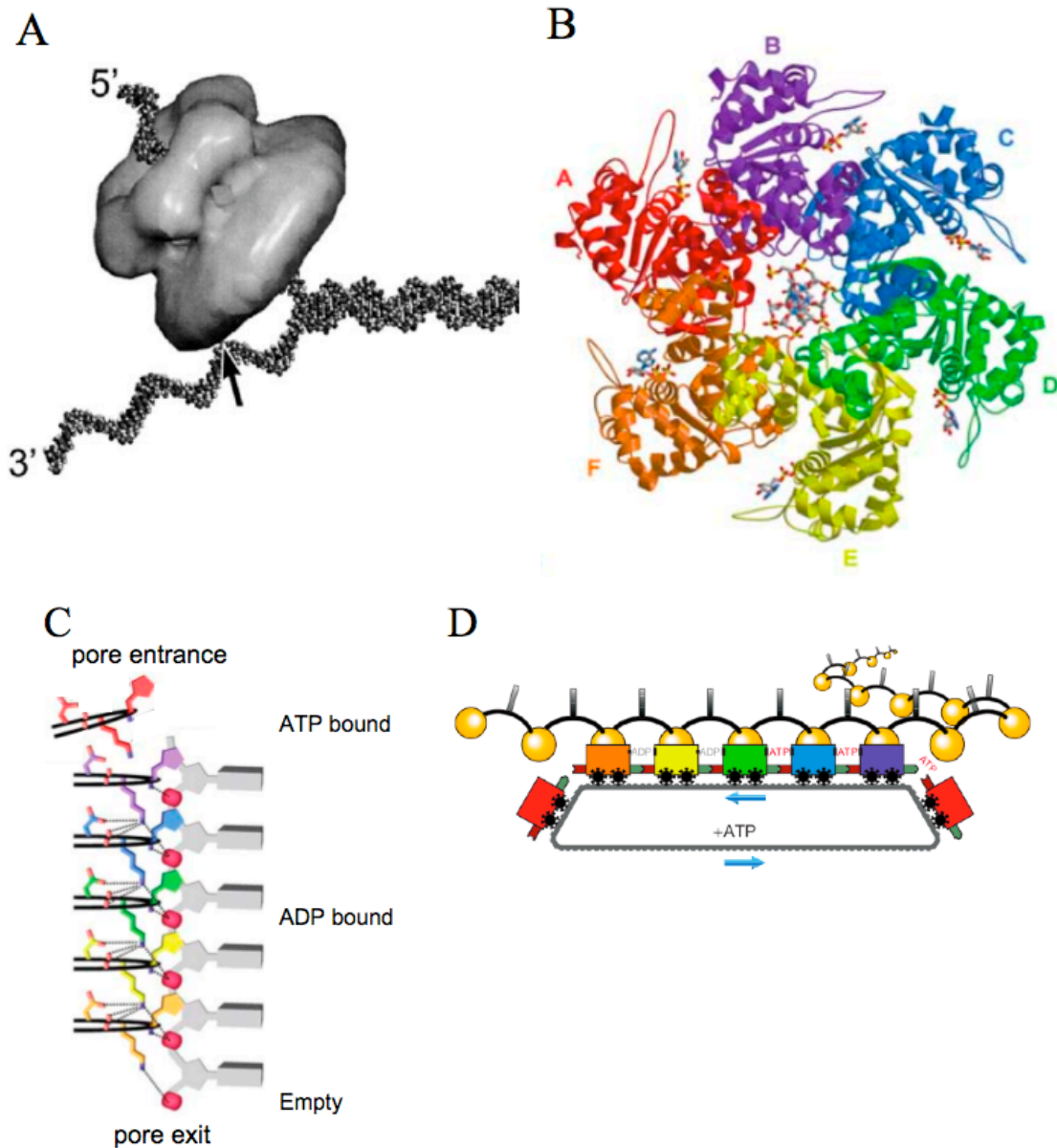
Figure 1.22: DNA translocation and unwinding by hexameric helicases

(A) Electron microscopy reconstruction of the T7 helicase from bacteriophage T7 in complex with DNA, revealing the passage of the 5' strand through the central pore of the ring and the exclusion of the 3' strand. The 3' strand docks onto a site on the outside of the protein (indicated by arrow). (B) Crystal structure of the hexameric E1 helicase from papillomavirus revealing a bound NA within the central pore and six NTP molecules at the interface between each subunit. (C) The DNA-binding hairpins (black loops) of the E1 helicase make contacts with the sugar-phosphate backbone of the 5' NA strand, the vertical position of which is correlated with the ATP-bound state of the protein. (D) The E1 helicase unwinds DNA by a 'coordinated escort' mechanism. Each subunit (wagon) binds one DNA nucleotide from the right (pore entrance) upon ATP binding and, upon ATP hydrolysis and subsequent release of ADP, carries the nucleotide to the left (pore exit). ATP binding repeats the process. Adapted from (Ahnert and Patel, 1997; Enemark and Joshua-Tor, 2006).

1.4   AIMS

This thesis aims to describe the characterisation of DNA repair proteins that are predicted to couple the energy from ATP binding and hydrolysis to DNA processing. The archaeal homologue of human XPD has been studied previously and this thesis aims to expand upon this work to understand the molecular basis of disease caused by mutation of the protein. This thesis also investigates the function of an XPD homologue, DinG, from the bacterium *Staphylococcus aureus*, a protein with a putative helicase and nuclease domain. The remaining three chapters of this thesis contribute to the current understanding of archaeal homologous recombination, firstly through the biochemical and biophysical characterisation of a DNA helicase, Hel308, and secondly by investigating the function of a RadA paralogues and putative DNA recombinase (Sso2452). Together, this thesis aims to contribute to the understanding of the diverse roles played by mechanochemical coupling proteins during DNA repair.

CHAPTER 2

MATERIALS AND METHODS

2.1 CLONING AND PROTEIN EXPRESSION

2.1.1 CLONING AND VECTORS

The pET28c-*xpd* vector, for expression of untagged, recombinant XPD from *Sulfolobus acidocaldarius* (Sac) was kindly provided by Dr J. Rudolf.

The Scottish Structural Proteomics Facility (SSPF) at the University of St Andrews provided the following vectors for polyhistidine-tagged, recombinant protein expression: Sso PBL2025 Hel308 (pDEST14-*hel308*), Sso2452 (pDEST14-*sso2452*) and SarDinG (pDEST14-*dinG*). The following genomic DNA was used to clone each gene sequence: Hel308, *Sulfolobus solfataricus* (Sso) strain PBL2025; Sso2452, *S. solfataricus* strain P2; DinG, *Staphylococcus aureus* (Sar) strain MRSA252. In each case, Gateway cloning was employed to insert the gene into the pDEST14 vector. The oligonucleotide primer sequences used to generate these constructs are noted in Appendix 1 (table A1.1).

Vectors encoding untagged, recombinant SsoRadA (pET19b-*radA*), SsoSSB (pET19b-*ssb*), SsoHjc (pET19b-*hjc*), SsoHje (pET19b-*hje*), SsoPCNA (pET19b-*pcna*) and SsoAlba1 (pET19b-*alba1*) were regenerated from frozen stocks produced by former lab members.

Plasmids were sequenced at The Sequencing Service, School of Life Sciences, Dundee.

2.1.2 SITE-DIRECTED MUTAGENESIS

Mutations of SacXPD, Sso PBL2025 Hel308 and SarDinG were generated using the Stratagene XL QuikChange Mutagenesis kit and the primers indicated in Appendix 1 (table A1.2). B. Petrovic-Stojanovska generated the SarDinG mutants. The Sso PBL2025 Hel308-646STOP mutant was provided by Dr J. Richards. DNA sequencing and mass spectrometry were used to confirm the introduction of base mutations.

2.1.3 PROTEIN EXPRESSION

All proteins were expressed in *Escherichia coli* BL21 Rosetta cells (Novagen). Cultures were grown at 37 ºC in Luria Bertani (LB) medium containing the appropriate antibiotic (ampicillin: 100 µg/ ml or kanamycin: 35 µg/ ml; both from Melford) until an OD_{600} of 0.8 – 1.0 was achieved. For expression of Sso PBL2025 Hel308, SsoRadA, SsoSSB, and SsoHjc, protein expression was induced by the addition of 0.2 mM IPTG (Melford) and growth was continued at 37 ºC for 3 h at 200 rpm. For SacXPD and SsoHje, protein expression was induced with 0.1 mM IPTG and growth was continued for ~14 h at 25 ºC and 200 rpm. For Sso2452, protein expression was induced with 0.4 mM IPTG and incubation at 25 ºC for ~ 14 h at 200 rpm. For SarDinG, 0.4 mM IPTG was used to induce protein expression and growth was then continued at 37 ºC for 4 h at 200 rpm. The same expression conditions were employed for wild-type (WT) and mutant forms of each protein.

2.2 PROTEIN PURIFICATION

Untagged recombinant WT and mutant SacXPD (Rudolf et al., 2006), SsoHjc (Kvaratskhelia and White, 2000), SsoHje (Middleton et al., 2003), SsoPCNA (Williams et al., 2006), SsoSSB (Wadsworth and White, 2001), SsoRadA (Ariza et al., 2005) and polyhistidine-tagged WT and mutant Sso PBL2025 Hel308 (Richards et al., 2008b) were purified as described previously.

Cell cultures containing expressed Sso2452 protein were harvested by centrifugation (6000 rpm) for 20 min at 4ºC (Beckman rotor JLA 8.1000). Pellets were resuspended in lysis buffer (20 mM Tris, pH 7.5 (Sigma), 500 mM NaCl (Fisher Scientific), 1 mg/ ml DNase (Sigma), 1 mg/ ml lysozyme (Sigma) and 1 mM benzamidine (Sigma)) and sonicated (Soniprep 150, MSE (UK) Ltd) for 6 minutes (2

min x 3) on ice. The cell lysate was centrifuged at 20,000 rpm (Beckman, JA 25.50 rotor) at 4 °C for 30 min, heat-treated (60 °C for 20 min) and then centrifuged for a further 30 min. Cleared lysate was filtered (Acrodisc® 0.45 µm syringe filter) and loaded onto a 5 ml HiTrap Chelating HP column (GE Healthcare) equilibrated with binding buffer (20 mM Tris, pH 7.5, 500 mM NaCl, 10 mM imidazole). The protein was eluted with a linear gradient of imidazole (500 mM) and protein-containing fractions were identified by SDS-PAGE (Invitrogen), pooled, and purified to homogeneity on a HiLoad 26/ 60 Superdex 200 size-exclusion column (GE Healthcare) equilibrated with gel-filtration buffer (20 mM Tris, pH 7.5, 500 mM NaCl, 1 mM ethylenediaminetetraacetic acid (EDTA; Sigma) and 1 mM DTT (Melford)). Pure protein was incubated with 200 ng/ µl tobacco etch virus (TEV) protease for ~ 14 h at 22 °C to remove the polyhistidine tag and cleaved protein was separated from remaining tagged protein using an equilibrated 5 ml HiTrap Chelating HP column. The column flow-through, containing untagged protein, was concentrated and analysed by MALDI-TOF and ESI mass spectrometry to confirm identity and integrity. Protein was diluted to contain 20% (v/v) glycerol (Sigma) and frozen in liquid nitrogen for storage at -80 °C.

To purify SarDinG, cultures containing expressed protein were harvested, lysed by sonication in lysis buffer (as above) and centrifuged at 20,000 rpm, for 30 min at 4 °C (Beckman, JA 25.50 rotor) to clear the lysate of precipitated and insoluble proteins. Protein was bound to a HiTrap 5 ml Chelating HP column equilibrated with binding buffer (as above) and eluted with a linear imidazole gradient (500 mM). All remaining purification steps are as described for Sso2452. The Walker A box mutant (K304A) and nuclease domain mutants (D10A/ E12A and Δnuclease) of DinG were expressed and purified as for the WT protein. Proteins were mixed with glycerol (20% v/v final volume), frozen in liquid nitrogen and stored at -80 °C.

2.2.1 DETERMINATION OF PROTEIN CONCENTRATION

Protein concentrations were calculated from the absorbance measured at 280 nm and the theoretical molar extinction coefficient obtained from ExPASy ProtParam analysis (http://ca.expasy.org/tools/protparam.html).

2.3 CIRCULAR DICHROISM (CD) SPECTROSCOPY

2.3.1 BACKGROUND TO CD SPECTROSCOPY

Plane-polarised light comprises two circularly polarised components, L (left-handed) and R (right-handed) that rotate counter-clockwise and clockwise, respectively, with equal magnitude (Kelly et al., 2005). Chiral (optically-active) chromophores absorb the L and R components of plane-polarised radiation to varying extents: providing absorption of L and R components does not occur or occurs to equal extents, the resulting radiation is polarised in the original plane; however, if differential absorption of the two components of light occurs, the resulting light would possess elliptical polarisation (figure 2.1). CD spectroscopy measures the extent to which a sample, e.g. pure protein, absorbs L and R radiation, and generally reports the absorption in terms of ellipticity (θ) in degrees (Kelly et al., 2005).
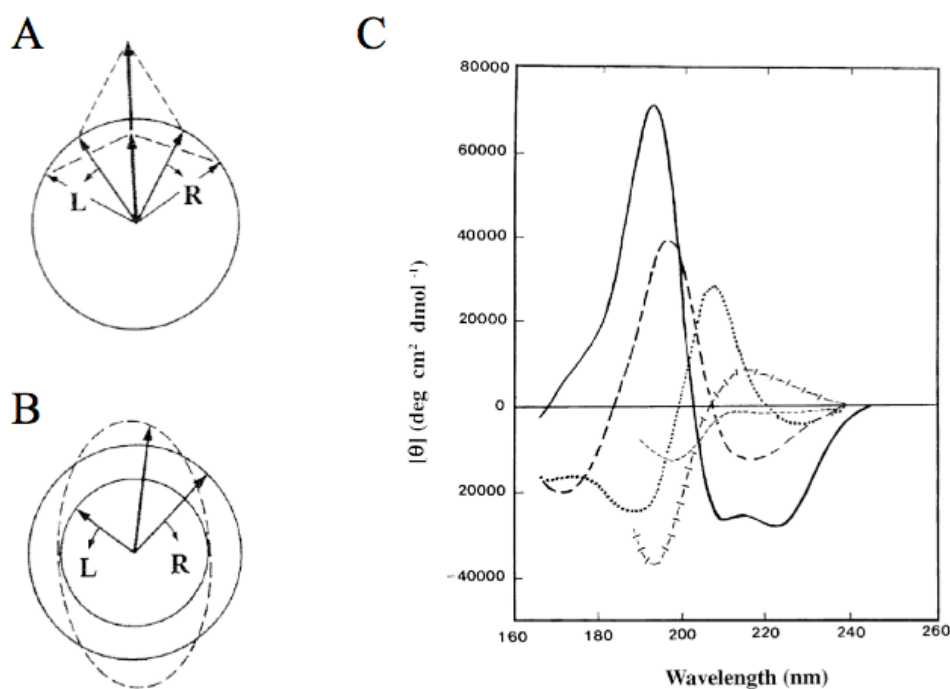
Figure 2.1: Basis of CD spectroscopy

CD spectroscopy measures the differential absorption of L and R light by a chiral material. If L and R are absorbed equally, the resulting light is polarised in the original plane (A); if L and R are absorbed differentially, the resulting radiation has elliptical polarisation (B). The secondary structure of a protein can be determined by absorption of far UV light (< 240 nm) by the peptide bond. The CD spectrum shown (C) reveals characteristic signals indicative of regular secondary structure. Solid line, α-helix; long dashed line, anti-parallel β-sheet; dotted line, type I β-turn; cross-dashed line, extended 3_1-helix or poly (Pro) II helix; short dashed line, irregular structure. Adapted from (Kelly et al., 2005).

45

Chiral chromophores absorb plane-polarised light and therefore generate a CD spectrum. In proteins, plane-polarised radiation is absorbed differentially by distinct structural elements, e.g. α-helices and β-sheets (figure 2.1 C) (Kelly et al., 2005). Particular regions of interest include the peptide bond (absorption below 240 nm), aromatic amino acids (absorption between 260-320 nM) and disulphide bonds (broad absorption around 260 nm) (Kelly et al., 2005). Information regarding the regular secondary structure of a protein can be obtained mainly from the absorption of plane-polarised radiation by the peptide bond.

2.3.2 CD MEASUREMENTS

For CD spectroscopy, all proteins were buffer-exchanged into CD buffer (10 mM Tris, pH 7.8, 200 mM NaF (Fisher Scientific), 0.5 mM EDTA). Spectra were recorded on the JASCO J-810 Spectropolarimeter using Spectra Manager Software. For near UV readings (450 - 260 nm), protein concentrations were adjusted to 1 mg/ ml and spectra were recorded at a scan rate of 20 nm/ min using a cuvette with a pathlength of 0.5 cm. For far UV measurements (260 - 180 nm), protein concentrations were adjusted to 0.5 mg/ ml and spectra were analysed at a scan rate of 50 nm/ min using a cuvette with a pathlength of 0.02 cm. For spin-labelled C86/ C284 Hel308, far-UV measurements were performed in the presence and absence of equimolar (0.5 mg/ ml) or 20X excess (10 mg/ ml) of ADP (Sigma)/ $MgCl_2$ (Fisher Scientific). In all cases, a control sample, lacking protein, was also performed (repeated six times). Spectral measurements were repeated six times for each protein, averaged and corrected for control sample values. Spectra were analysed using the DichroWeb (Lobley et al (2002)) server (http://www.cryst.bbk.ac.uk/cdweb/html/home.html).

2.4 ELECTRON PARAMAGNETIC RESONANCE (EPR)

2.4.1 BACKGROUND TO EPR

Every electron possesses a charge and spin quantum number (s) of ½ that endows the particle with a magnetic dipole moment (μ) (figure 2.2 A). This magnetic property of the electron forces it to behave like a bar magnet such that, in the presence of an external magnetic field (B_0), the projection of the electron spin (m_s) lies parallel ($m_s = -1/2$) or antiparallel ($m_s = +1/2$) to the external magnetic field B_0 (figure 2.2 B). These two alignments result from a splitting of energy (E) levels such that electrons

aligned parallel to the external field lie in a lower energy state than those lying antiparallel. The extent of the energy splitting is directly proportional to the strength of the magnetic field, i.e. by increasing the strength of the magnetic field, the gap between the m_s +1/2 and m_s -1/2 energy states is increased. When the strength of the magnetic field tunes the difference between the energy levels to the energy of the microwave frequency, as represented by the double arrow in figure 2.2 C, the "field for resonance" has been achieved (Weil et al., 1994).
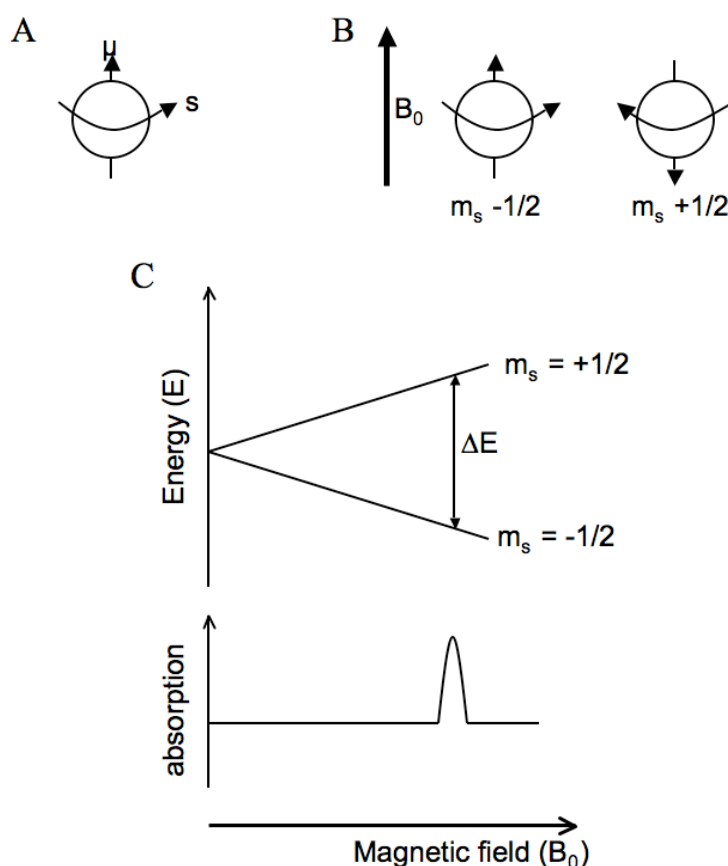
Figure 2.2: Fundamental concepts of EPR

(A) A unpaired electron possesses a spin (s) and a magnetic moment (μ) such that, when an external magnetic field (B_0) is applied to the electron, it aligns parallel or antiparallel to the applied field (B). (C) Electrons with m_s +1/2 have a higher energy level than those of m_s -1/2, and EPR measures the energy difference (ΔE) between these two states. The magnetic field is increased until it tunes the energy gap between the electron spin states to the frequency of the external microwave source, resulting in resonance (depicted by the vertical double-arrow). When electrons establish resonance, net absorption of energy occurs and this can be measured using EPR. (A-B) adapted from the thesis of Dr J. Rudolf; (C) adapted from http://www.bruker-biospin.com/cw.html.

Conventional EPR measurements are typically performed using one-dimensional continuous wave (cw)-EPR, which relies on a varying magnetic field and constant microwave frequency (Weil et al., 1994). When the sample is exposed to the microwave frequency, a photon with electromagnetic energy is released. As the magnetic field increases, the energy difference between the spin states of the electron increases until ΔE equals the energy of the microwave photon (i.e. resonance occurs), as described above. This is the basis of Planck's law that states that electromagnetic energy will be absorbed providing ΔE = hν, where h is Planck's constant and ν is the frequency of the radiation (Weil et al., 1994). Therefore, at the field for resonance, an unpaired electron can move between the two spin states by either absorbing or emitting energy. According to the Maxwell-Boltzmann distribution, the majority of electrons lie in the lower energy state such that, during resonance, net energy absorption occurs (Weil et al., 1994). This absorption can be measured by EPR spectroscopy, and processed to produce a spectrum (figure 2.2 C).

Unfortunately, the spatial resolution cw-EPR is limited to distances of 8-20 Å, with large distances compounded by overlapping spectra (Banham et al., 2008). Advances in the field have minimised these issues by employing pulsed methods, such as the pulsed electron-electron double resonance (PELDOR) technique. By employing a 4-pulsed sequence of PELDOR, overlapping spectra that are observed by cw-EPR can be separated and distance parameters easily deciphered. PELDOR spectroscopy provides a means of measuring the dipole-dipole interaction between electron spin centres separated by 15 and 80 Å, distances relevant to the study of proteins (Prisner et al., 2001; Jeschke, 2002). In a typical PELDOR spectroscopy, a paramagnetic sample is exposed to a short (<20 ns) and intense (>300 W) pulse of microwave energy. As a consequence, the sample emits a signal of microwave energy, known as an electron spin echo (ESE), and the kinetics of the decay of this signal between pulses is measured as a function of time. This data can be subsequently processed by Fourier transform into the frequency domain, or alternatively (as performed by Olav Schiemann in the EPR experiments presented in this thesis) using Tikhonov regularisation to provide a measure of the distance between the unpaired electrons (http://www.bruker-biospin.com). Modulation of the ESE decay relies on the dipolar interactions between unpaired electrons and is very sensitive to changes in the local environment (http://www.bruker-biospin.com); for instance, fast dampening of the modulation of the

ESE decay suggests that a sample has undergone conformational changes such that the distance between the unpaired electrons has increased. Thus, EPR can facilitate the investigation of the local structure and dynamics of biological samples.

2.4.2   SPIN-LABELLING OF HEL308

Proteins were dialysed into labelling buffer (10 mM Tes (Sigma), pH 7.4, 300 mM NaCl) and 50 μM of protein was then incubated with 250 μM [(1-oxyl-2,2,5,5-tetramethylpyrroline-3-methyl)-methanethio-sulfonate] spin label (MTSL; Sigma), in a final volume of 2.5 ml, for 2 h at 4°C. Thiol-reactive MTSL binds the protein (figure 2.3), increasing the molecular mass by 185 daltons.

side-chain R1
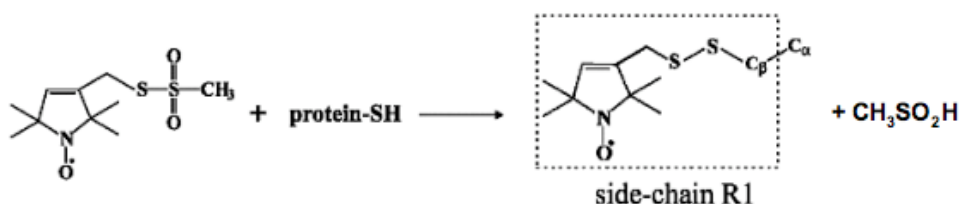
Figure 2.3: Site-directed spin labelling
Site-directed spin labelling involves the formation of a covalent bond between the thiol group (-SH) of an amino acid, typically cysteine, and a paramagnetic spin label, such as MTSL (left). Attachment of MTSL increases the protein mass by 185 Da.

Unincorporated label was removed by gel filtration using disposable PD-10 desalting columns (GE Healthcare) equilibrated with labelling buffer. Protein-containing fractions were determined by measuring the absorbance at 280 nm, pooled and buffer-exchanged into labelling buffer containing deuterium oxide (D_2O). For EPR spectroscopy, samples were concentrated to provide 300 μM protein, 600 μM MTSL in a final volume of 50 μl and then diluted to 100 μl with deuterated ethylene glycol (Cambridge Isotope Laboratories). Where appropriate, ATP/ ADP/ AMPPNP (Sigma), magnesium chloride and/ or oligonucleotides (Operon), in D_2O, were added to the MTSL-labelled protein samples such that the final EPR sample (100 μl) contained 150 μM protein, 300 μM MTSL, 3 mM (or 150 μM where 1:1 ratio with protein was required) ATP/ ADP/ AMPPNP, and/ or 3 mM (or 150 μM where 1:1 ratio with protein was required) magnesium chloride and/ or 750 μM DNA (34mer; Appendix 1, table A1.8) and 50 % (v/v) ethylene glycol (in D_2O) in deuterated labelling buffer. Samples were frozen rapidly by immersing in an ether solution (comprising 1

methylcyclohexane: 4 iso-pentane) cooled with liquid nitrogen. Samples were stored in liquid nitrogen until required.

2.4.3   PELDOR EPR MEASUREMENTS

PELDOR EPR measurements were performed by Dr Olav Schiemann. All PELDOR spectra were recorded on a Bruker ELEXSYS E580 pulsed X-band EPR spectrometer, with a standard flex line probe head housing a dielectric ring resonator (MD4). The instrument was equipped with a continuous flow helium cryostat (CF935) and temperature control system (ITC 502), both from Oxford Instruments. The second microwave frequency was coupled into the microwave bridge using a commercially available setup from Bruker. All pulses were amplified via a pulsed traveling wave tube (TWT) amplifier (117X) from Applied Systems Engineering. The resonator was over-coupled to a quality factor Q of about 50. PELDOR experiments were performed with the pulse sequence p/2(n_A)-t_1-p(n_A)-(t_1+t)-p(n_B)-(t_2-t)-p(n_A)-t_2-*echo*. The detection pulses (n_A) were set to 16 ns for the p/2 and 32 ns for the p pulses and applied at a frequency 80 MHz higher than the resonance frequency of the resonator. The pulse amplitudes were chosen to optimise the refocused echo. The p/2-pulse was phase-cycled to eliminate receiver offsets. The pump pulse (n_B) with a length of 28 ns was set at the resonance frequency of the resonator. The field was adjusted such that the pump pulse is applied to the maximum of the nitroxide spectrum, where it selects the central $m_I = 0$ transition of A_{zz} together with the $m_I = 0, \pm 1$ transitions of A_{xx} and A_{yy}. The pulse amplitude was optimised to maximise the inversion of a Hahn-echo at the pump frequency. All PELDOR spectra were recorded at 50 K with an experiment repetition time of 4 ms, a video amplifier bandwidth of 20 MHz and an amplifier gain of 54 dB. t_1 was set to 380 ns and t_2 varied from 2500 to 6000 ns. About 300 scans were accumulated with time increments Dt of 8 ns or 16 ns giving an approximate measurement time of 18 to 48 hours necessary to obtain a signal-to-noise ratio > 200:1. Proton modulation was suppressed by addition of 8 spectra of variable t_1 with a Dt_1 of 8 ns. The obtained time traces were divided by a mono-exponential decay to eliminate intermolecular contributions and renormalized. Distance distributions were obtained from the background corrected data by using the program DeerAnalysis.

2.5 SINGLE-MOLECULE FLUORESCENCE RESONANCE ENERGY TRANSFER (SMFRET) STUDIES

Luminescence, including both fluorescence and phosphorescence, describes the emission of a photon from an electronically excited species (Lakowicz, 2006). For instance, electrons, moving in orbits around the nucleus of an atom, are able to absorb energy from an externally applied source of electromagnetic radiation such that they are promoted from an electronically stable ground state to an "excited" higher energy state. This excited state is unstable, dictating that the electron emits the absorbed energy as a photon and returns to its original low energy state (figure 2.4).
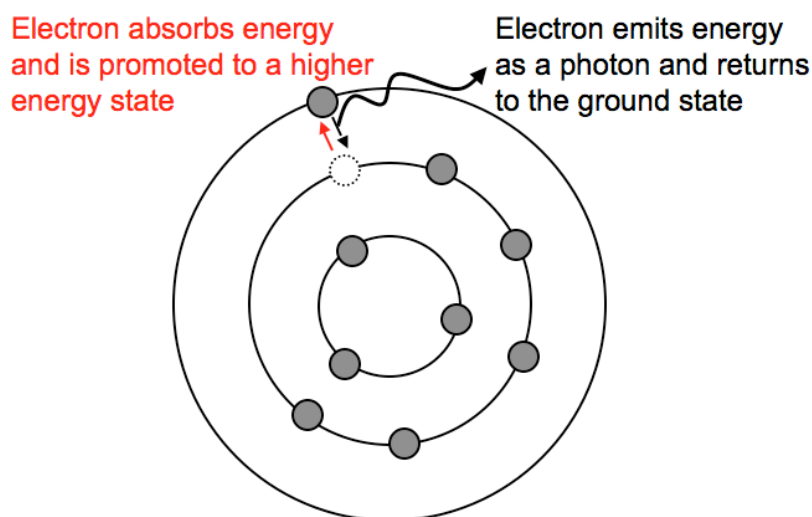
Figure 2.4: Principle of luminescence

Electrons (small grey circles) move in orbitals (depicted as large, open circles) around the nucleus of an atom, arranged according to increasing energy levels. An electron can absorb electromagnetic radiation, promoting it from a ground state to an unstable higher energy state (indicated by red arrow). The unstable nature of this promoted state means that the electron eventually emits the absorbed energy as a photon (wavy black arrow) and falls back into the low energy state (depicted by straight, black arrow).

FRET relies on the excitation of electrons within fluorophores, which are typically aromatic molecules such as cy3 and cy5 (figure 2.5). Each fluorophore possess a unique absorption and emission spectrum (figure 2.5 A), corresponding to the wavelength of light at which it optimally absorbs and emits electromagnetic radiation, respectively. The process of FRET relies on the overlap of the donor emission spectrum with the excitation spectrum of the acceptor, as illustrated in figure 2.5 B. Excitation of the donor molecule may, therefore, result in the transfer of excitation energy to the

acceptor species via long-range dipole-dipole interactions (Ha, 2001; Lakowicz, 2006). Thus, during FRET, the energy absorbed by the donor is emitted by the acceptor at a higher wavelength.
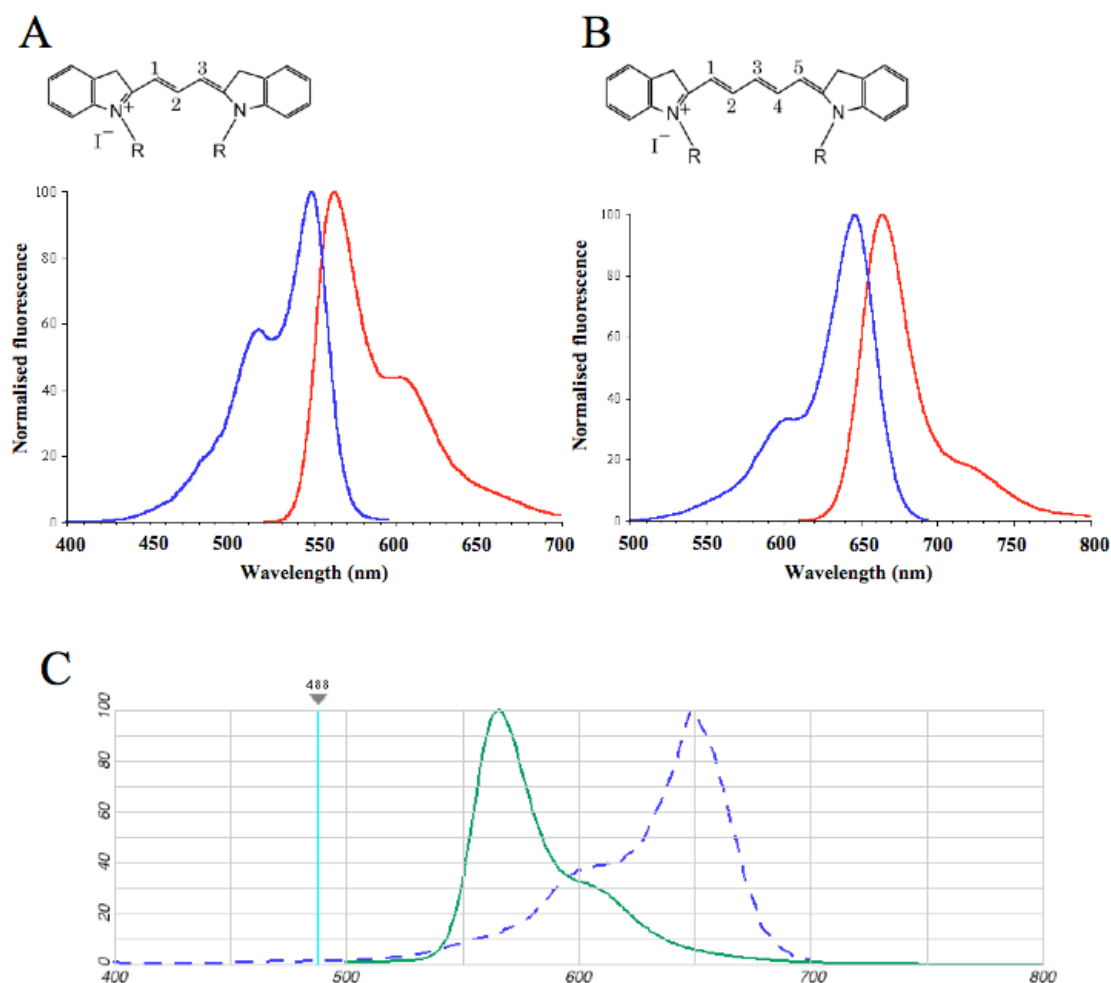
Figure 2.5: Excitation and emission spectral characteristics of cy3 and cy5

Cy3 (A) and cy5 (B) are aromatic fluorophores, typically used during smFRET experiments as a donor-acceptor pair, respectively. The excitation (blue trace) and emission (red trace) spectra of each fluorophore are shown (taken from Invitrogen). (C) Spectrum displaying the overlap between cy3 emission (green trace) and cy5 excitation (blue dashed line). This overlap permits FRET between cy3 and cy5 fluorophores. (C) was generated using Fluorescence SpectraViewer (Invitrogen).

FRET has been described as a "spectroscopic ruler" because the efficiency with which energy is transferred between the donor and acceptor molecules is distance-dependent (Ha, 2001). These parameters are related to each other by the Förster equation (Equation 1 below) (Ha, 2001; Lakowicz, 2006):

$$E = \frac{1}{1 + (R/R_0)^6} \qquad \text{Equation 1}$$

(E refers to FRET efficiency, R is the distance between the donor and acceptor, and R_0 is the Förster distance). The Förster distance is described as the distance at which 50% of the energy is transferred and is approximately 60 Å for the cy3-cy5 pair (Selvin, 2008). Given that the efficiency of FRET decreases by the sixth power of the distance between the donor and acceptor dyes, this technique provides a very sensitive measure of distances: very small distance changes between donor and acceptor molecules produce large changes in FRET efficiency.

2.5.1 DNA AND PROTEIN LABELLING WITH CY-DYES

Oligonucleotides (Appendix 1, table A1.6) were generated with functional groups that permitted the attachment of an amine-reactive cy3 dye (Amersham). Where indicated, oligonucleotides were also synthesised with a 3' biotin moiety. Oligonucleotides were purified as described in section 2.7.1. In a final volume of 100 µl, 250 µg cy3 (in 14 µl DMSO (Sigma)) was added to 4 µl DNA (stock 25 µg/ µl) in 0.1 M tetraborate buffer (pH 8.5; Sigma). Reactions were incubated overnight at room temperature, prior to purification on denaturing 20 % (v/v) polyacrylamide: TBE gels at 22 W for 4 h to separate labelled from unlabelled DNA. Labelled DNA was excised from the gel, incubated in TE buffer (10 mM Tris, pH 7.5, 1 mM EDTA) overnight at 4 ºC and ethanol precipitated. DNA was stored at -20 ºC as a dried pellet and resuspended in ultrapure water as required. The concentration of DNA and cy3 dye was determined from the absorbance at 260 nm and 550 nm, respectively, and their respective extinction coefficients (DNA, determined from http://biophysics.idtdna.com/; cy3 dye, 150000 $M^{-1}cm^{-1}$). The labelling efficiency of the reaction was subsequently determined.

Site-directed mutagenesis was employed to substitute residue 595 (serine) for cysteine in Sso PBL2025 Hel308. The purified protein (50 µM) was incubated (with shaking) at room temperature for 1 h in a final volume of 2.5 ml containing labelling buffer (10 mM Tes, pH 7.4, 300 mM NaCl) and 250 µg cy5 dye (in DMSO). The sample was loaded onto a disposable PD-10 desalting column (GE Healthcare) equilibrated with labelling buffer to separate unlabelled protein, labelled protein, and free label. Measuring the absorbance at 280 nm and 650 nm identified fractions containing labelled protein. These fractions were pooled, concentrated and exchanged

into gel filtration buffer (200 mM Tris, pH 7.5, 500 mM NaCl, 1 mM EDTA and 1 mM DTT). The concentrations of protein and dye were determined using the absorbance at 280 nm and 650 nm, respectively, and their respective extinction coefficients (Hel308, 101190 $M^{-1}cm^{-1}$; cy5 dye, 250000 $M^{-1}cm^{-1}$), which in turn permitted an assessment of labelling efficiency. Labelled protein was stored at 4 ºC.

2.5.2 CLEANING OF QUARTZ SLIDES

Standard quartz slides (Deckgläser) were cleaned as outlined in table 2.1. In all cases, H_2O refers to ultrapure water and potassium hydroxide (KOH) (Fisher Scientific) was filtered before use. Methanol, hydrochloric acid (HCl), acetone and hydrogen peroxide (H_2O_2) were from Fisher Scientific. Hellmanex® II (Hellma) is a cleaning concentrate specifically adapted for use in photometric measurements. For stages 2-11 (refers to table 2.1), samples were immersed in the reagent and sonicated for the indicated incubation time.

Table 2.1: Protocol for cleaning quartz slides.

Stage	Reagent	Incubation time (min)
1	1 H_2O : 3 HCl : 6 H_2O_2	120
2	Hellmanex® II	15
3	H_2O	5
4	Acetone	15
5	H_2O	5
6	1 M KOH	15
7	H_2O	5
8	Methanol	15
9	H_2O	5
10	1 M KOH	15
11	H_2O	5

2.5.3 AMINOSILATION OF QUARTZ SLIDES AND COVERSLIPS

Cleaned slides were dried with nitrogen gas and heated, using a blow-torch, for 6 min (3 min for each slide face). Slides and coverslips (Deckgläser) were then immersed in methanol and dried (with nitrogen) before a 5-min incubation (with

sonication) in aminosilation solution (50 ml methanol (Fisher-Scientific), 2.5 ml acetic acid (Fisher-Scientific), 200 µl aminosilane (Fluorochem)). Slides and coverslips were washed with water, followed by a 50: 50 mix of water and methanol, and then methanol before drying with nitrogen.

2.5.4 PEGYLATION OF QUARTZ SLIDES AND COVERSLIPS

Biotinylated PEG (2 mg) (Laysan Bio Inc.) and non-biotinylated PEG (100 mg) ((Laysan Bio Inc.) were dissolved in 320 µl of PEG buffer (100 mM sodium bicarbonate (Sigma), pH 8.5) and centrifuged for 1 min at 13,000 rpm. The supernatant ('PEG solution') was decanted into a fresh eppendorf tube and 65 µl was added to the surface of each slide (after aminosilation). A coverslip was carefully positioned over the PEG solution, on top of the slide, and the entire assembly was incubated at room temperature for 3 h under low-light conditions. Both slide and coverslip were then cleaned with water/ methanol: water (50: 50)/ methanol, and then dried using nitrogen gas. Using double-sided tape, the PEGylated surface of the coverslip was mounted onto the PEGylated surface of the slide and the assembly was then wrapped in foil, stored in a 50 ml plastic (sterile) tube and left at 4 °C until required.

2.5.5 TOTAL INTERNAL REFLECTION (TIR) MICROSCOPY

smFRET experiments of Hel308 were performed by wide-field microscopy exploiting total internal reflection (TIR). The experimental set-up for single-molecule studies (figure 2.6) included a CL-2000 diode pumped crystal laser (CrystaLaser), an Olympus IX 71 inverted TIR microscope, and a CCD camera.
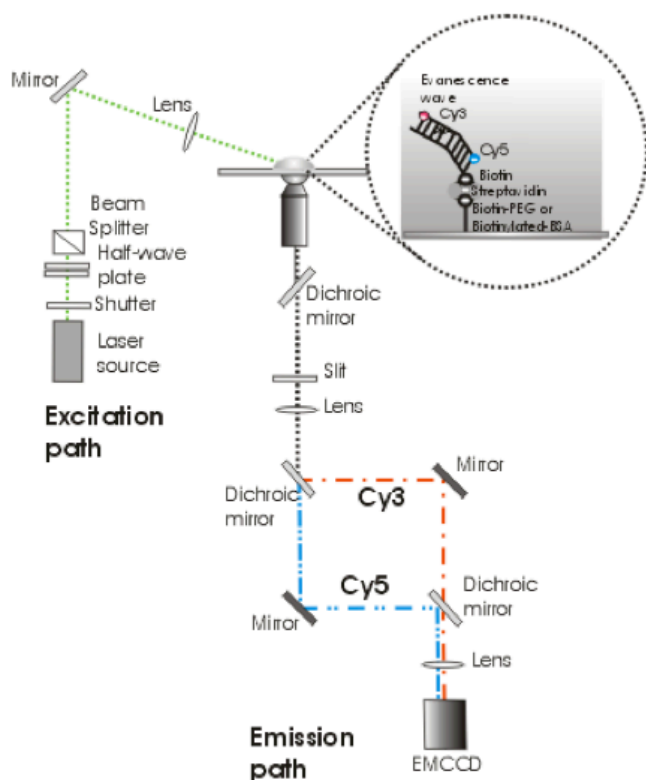
Figure 2.6: TIR microscopy apparatus for single-molecule studies

The diagram details the path of light from the excitation source to the microscope slide, containing immobilised DNA possessing a cy3- and cy5-dye, and generation of the evanescence wave. The path taken by the resulting emission light, from both cy3- and cy5-molecules, to the CCD detection camera is shown.

TIR relies on the induction of an evanescence wave, which is generated when the incident beam of light bends (or refracts) upon encountering the interface between two media, e.g. a glass prism and water (Ross et al., 2007). However, the illuminating beam can either bend (to generate the evanescence wave) or be reflected upon encountering the media interface depending on the difference between the refractive indices of the two media and the angle of incident light (Ross et al., 2007). A typical smFRET experiment involving Hel308 included a quartz layer (consisting of a quartz prism, oil, and quartz microscope slide) of a high refractive index (~ 1.43) and a layer of water of low refractive index (~ 1.33) in direct contact with the base of the prism (figure 2.7 A). Providing light from the excitation source meets the interface between the two media at a sufficiently high angle, known as the critical angle, its direction becomes parallel to the interface (Ross et al., 2007). When the incident light strikes the interface at an angle greater than the critical angle, it is reflected back into the quartz layer creating an evanescence wave within the aqueous sample. The evanescence field excites fluorophores only within 100 nm (along z-axis) of the biological sample because the intensity of the evanescence field decays exponentially with increasing distance from the interface (Ross et al., 2007). This has particularly important implications for single molecule detection: since only a very thin layer of the biological specimen is excited by the evanescence field, visualisation of single

molecules at a higher signal-to-noise ratio is achieved since the effective volume of the sample is reduced (Ha, 2001). Furthermore, a wide-field microscopy technique utilising TIR allows large numbers of single molecules to be visualised simultaneously, which is particularly important when monitoring irreversible or rare catalytic events (Ha, 2001; Ross et al., 2007). For instance, the helicase activity of Hel308 is energy-dependent. Therefore, single-molecule detection of DNA unwinding by the helicase should be performed by injecting ATP/ $MgCl_2$ in real-time. The ability to observe hundreds of reactions in a single field of view becomes crucial to determine reaction kinetics during such experiments.
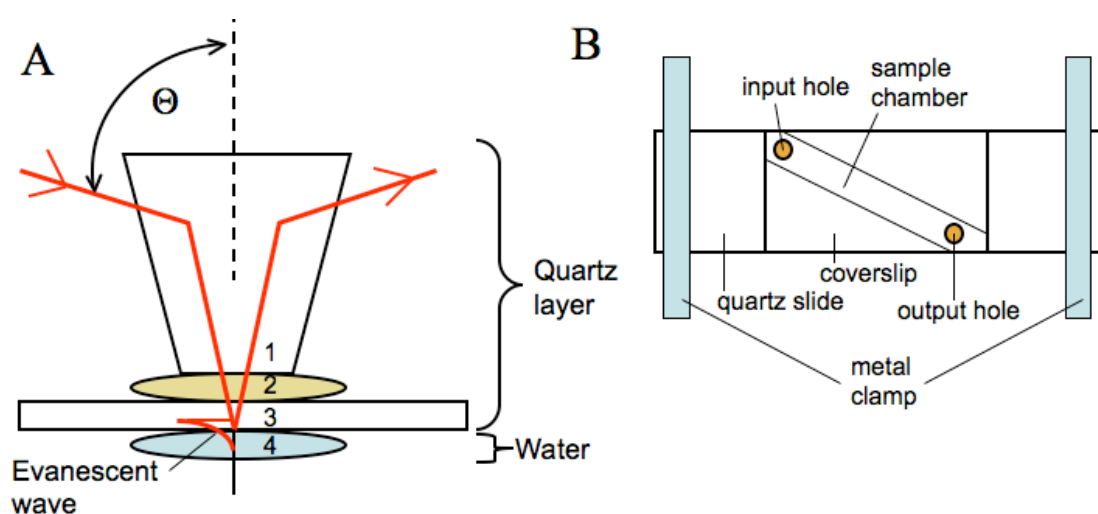
Figure 2.7: Basis of TIR fluorescence microscopy

(A) TIR fluorescence microscopy with Hel308 involved a quartz layer consisting of a quartz prism (1), drop of oil (2) and quartz microscope slide (3). The slide was separated from the objective by a drop of water (4). When the excitation beam (red) strikes the interface between the quartz and water layer at an angle larger than the critical angle (Θ), the evanescence field of excitation is generated within the biological sample. Adapted from (Ross et al., 2007). (B) Slide configuration for smFRET spectroscopy.

2.5.6   REAGENTS FOR SINGLE-MOLECULE ASSAYS

The reagents required for all the single-molecule experiments described in this thesis are noted in Table 2.2.

Table 2.2: Reagents for single-molecule assays

Reagent	Constituents
T50 buffer	50 mM Tris, pH 7.5, 50 mM NaCl
T50-glucose buffer	50 mM Tris, pH 7.5, 50 mM NaCl, 6% (v/v) glucose (Sigma)
Gloxy solution	100 µl T50 buffer, 20 µl glucose-catalase (Roche), 10 mg glucose-oxidase (Sigma)
Trolox solution	9 ml methanol, 1 ml ultrapure water, 150 mg trolox (Sigma), adjust to pH 9.5 with sodium hydroxide (Fisher Scientific)
Imaging buffer	Per 100 µl: 98 µl T50-glucose buffer, 1 µl gloxy solution, 1 µl trolox solution

2.5.7   CALIBRATION OF CY3 AND CY5 CHANNELS

Prior to each experiment, a prepared slide of immobilised fluorescent beads (provided by Dr Carlos Penedo, University of St Andrews) was used to calibrate the microscope and ensure the cy3 and cy5 channels were aligned appropriately. Excitation of the fluorescent beads at 532 nm produced sufficient intensity to observe individual beads in both the cy3 and cy5 channel. MATLAB software was subsequently used to select three paired cy3/ cy5 spots (each paired spot corresponds to the same fluorescent bead) to enable the software to appropriately re-align the channels to the same position.

2.5.8   SINGLE-MOLECULE ASSAYS

At the start of each experiment, prepared slides were treated with Epoxy resin (Araldite® Adhesives) to create a sealed chamber through which aqueous material could be injected onto the slide (figure 2.7 B). To confirm the absence of fluorescent impurities, each slide was washed with imaging buffer and mounted on the microscope stage and excited at 532 nm (cy3 compatible) and 638 nm (cy5 compatible) using a CL-2000 diode pumped crystal laser. The microscope stage was moved in the z direction to focus the sample within the evanescence field, and moved in the x/ y directions to visualise the slide surface (each imaged area had dimensions of approximately 100 x 50 µm).

Slides were subsequently incubated with 0.2 mg/ ml streptavidin (Sigma) (in T50 buffer) at room temperature for 15 min and then washed with T50 buffer. Biotinylated, cy3-DNA (200 pM), diluted in T50 buffer, was injected onto the slide and

incubated at room temperature for a further 15 min. Slides were washed with imaging buffer, mounted on the microscope stage, and imaged at 532 nm to verify appropriate density of DNA across the slide. Solutions of protein and/ or ATP and/ or magnesium chloride were incubated with the slide in the presence of imaging buffer, as specified in chapter 6. Movies of 2000 frames were recorded at a rate of 1 frame/ 50 ms and processed using MATLAB software.

2.6   BULK FRET EXPERIMENTS

Bulk FRET experiments were performed at 20 ºC using a Varian Cary Eclipse fluorimeter equipped with automatic polarisers. DNA and protein molecules were labelled with cy3 and cy5, respectively, as described in section 2.5.1. For direct binding experiments, DNA (40 nM) was added to FRET buffer (20 mM Hepes, pH 7.5, 150 mM NaCl, 1 mM DTT, 0.01% (v/v) Triton-X100) to give a total volume of 150 μl. The sample was excited at 547 nm and emission was collected between 557 and 700 nm. Cy5-labelled protein was titrated into the cy3-DNA solution and, following each titration, the cy3 dye was excited (547 nm) and emission collected (as above). For competitive binding experiments, 30 nM cy3-ssDNA and 30 nM cy5-Hel308 were incubated in FRET buffer in a total volume of 150 μl prior to the titration of unlabelled protein. Before the titration of unlabelled protein, and after each subsequent addition of unlabelled DNA, the sample was excited at 547 nm and emission was collected between 557 – 700 nm. For both direct and competitive titrations, fluorescence emission values were plotted against labelled or unlabelled protein concentration, respectively. For both direct and competitive binding experiments, a control sample was performed in the absence of cy5 molecules; the ratio of the fluorescence emission between 560 and 567 nm (cy3) and between 650 and 680 nm (cy5) from the control sample was subtracted from the donor: acceptor emission ratio obtained in the presence of cy5 molecules. The corrected ratio was plotted against labelled (direct titration) or unlabelled (competitive titration) protein concentration.

For bulk FRET measurements of DNA translocation by Hel308, 40 nM cy3-ssDNA DNA and 250 nM cy5-protein were incubated in FRET buffer in a final volume of 150 μl. Multiwavelength scanning was employed whereby both cy3 and cy5 fluorophores were excited at 547 nm and emission was collected simultaneously at 570 nm and 670 nm, respectively, for indicated time periods. During excitation, ATP/

MgCl$_2$ was added (to a final concentration of 1 mM; 15 µl) at indicated time points. Prior to the addition of ATP/ MgCl$_2$, the sample was excited at 547 nm and emitted light was collected between 557 and 700 nm. This was repeated after the addition of ATP/ MgCl$_2$.

Bulk FRET unwinding studies were performed as outlined for translocation experiments with the exception that 40 nM cy3-labelled 3'overhang DNA was incubated with 250 nM cy5-protein in a final volume of 150 µl.

2.7 GENERATION OF DNA SUBSTRATES AND MARKERS

2.7.1 PURIFICATION OF OLIGONUCLEOTIDES

Oligonucleotides (Operon) were denatured in 50 % (v/v) formamide (Promega) at 80 ºC for 10 min prior to incubation on ice. A denaturing polyacrylamide: TBE gel (12 % (v/v) acrylamide (Flowgen Bioscience), 1x TBE, 7 M urea (Melford), 0.05 % (v/v) APS (Sigma) and 0.15 % (v/v) TEMED (Fluke BioChemika)) was pre-run for 1 h at 22 W before oligonucleotide samples were loaded. Electrophoresis continued at 22 W for a further 90 min to separate DNA fragments. The relevant DNA bands were excised from the gel and incubated overnight at 4 ºC in TE buffer. DNA was subsequently purified by ethanol precipitation, resuspended in ultrapure water and stored at – 20 ºC.

2.7.2 PURIFICATION OF PLASMID DNA

For D-loop assays (see below), the double-stranded supercoiled plasmid pUC19 was purified by caesium chloride/ ethidium bromide density gradient centrifugation. *E. coli* BL21 Rosetta cells containing the pUC19 plasmid (regenerated from frozen lab stocks) were grown in LB media at 37 ºC at 200 rpm for ~ 14 h. Cells were harvested by centrifugation (6000 rpm, 4 ºC, 20 min, Beckman rotor JLA 8.1000) and plasmids were purified according to the Plasmid MaxiPrep kit (Qiagen). Purified DNA was resuspended in 2 ml prior to the addition of 2.4 g caesium chloride (Sigma) and 100 µg ethidium bromide (Fluke BioChemika). Samples were incubated at room temperature for 5 min and then split between two TL100 quickseal centrifugation tubes (Beckman). Tubes were sealed and centrifuged at 80,000 rpm at 20 ºC for ~ 14 h (Beckman Optima, TL-100 rotor) to separate bacterial chromosomal DNA from the pUC19 plasmid (visualised as two bands after centrifugation, the pUC19 plasmid band

located beneath the bacterial DNA). The pUC19 plasmid DNA was extracted using a 25 G hypodermic needle and added to an equal volume of butan-2-ol (Fisher Scientific). The mixture was vortexed, centrifuged (13,000 rpm, 4 ℃, 2 min) and the resulting supernatant was discarded. This was repeated four times, or until the ethidium bromide had been completely removed. To the remaining pellet, 1 ml ultrapure water and 0.6 volumes of isopropanol (Fisher Scientific) were added and the sample was incubated at room temperature for 20 min. The sample was centrifuged again, resuspended in 400 µl ultrapure water and concentrated by ethanol precipitation.

2.7.3 ASSEMBLY AND PURIFICATION OF DNA SUBSTRATES

Oligonucleotides (Appendix 1, table A1.3) used for constructing DNA substrates for catalytic assays were 5' [γ-^{32}P] ATP (Amersham) end labelled using polynucleotide kinase (PNK; Fermentas) and assembled into various structures by slow cooling from 95 ℃ to room temperature overnight. Ficoll-based gel-loading dye (30% (v/v) ficoll (Sigma), 0.25% (w/v) bromophenol blue (Promega), 0.1% (v/v) xylene cyanol (Sigma)) was added to annealed samples, which were subsequently purified on native 12 % (v/v) polyacrylamide: TBE gels (12 % acrylamide, 1x TBE, 0.05 % (w/v) APS, 0.15 % (v/v) TEMED), eluted overnight at 4 ℃ into TE buffer and ethanol precipitated. DNA substrates were resuspended in ultrapure water and stored at -20 ℃.

2.7.4 GENERATION OF MAXIM-GILBERT A+G MARKERS

Markers were generated from 5' [^{32}P]-end labelled oligonucleotides. To a sterile microcentrifuge tube, 50 ng of labelled DNA was incubated with 1 µg calf thymus DNA and 1 µl (4 % v/v) formic acid in TE buffer (final volume of 11 µl) at 37 ℃ for 25 min. The tube was placed on ice and 150 µl 1M piperidine (Sigma) was added. Reactions were incubated at 90 ℃ for 30 min and then placed on ice again. N-butanol (1 ml) was added to each reaction prior to centrifugation (13,000 rpm, 4 ℃, 2 min). The supernatant was removed, 150 µl 1% (w/v) SDS was added and the sample was centrifuged again. The supernatant was discard and 0.5 ml of butan-2-ol was used to rinse the pellet. The supernatant was discarded and the pellet was dried under vacuum for 10 min. The pellet was resuspended in 10 µl formamide-based gel-loading dye (95% (v/v) deionised formamide, 0.025% (w/v) bromophenol blue, 0.025% (w/v) xylene cyanol FF, 5 mM EDTA, pH 8.0, 0.025% (w/v) SDS) and stored at – 20 ℃.

2.8 CATALYTIC ASSAYS

2.8.1 HELICASE ASSAYS

Helicase reactions contained helicase buffer (20 mM MES, pH 6.8, 1 mM DTT), 0.1 mg/ ml BSA, 10 nM 5' [^{32}P]-end labelled DNA and 1 mM ATP/ MgCl$_2$. Reactions were incubated at the assay temperature (as stated in results chapter) for 1 min and then started by the addition of protein (concentration stated in results chapter) to give a final volume of 80 µl. Along an appropriate time course, samples (10 µL) were taken and immediately added to 20 µL chilled STOP buffer (10 mM Tris, pH 8.0, 5 mM EDTA, 5 µM competitor DNA, 0.5 % (w/v) SDS, and 1 mg/ ml proteinase K (New England Biolabs)). Samples were incubated at room temperature for 15 min to allow protein digestion by proteinase K and then added to ficoll-based gel-loading dye and separated by electrophoresis on a native 12% polyacrylamide:TBE gel for 3 hours at 130 V. Gels were exposed overnight and processed by phosphorimaging (Fuji FLA5000). ImageGauge software (Fuji) was used to quantify the extent of DNA unwinding. DNA unwound (%) was calculated: (unwound/ total) x 100, and plotted against time using KaleidaGraph software.

For all helicases assays, three controls samples were included: a size marker was generated by incubating the DNA substrate at 95 ºC for 5 min in the presence of STOP buffer to prevent substrate reannealing; an end-point reaction performed in the absence of protein provided a measure of substrate stability; an end-point reaction performed in the absence of ATP/ MgCl$_2$ assessed the ATP-dependence of the unwinding reaction.

To determine the effect of DNA binding proteins on the ability of Hel308 to unwind DNA, DNA-binding proteins (final concentration 10 µM) were incubated with the DNA for 5 min prior to the addition of Hel308 (0.5 µM). Reactions continued as described above.

To determine the effect of pH on DinG helicase activity, the buffering component of the helicase buffer (above) was changed to Mes [pH 6.5], Tris [pH 7.0] or Tris [pH 7.5].

To determine the heat-stability of XPD, protein (stock solution of 2 µM) was incubated at indicated temperatures for 20 min. Helicase assays (20 mM MES, pH 6.8, 1 mM DTT, 0.1 mg/ ml BSA, 10 nM 5' [^{32}P]-end labelled DNA and 1 mM ATP/ MgCl$_2$) were incubated at the same temperature used to heat-shock XPD and reactions

were initiated by the addition of heat-shocked protein. Reactions proceeded for 10 min prior to the addition of stop solution (as above).

2.8.2   STREPTAVIDIN DISPLACEMENT ASSAYS

Streptavidin displacement assays were performed at 37 ºC or 45 °C (as stated in results chapters) in helicase buffer, 0.1 mg/ ml BSA, 1 mM ATP/ MgCl$_2$, 300 nM streptavidin (Sigma) and 10 nM 5' [^{32}P]-end labelled DNA (B50) conjugated to biotin at either the 5' or 3' end. Reactions were incubated for 5 min to allow streptavidin to bind to biotinylated DNA and then free biotin (6 µM) was added to trap any streptavidin displaced during the reaction. Reactions were initiated by the addition of protein (concentrations stated in results chapters) to give a final volume of 80 µl. At appropriate time points, 10 µl aliquots of reaction mixture were removed and added to 10 µl chilled stop solution (100 mM Tris, pH 7.5, 200 mM EDTA, 1 M NaCl, 10 µM non-biotinylated B50 oligonucleotide). Samples were added to ficoll-based gel-loading dye and separated on native 12% polyacrylamide: TBE gels at 130 V for 3 h. Gels were exposed overnight and processed by phosphorimaging. The extent of streptavidin displacement was quantified using ImageGauge software: (displaced product/ total) x 100.

To determine the effect of DNA binding proteins on the ability of Hel308 to displace streptavidin from biotinylated DNA, DNA-binding proteins (final concentration 10 µM) were incubated with the DNA for 5 min prior to the addition of Hel308 (0.5 µM). Reactions proceeded as described above.

2.8.3   NUCLEASE REACTIONS

DinG nuclease activity was determined at 37 ºC in reactions comprising nuclease buffer (20 mM Tris (pH 7.5), 50 mM NaCl, 0.1 mg/ ml BSA), 5 mM MgCl$_2$ and 10 nM 5' [^{32}P]-end labelled DNA. Where indicated, ATP was also included at varying concentrations. Samples were initiated by the addition of 500 nM DinG to a final volume of 80 µl. At indicated time points, 10 µl aliquots of reaction mixture were added to 10 µl formamide-based gel-loading dye and incubated at 95 ºC for 5 min. Reactions were separated on denaturing 20% polyacrylamide: TBE gels at 95 W for 1.5 hr prior to visualisation by phosphorimaging and analysis using ImageGauge software.

To analyse the ATP-dependence of DinG nuclease activity, 10 nM 5' [^{32}P]-end labelled DNA was incubated in nuclease buffer, 0.1 mg/ml BSA, 5 mM MgCl$_2$, and indicated concentrations of ATP. Reactions were incubated at 37 °C prior to initiation by DinG. Reactions were continued for 60 min before being stopped by the addition of formamide-based gel-loading dye and separated and processed as described above.

In all cases, a control reaction was performed at 37 °C for 60 min in the absence of DinG to ensure that the assay was free from contaminant nucleases.

2.8.4 STRAND-EXCHANGE REACTIONS

Strand-exchange reactions (80 µl) were performed in 50 mM Hepes-HCl, pH 7.4, 0.1 mg/ ml BSA, 100 mM NaCl, 100 mM KCl, 5 mM ATP/ MgCl$_2$, 0.75 µM (in nucleotides) 5' [^{32}P]-end labelled ssDNA ('X50' oligonucleotide; sequence in Appendix 1, table A1.3) and 4 µM protein. Following 3 min of incubation at 60 °C, 3.75 µM (in nucleotides) dsDNA ('X26-50' annealed to 'B25' oligonucleotide) was added to initiate the reaction. At indicated time points, samples (10 µl) were added to chilled stop solution (10 mM Tris, pH 7.5, 20 mM EDTA, 100 mM NaCl, 0.5% (w/v) SDS, 1 mg/ ml proteinase K) and incubated at room temperature for 15 min to allow protein digestion. Ficoll-based gel-loading dye was added to samples, which were then separated on a native 12% (v/v) polyacrylamide: TBE gel at 130V for 3 h. Gels were exposed overnight, processed by phosphorimaging and quantified using ImageGauge software ((strand exchange product/ total) x 100).

To study the effect of additional proteins, 5' [^{32}P]-end labelled ssDNA was incubated at 60 °C for 3 min with SSB (10 µM), Alba1 (10 µM), or RadA (4 µM). Following a further 3-min incubation, dsDNA was added to initiate the reaction. Appropriate controls were included for each assay, including two end-point assays (reactions stopped 10 min after initiation), one performed in the absence of protein and a second performed in the absence of ATP/ MgCl$_2$.

2.8.5 D-LOOP FORMATION ASSAYS

D-loop reactions (50 µl) contained 5' [^{32}P]-end labelled 80mer ssDNA (3 µM in nucleotides) and 5 µM SsoRadA, RecA or Sso2452 in D-loop buffer (50 mM Hepes-HCl, pH 7.4, 1 mM DTT, 0.1 mg/ ml BSA, 100 mM NaCl, 100 mM KCl, 2 mM ATP and 15 mM MgCl$_2$). After 5 min at 60 °C (or 37 °C for RecA), the reaction was

initiated by the addition of supercoiled pUC19 dsDNA (300 μM in nucleotides) and 10 μl samples were added to ½ volume of chilled stop solution (10 mM Tris, pH 7.4, 20 mM EDTA, 100 mM NaCl, 3 % (w/v) SDS, 1 mg/ ml proteinase K and 100 mM MgCl$_2$) at specified time points (1, 5, 10, 20 and 30 min for Sso2452 and RadA; 1, 2, 4, 8, 16 min for RecA). Following a 15-min incubation at room temperature to allow for protein digestion, samples were mixed with 1/5 volume of loading dye (70 % (v/v) glycerol and 0.1 % (w/v) bromophenol blue) and analysed by electrophoresis through 0.8 % (w/v) agarose-TBE gels at 4 V/ cm for 3 h. Gels were dried onto 3 mm Whatman paper at 50 ºC for 2 h and visualised by autoradiography.

In order of addition experiments, ssDNA was preincubated with protein 1 for 5 min prior to the addition of protein 2. In the event of introducing protein 3, reactions were incubated for a further 5 min prior to addition of dsDNA to initiate the reaction. Reactions were stopped after 1, 5, 10, 20 and 30 min. Final protein concentrations were as follows: Sso2452 and RadA, 5 μM; SSB, 1 μM; Hel308, 1 μM.

To investigate the concentration-dependent effect of Sso2452 inhibition of RadA-catalysed D-loop formation, samples were incubated with increasing concentrations of Sso2452 (0.5, 1, 2, 4 and 6 μM). Reactions were initiated by dsDNA and incubated for 30 min prior to the addition of chilled stop solution.

2.9   ATPASE HYDROLYSIS REACTIONS

Malachite Green Phosphate Assay kit standard curves were produced according to the BioAssay Systems protocol. ATPase assays were performed in a final volume of 300 μl containing 20 mM Mes, pH 6.5, 1 mM DTT, 0.1 mg/ ml BSA, 100 mM KCl, and specified concentrations of protein and DNA (ΦX174 virion or RFI DNA, New England Biolabs). Reactions were incubated at the indicated assay temperature for 1 min and initiated by 1 mM ATP/ MgCl$_2$. At indicated time points, 40 μl samples were taken and immediately added to 40 μl of 0.3 M chilled perchloride acid on a 96-well plate. Samples were equilibrated to room temperature prior to the addition of malachite green (20 μl), and, following a 12-min incubation at room temperature, the absorbance at 650 nm was measured on a SpectraMAX 250 Microplate Reader (Molecular Devices). For each reaction, a blank without protein was quantified and subtracted as background from sample reactions. All experiments were carried out in triplicate.

2.10   FLUORESCENCE ANISOTROPY

2.10.1   PRINCIPLES OF FLUORESCENCE ANISOTROPY

Fluorescence anisotropy is based on the selective excitation of fluorophores by polarised light. Fluorescent molecules absorb and emit light oriented along a particular axis. These are referred to as the absorption and emission dipoles, respectively (Lakowicz, 2006). In a homogeneous solution, fluorophores are randomly oriented. Upon excitation with plane-polarised light, fluorescent molecules whose absorption dipoles lie parallel to the direction of incident light become polarised, or selectively-oriented (Lakowicz, 2006). Selective polarisation of a fraction of fluorophores within a randomly oriented population leads to partially polarised fluorescence emission (figure 2.8 B).

During this study, fluorescently-labelled DNA molecules were excited with vertically polarised light and the emission was collected through automatic polarisers (figure 2.8 A). The emission polariser can be oriented parallel (II) or perpendicular ($I_\perp$) to the plane of polarised excitation, resulting in observed intensity referred to as I_{II} or $I_\perp$, respectively. These intensity values are used to calculate the anisotropy (r), as described in equation 2 (Lakowicz, 2006).

$$\text{Equation 2:} \qquad r = \frac{I_{II} - I_\perp}{I_{II} + 2I_\perp}$$

Fluorescence anisotropy measurements were performed under "magic angle" conditions whereby the emission polariser was positioned 54.7 ° from the vertical position to increase the intensity of horizontal emission two-fold over vertical emission. "Magic angle" conditions ultimately ensure that the sum of the measured intensities of the horizontal and vertical emission equal the theoretical total fluorescence intensity. Importantly, the measured ratio between horizontally and vertically polarised light deviates from the actual ratio by a factor of G. This results from the individual polarisation characteristics of monochromators (Lakowicz, 2006). Therefore, the fluorimeter automatically calculates the G factor during each anisotropy measurement to obtain the true anisotropy value.
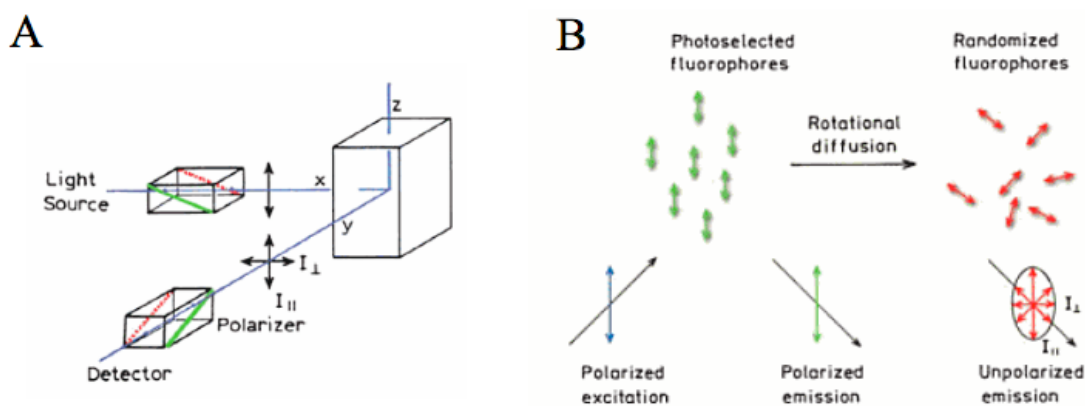
Figure 2.8: Principles of fluorescence anisotropy

(A) Vertically-polarised light is used to excite fluorescent molecules oriented parallel to the plane of light. Emission polarisers positioned to detect vertically-polarised (I_{II}) and horizontally-polarised ($I_{\perp}$) light record the intensity of the fluorescence emission. (B) Polarised light excites fluorophores oriented parallel to the plane of light resulting in emission polarisation. Rotational diffusion of fluorescent molecules creates a population of randomly oriented molecules, decreasing emission polarisation. Adapted from (Lakowicz, 2006).

Several phenomena can influence the measured anisotropy, the most common of which is rotational diffusion (figure 2.9). Rotational movement of a fluorescent molecule during its fluorescent lifetime results in a shift in the emission dipole of the molecule causing emission depolarisation, or a decrease in measured anisotropy (Lakowicz, 2006). The effects of rotational diffusion decrease upon fluorophore interaction with a macromolecule, e.g. a protein. This occurs because the rotational motion of molecules is related to its size, such that large molecules rotate more slowly than small molecules (Lakowicz, 2006). Hence, if a fluorescently-labelled oligonucleotide binds to a protein, the rate of rotation decreases with a concomitant increase in anisotropy. Since an increase in anisotropy is directly correlated to DNA:protein binding, such experiments permit the determination of DNA binding affinities of proteins.
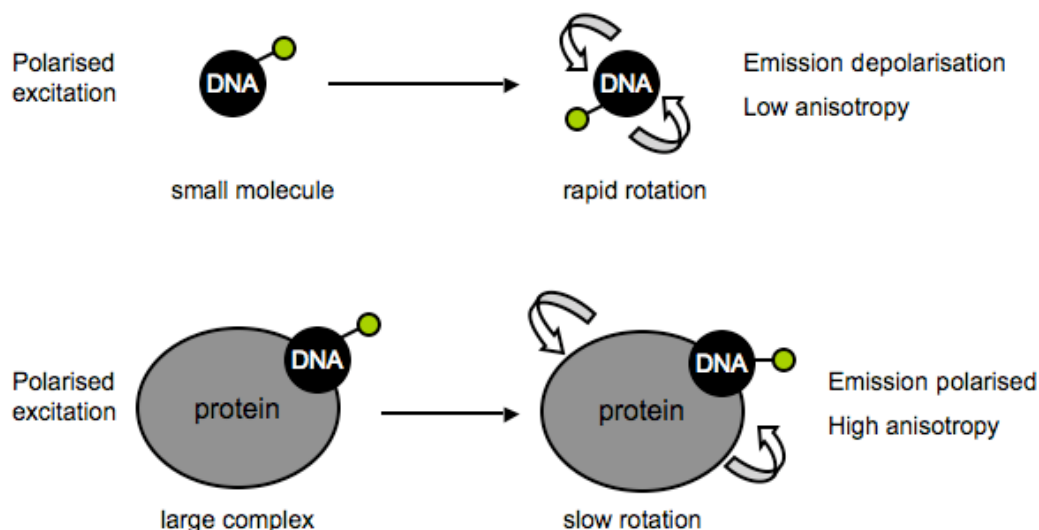
Figure 2.9: Effect of rotational diffusion on measured anisotropy

Fluorescein-labelled DNA (green circle, fluorescein) tumbles rapidly in solution resulting in emission depolarisation, and hence low anisotropy. Upon protein binding (grey circle), the larger complex tumbles more slowly in solution resulting in emission polarisation and increased anisotropy.

2.10.2 ANISOTROPY MEASUREMENTS

Anisotropy measurements were performed using a Varian Cary Eclipse fluorimeter equipped with automatic polarisers. All DNA molecules were fluorescein-labelled (Operon) and, therefore, excitation was performed at 490 nm and emission spectra collected at 535 nm. Unless otherwise stated, direct titrations were performed by incubating 20 nM fluorescein labelled DNA (Appendix 1, table A1.7) at 20 ºC in 150 μl anisotropy buffer (20 mM Hepes (Sigma), pH 7.5, 100 mM NaCl, 1 mM DTT, 0.01% (v/v) Triton-X100 (Sigma)). After 5 min of equilibration, protein was titrated into the DNA solution, and anisotropy and total fluorescence intensity (under 'magic angle' conditions (Reid et al., 2001) were measured after each protein addition. Anisotropy in the absence of protein was subtracted from each data point and data were corrected for the effect of dilution. Each titration was repeated in triplicate. Data were fitted to Equation 3, using KaleidaGraph:

Equation 3:

$$A = A_{min} + [(D + E + K_D) - \{(D + E + K_D)^2 - (4DE)\}^{1/2}](A_{max} - A_{min})/(2D)$$

(A, anisotropy; E, total protein concentration; D, total DNA concentration; A_{min}, anisotropy of free DNA; A_{max}, maximum anisotropy of the DNA-protein complex; K_D, the dissociation constant (Reid et al., 2001). Equation 3 assumes a 1:1 interaction

between protein and DNA; however, binding of multiple protein molecules to a single DNA molecule is likely and thus the equation produces an 'apparent' dissociation constant.

For competition assays, the direct titration of protein into a solution containing fluorescein-labelled DNA was performed (as described above) followed by a titration of unlabeled DNA (of the same sequence as the fluorescein-labelled DNA). Upon each addition, anisotropy and total fluorescence intensity (under 'magic angle' conditions) were measured and the data were fitted to Equation 4:

Equation 4:

$$A = -[A_{min} + [(D + E + K_D) - \{(D + E + K_D)^2 - (4DE)\}^{1/2}](A_{max} - A_{min})/(2D)]$$

(parameters as described above).

In all cases, G factors were automatically calculated to correct for differences in response to vertical and horizontal polarized light.

2.11 ELECTROPHORETIC MOBILITY SHIFT ASSAY

For DNA-binding studies, increasing concentrations of protein were titrated into binding buffer (20 mM Tris, pH 7.5, 50 mM NaCl, 1 mM EDTA, 0.1 mg/ ml BSA) containing 10 nM DNA substrate ('B50' oligonucleotide; sequence stated in Appendix 1, table A1.3) and, where indicated, 5 mM $MgCl_2$ and/ or 2 mM ATP. Binding reaction samples (15 µl final volume) were incubated at 37 ºC for 10 min and then cooled on ice. Samples were mixed with ficoll-based gel loading dye and separated on native 8% (v/v) polyacrylamide:TBE gels at 130 V for 2.5 h. Gels were exposed overnight, processed by phosphorimaging and quantified using Fuji Image Gauge software. A control reaction was performed in the absence of protein.

2.12 ANALYTICAL GEL FILTRATION

Analytical size-exclusion chromatography was performed using a Superdex 200 10-300 column (GE Healthcare) equilibrated with 50 mM Tris, pH 7.2, 150 mM NaCl, 1 mM EDTA and 1 mM DTT. To calibrate the column, proteins of known molecular weight were loaded onto the column and their elution volumes noted. The following standard proteins were used: blue dextran (2000 kDa), β-amylase (200 kDa), alcohol dehydrogenase (150 kDa), albumin (66 kDa), carbonic anhydrase (29 Da), and cytochrome C (12.4 kDa). The elution volume (Ve) of blue dextran (31.2 ml)

determined the void volume (Vo) of the column. A standard curve, of equation y = 4.8239 – 1.6546x (R = 0.99666), was obtained by plotting Ve / Vo against the logarithm of the molecular weight of the proteins (Appendix 5).

Proteins of interest (4 mg/ ml) were loaded onto the equilibrated column either alone or in complex with a 34-nucleotide ssDNA (80 µM; Appendix 1, table A1.8). In the latter case, proteins were incubated (with shaking) with DNA for 1 h at room temperature. Elution volumes were determined from the absorbance at 280 nm, and molecular masses were calculated using the standard curve generated from the elution volumes of standard proteins. In all cases, samples from each eluted peak were analysed by SDS-PAGE to confirm protein identity.

2.13   PROTEIN INTERACTIONS

Protein: protein interactions were identified, as depicted in figure 2.10, using a 1 ml HiTrap Chelating HP column (GE Healthcare) equilibrated with interaction buffer (100 mM Tris (pH 7.5), 50 mM NaCl, 20 mM imidazole, 0.5 mg/ ml ethidium bromide (Fluka)). A polyhistidine-tagged bait protein (100 µM) was incubated with an untagged target protein (100 µM), in the presence of 0.5 mg/ ml ethidium bromide, at room temperature for 1 hr.
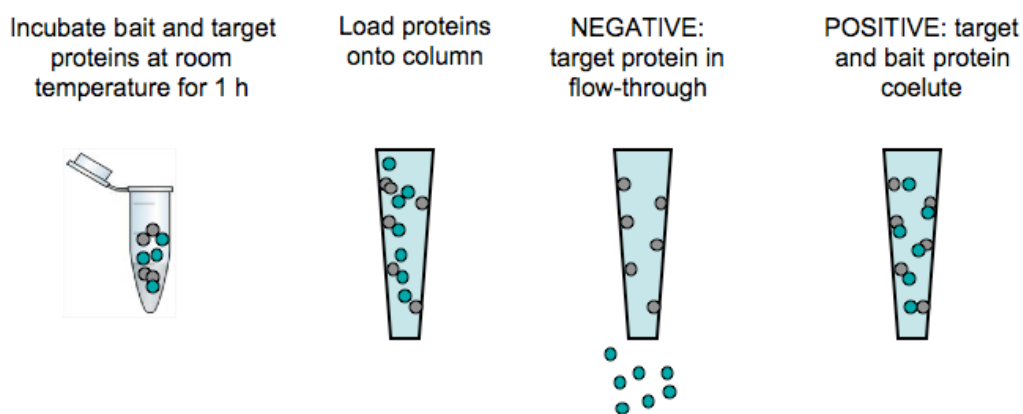
Figure 2.10: Experimental set-up for interaction study

A polyhistidine tagged bait protein (grey circle) was incubated with an untagged target protein (green circle), at equimolar concentrations, for 1 h at room temperature in the presence of ethidium bromide. The protein mixture was loaded onto a 1 ml HiTrap Chelating HP column and the column was then washed thoroughly to remove unbound material (flow-through collected). Bound proteins were subsequently eluted from the column. If the bait protein does not interact with the target protein, the latter would be found in the flow-through; alternatively, if the bait interacts with the target protein, the proteins would co-elute.

The sample was loaded onto the column in 5 column volumes (CV) of interaction buffer, and the column was subsequently washed with a further 5 CV of buffer to remove unbound materials (flow-through was collected and concentrated). Bound material was eluted with 5 CV of elution buffer (100 mM Tris, pH 7.5, 100 mM NaCl, 500 mM imidazole). The flow-though and eluted fractions were mixed with 4x NuPAGE LDS Sample Buffer (Invitrogen) and analysed by SDS-PAGE and/ or Western blotting.

2.14 WESTERN BLOTTING

Protein: protein interaction experiments revealed a potential interaction between Sso PBL2025-Hel308 and SsoHjc. After analysis by SDS-PAGE, samples were transferred to nitrocellular membrane using the iBlot Gen Transfer system (Invitrogen), which was subsequently incubated in blocking buffer (5% (w/v) milk, 0.1 % (v/v) Tween 20 in PBS buffer) for 1 h at room temperature. Membranes were rinsed in PBS buffer and then incubated overnight at 4 °C in PBS buffer containing 1% (w/v) milk, 0.1 % (v/v) Tween 20 and polyclonal antibodies against SsoHjc (diluted 1:10000 in PBS buffer; obtained from Sheep antiserum; produced by Alba Bioscience). The membrane was washed with blocking buffer (3x 10 min, room temperature) and then incubated for 30 min at room temperature with the secondary antibody (Anti-Goat, Pierce ImmunoPure® Antibody) diluted 1:200000 in PBS buffer containing 1% (w/v) milk and 0.1 % (v/v) Tween 20 (1:200000; Anti-Goat, Pierce ImmunoPure® Antibody). The membrane was washed in blocking buffer (3x 10 min) and rinsed in PBS (3x 5 min). Chemiluminescence detection was performed using SuperSignal® West Pico Chemiluminscent Substrate or Femto Maximum Sensitivity Substrate kits (Pierce). The signal was visualised using the Fuji Luminescent Image Analyser LAS-1000 and Image Reader software.

2.15 BATHOPHENANTHROLINE ASSAY

The number of irons bound to SacXPD was determined using a colorimetric assay (Lewis, 1971; Pieroni et al., 2001). Briefly, 100 µl of 50 µM protein in buffer (20 mM Tris (pH 6.5), 200 mM NaCl, 1 mM DTT, 1 mM EDTA) was mixed with 30 µl concentrated HCl and incubated at 100 °C for 15 min. A control reaction was set-up and contained buffer only. The reaction mixture was centrifuged at 13000 x g for 2 min

(Eppendorf) and 100 µl of the supernatant were taken and added to 1.3 ml 500 mM Tris (pH 8.5), freshly prepared 5% ascorbic acid (100 µl) (Fisher Scientific) and 0.1% (v/v) bathophenanthroline sulfonate (400 µl) (Sigma). Samples were thoroughly mixed and incubated for 1 h at room temperature. The number of iron bound to XPD was quantified by measuring the absorbance at 535 nm against the control reaction and the molar extinction coefficient of the ferrous iron-bathophenanthroline sulfonate complex ($22369 \; L \times mol^{-1} \times cm^{-1}$).

The volume taken for analysis was 100 µl (out of the total volume of 130 µl, i.e. 100 µl protein solution + 30 µl HCl). Given that the final reaction volume was 1.9 ml, Equation 5 was employed to calculate the total iron content (n_{Fe}), in mol, for 5000 pmol XPD (100 µl of a 50 µM solution) using the absorbance (A) at 535 nm. The amount of iron ions per XPD molecule was calculated by dividing the total iron content (obtained from Equation 5) by 5000 pmol.

Equation 5:

$$n_{Fe} = \frac{A}{22,369} \frac{mol}{L} \times 1.9 \cdot 10^{-3} \, L \times 1.3 \; \text{(for a cuvette with 1cm path length)}$$

2.16   REVERSE TRANSCRIPTASE-POLYMERASE CHAIN REACTION (RT-PCR)

2.16.1   BACKGROUND TO RT-PCR

In RT-PCR, fluorescent reporter molecules, such as SYBR® green I, are included in the reaction mixture. During amplification, SYBR® green I binds to dsDNA products and fluoresces. The BioRad iCycler measures the fluorescence intensity and records a Ct value for the reaction, which corresponds to the number of PCR cycles required for the fluorescence intensity to cross a concentration threshold defined by the iCycler software using the background intensity of the first 10 cycles. Since fluorescence intensity is directly proportional to transcript amplification, the Ct value provides a measure of transcript concentration. Specifically, the greater the Ct value, the lower the concentration of transcript (figure 2.11). Assuming the primer efficiency is 100%, transcript concentration halves with each Ct value increase of 1.

LOW TRANSCRIPT ABUNDANCE
LOW FLUORESCENCE
HIGH Ct VALUE

HIGH TRANSCRIPT ABUNDANCE
HIGH FLUORESCENCE
LOW Ct VALUE
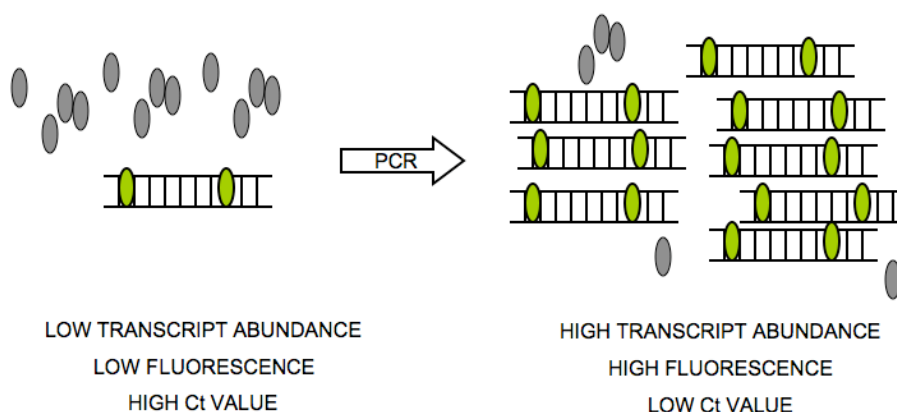
Figure 2.11: Basis of RT-PCR measurements using SYBR® Green I

SYBR® Green I fluoresces upon interaction with dsDNA: high DNA concentrations produce high fluorescent intensity. Ct values represent the point at which the fluorescent intensity exceeds a threshold limit. Thus, high DNA concentrations produce lower Ct values. SYBR® green I bound to dsDNA, green oval; Free SYBR® green I, grey oval.

2.16.2 RNA EXTRACTION FROM *S. SOLFATARICUS*

Frozen stocks of *S. solfataricus* cells (200 µl) were used to inoculate 50 ml *Sulfolobus* media. Cultures were grown overnight at 80 ºC and then transferred to 1 L *Sulfolobus* media (pre-warmed to 80 ºC; recipe stated in Appendix 4) and grown until an OD_{600} 0.2 was achieved. RNA was extracted from *S. solfataricus* cultures according to the RNeasy Mini Kit (Qiagen), with the exception that proteinase K (100 mg/ ml) was used instead of lysozyme to promote efficient cell lysis.

2.16.3 RT-PCR MEASUREMENTS

Quantitative RT-PCR was carried out using the BioRad iQ5 thermocycler using a BioRad iScript One-Step RT-PCR with SYBR® Green 1 Kit (BioRad) in accordance with the manufacturer's directions. Briefly, reactions were set-up in a 96-well plate and contained 12.5 µl SYBRmix, 0.2 µM forward primer, 0.2 µM reverse primer (gene-specific primers stated in Appendix 1, table A1.5), 0.5 µl iScript and ultrapure water to give a final volume of 24 µl. To each reaction, 1 µl RNA (100 ng) was added and plates were then covered with Microseal® (B-film) (Biorad). The following PCR programme was used: 50 ºC for 10 min (1 cycle); 95 ºC for 5 min (1 cycle); 95 ºC for 10 s, 54 ºC for 30 s, 95 ºC for 1 min, 55 ºC for 1 min (35 cycles); 55 ºC+ for 1 min (80 cycles; temperature increases 0.5 ºC/ cycle to produce the melt curve); 4 ºC

Reactions were carried out in triplicate. Ct values were measured and the ratio of gene expression in UV-irradiated over control samples was quantified as described by (Pfaffl, 2001) using gene *sso0961* as a control, as its expression has been shown to remain unchanged after UV irradiation.

The amplification efficiency of the gene-specifc primers was determined by gene amplification from neat DNA and 1/ 10, 1/ 100 and 1/ 1000 genomic DNA dilutions. Reactions included 10 μl iQ SYBR Green Supermix, 1 μM each of the forward and reverse primer and ultrapure water to give a final volume of 19 μl. To each reaction, 1 μl of DNA (of the appropriate dilution) or 1 μl ultrapure water (control) was added and plates were sealed with microseal® (Biorad). Amplification of DNA was achieved using the following PCR programme: 95 ºC for 3 min (1 cycle); 95 ºC for 30s, 52 ºC for 30 s, 72 ºC for 30 s (40 cycles); 72 ºC for 5 min (1 cycle); 95 ºC for 1 min (1 cycle); 55 ºC+ for 1 min (80 cycles; temperature increases 0.5 ºC/ cycle to produce the melt curve); 4 ºC. Primer efficiencies of 1.06, 1.00, 1.04 and 1.08 were obtained from the *radA, sso0777, sso2452* and *sso1861* primer sets, respectively.

2.17   *IN VIVO* CELL SURVIVAL ASSAYS

S. aureus cells containing a disrupted *dinG* gene were generated by Dr S. Graham (University of St Andrews), according to the TargeTron™ gene knockout methodology (Sigma) (figure 2.12). Briefly, to disrupt translation of DinG, a group II intron was mutated to include sequences that re-target the intron to the gene of interest. This was achieved by PCR, using three DinG-specific primers, IBS, EBS2, and EBS1d, and a universal primer (EBS universal) (Appendix 1, table 9). The PCR product (350 bp) was digested using *Hind*III and *Bsr*GI and ligated into the linearised vector, pNL9164. Group II introns must form a complex with reverse transcriptase (RT) to produce an RNA-protein complex, or RNP, prior to insertion within the gene of interest. Therefore, the pACD4K-C expression vector also encodes reverse transcriptase. The plasmid was transformed in *S. aureus* RN4220 host cells to enable expression of the RNA-protein complex (RNP). The RNP complex scans the genomic DNA of *S. aureus* to locate the specific insertion site and, upon target recognition, insertion of intron RNA occurs followed by reverse transcription, to generate cDNA, and DNA repair to regenerate any damaged DNA ends. Screening by PCR using gene-specific primers (Appendix 1, table A1.9: 'DinG_f KO' and 'DinG_r KO') was used to confirm the

presence of the intronic insertion within the *dinG* sequence (performed by Dr S. Graham). Finally, to prevent intron splicing, plasmids were cured from the host according to the manufacturer's protocol (Sigma).

WT and *dinG*-disrupted *S. aureus* cells were grown in TSB media (Appendix 4) to an OD_{600} of 0.65. Cells were diluted 1: 20,000 in TSB media prior to the addition of DNA damaging agents. Cultures (900 µl) were treated with 100 µl nalidixic acid, mitomycin C or 4-nitroquinoline 1-oxide (to give final concentrations of 0, 0.25, 0.5, 1, 2, 4 or 8 µM) or subjected to UV radiation (0, 10, 20, 40, 60, 80 or 100 Joules/ m^2) followed by incubation at 37 ºC with shaking (600 rpm). After 10 min, cells (10 µl) were plated on TSB agar (Appendix 4) and incubated overnight at 37 ºC. Colonies were then counted and plotted.
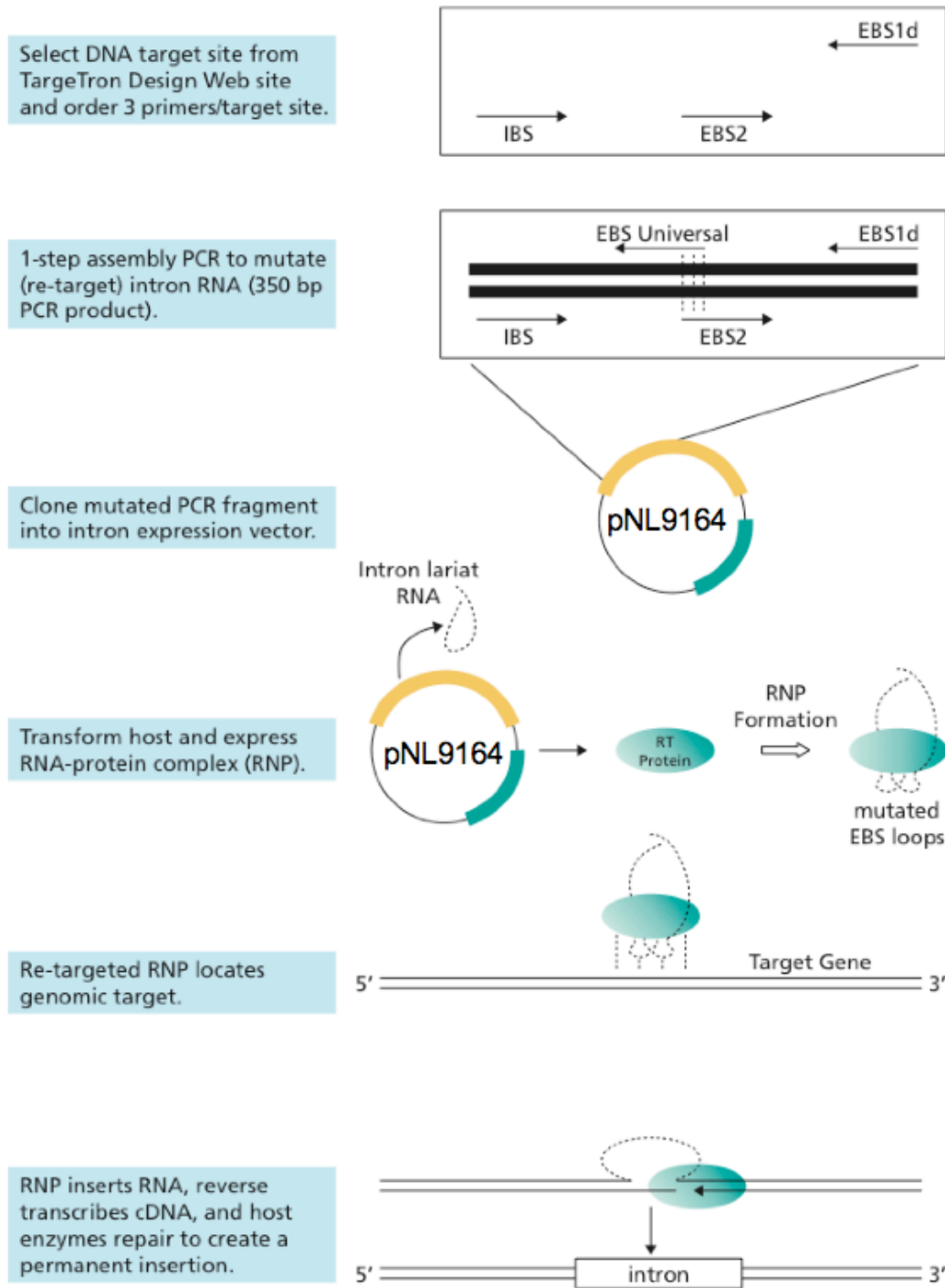
Figure 2.12: TargeTron^TM Gene Knockout System

TargeTron^TM gene knockout system (Sigma-Aldrich) methodology. Described fully in text.

CHAPTER 3

MOLECULAR BASIS OF XPD-RELATED DISEASES

3.1   INTRODUCTION

XPD (or Rad3 in *S. cerevisiae*) is a SF2 DNA helicase with 5'-3' polarity (Rudolf et al., 2006) and, in eukaryotes, forms part of the 10-subunit TFIIH (transcription factor IIH) complex. The TFIIH complex plays a central role in transcription initiation and in the nucleotide excision repair (NER) pathway where it catalyses localised unwinding around the transcription promoter site or around regions of DNA damage, respectively. The helicase activity of XPD is not required for transcription initiation, but is essential to NER. The TFIIH complex comprises a second DNA helicase, XPB, of opposite polarity to XPD and its unwinding activity is essential during both transcription initiation and NER (Tirode et al., 1999; Coin et al., 2007).

Mutations of the human *xpd* gene are associated with three genetic diseases, namely *xeroderma pigmentosum* (XP), Cockayne syndrome combined with XP (XP/ CS) and trichothiodystrophy (TTD) (figure 3.1). Despite arising from the mutation of a single gene, these conditions are associated with a wide spectrum of symptoms. XP is characterised by skin abnormalities of varying severity and a 2000-fold increased propensity to skin cancer due to defective repair of UV-induced DNA damage via the NER pathway (Lehmann, 2001). In severe cases, neurological abnormalities have been reported, resulting from premature neuronal death. XP/ CS is a multisystem disorder associated with sun sensitivity but, unlike patients with XP only, the combined XP/ CS condition does not lead to changes in skin pigmentation nor an increased risk of cancer

(Lehmann, 2001). In addition, this condition is characterised by dwarfism, and skeletal and retinal abnormalities (Lehmann, 2001). The phenotypes associated with this condition result from protein mutations that affect the transcription and NER pathways of the cell. The defining feature of TTD is sulphur-deficient brittle hair resulting from reduced levels of cysteine-rich matrix proteins. TTD patients exhibit premature ageing, mental retardation, and dwarfism but, as for XP/ CS, the condition is not associated with skin cancer (Lehmann, 2001). The phenotypes of TTD and XP are quite distinct, a likely consequence of the dual roles of TFIIH in transcription initiation and DNA repair. It has been proposed that XP is the consequence of *xpd* mutations affecting the DNA repair role of TFIIH, while those affecting transcription would result in TTD (Bootsma and Hoeijmakers, 1991).
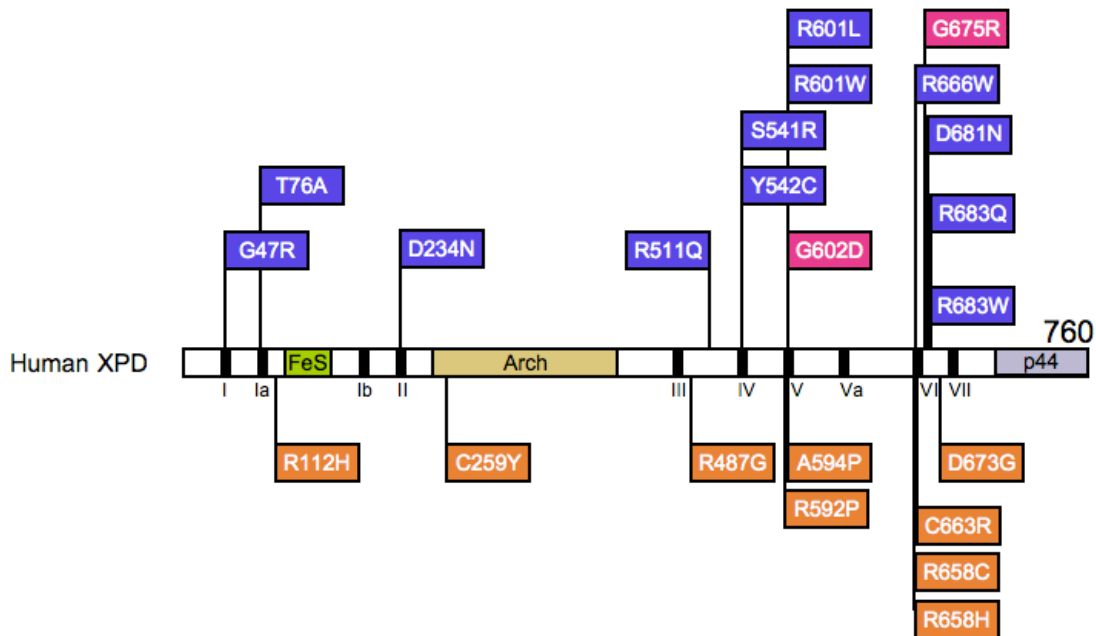
Figure 3.1: Mutations of XPD associated with human disease

Cartoon showing the domain organisation and conserved helicase motifs (black rectangles) of human XPD (760 amino acids). Residues targeted by mutation in XP (blue), XP/ CS (pink) and TTD (orange) are indicated. The FeS cluster, Arch and p44-interacting domains of human XPD are highlighted. Adapted from (Lehmann, 2001).

Within TFIIH, human XPD acts as a functional bridge linking the TFIIH core subunits with the Cdk (cyclin-dependent kinase)-activating kinase (CAK) complex (figure 3.2). This is achieved by two key interactions: the N-terminus of XPD interacts with the MAT1 subunit of the CAK complex (Busso et al., 2000; Feaver et al., 2000), while the C-terminus of XPD binds to the p44 subunit of TFIIH (Tirode et al., 1999).

Interaction of the core TFIIH subunits with the CAK complex to form the TFIIH holoenzyme is only required for transcription. There are no currently known disease-related mutations of XPD that abolish the interaction with the MAT1 protein. Instead, most of the known human mutations in XPD cluster at the C-terminus of the protein (figure 3.1). It is predicted that many of these mutations will influence the ability of XPD to bind to p44 and therefore affect the stability of the TFIIH complex and, consequently, affect transcription initiation and NER. The interaction between XPD and p44 results in a 10-fold stimulation of the helicase activity (Coin 1998); it follows that mutations that occur within the p44-interacting domain of XPD may affect the ability of the enzyme to unwind DNA during NER. Several of the C-terminal XPD mutations occur within the canonical helicase motifs (figure 3.1) and, as such, are predicted to affect the biochemistry of the protein, including its ability to bind ATP and DNA.
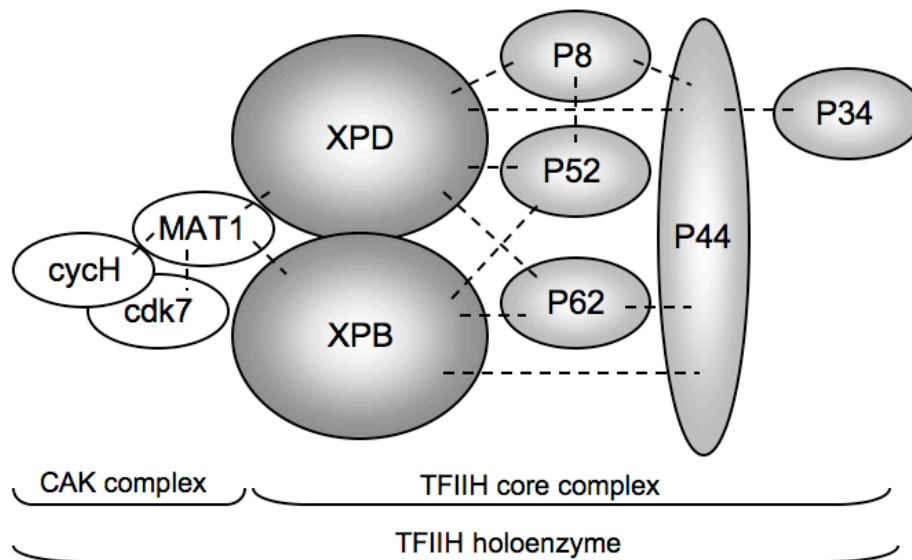
Figure 3.2: The TFIIH holoenzyme
Eukaryotic TFIIH is a 10-subunit complex composed of a core complex and the Cdk-activating kinase (CAK) complex. The core complex includes XPD and XPB, two SF2 DNA helicases of opposing polarity, P8, P52, P62, P44 and P34 and can function independently of the CAK complex in NER. During transcription initiation, the core complex must interact with the CAK complex via MAT1 to form the TFIIH holoenzyme. Dashed lines represent pair-wise interactions (Iyer et al., 1996; Busso et al., 2000; Seroz et al., 2000; Coin et al., 2006).

Homologues of XPD and XPB have been identified in most archaeal organisms (figure 3.3). In contrast to the eukaryal proteins, archaeal XPD and XPB are functionally independent proteins and do not assemble into a large TFIIH-like complex.

White and coworkers reported that XPD from the crenarchaeon *S. acidocaldarius* (SacXPD) catalysed strand separation with 5'-3' polarity, an activity that was dependent upon a stable iron-sulphur (FeS) binding domain near the N-terminus (figure 3.3) (Rudolf et al., 2006). Cluster coordination is achieved by three cysteine residues that are conserved among all archaeal and eukaryotic XPDs. Other than these cysteine residues, few other amino acids of the FeS binding domain are conserved among the archaea and eukaryotes.

In this study, the role of two conserved residues (R55 and E66) on the stability of the FeS cluster of archaeal XPD was investigated. These residues were selected prior to the elucidation of the XPD crystal structure and were anticipated to lie in relatively close proximity to the FeS cluster based on an amino acid sequence alignment (figure 3.3 B). In addition, this chapter investigates the molecular basis of disease-causing *xpd* mutations. Biochemical and structural characterisation of these mutants was performed using *xpd* from *S. acidocaldarius* and *S. tokodaii*, respectively. This chapter outlines the use of an archaeal protein to aid understanding of a more complex eukaryotic system that is involved in human disease. Unless otherwise stated, 'XPD' refers to XPD from *S. acidocaldarius* (SacXPD). The results shown in this chapter have been published in (Liu et al., 2008).

A

```
MOTIFS      I        Ia       Ib      II      III     IV       V        Va     VI       VII
SacXPD   APTGSGKT  LFVVRT  LTYPYF  DEAHNL  LMSGTL  VFPSY  VGKGKLAEG  GIPY  QAIGRAIR  DKRF
StoXPD   SPTGSGKT  IFVVRT  ITYPYF  DEAHNI  LMSGTL  VFPSY  VGRGKITEG  GVPY  QAIGRSIR  DKRY
hXPD     MPSGTGKT  IYCSRT  YSYHYL  DEAHNI  ITSGTL  FFTSY  VARGKVESG  GVPY  QAIGRSIR  DKRF
SceXPD   MPSGTGKT  IYCSRT  YSYHYL  DEAHNI  ITSGTL  FFPSY  VARGKVSEG  GIPF  QCVGRAIR  DRRF
```

SacXPD — I Ia FeS Ib II Arch III IV V Va VI VII — 551

StoXPD — I Ia FeS Ib II Arch III IV V Va VI VII — 540

hXPD — I Ia FeS Ib II Arch III IV V Va VI VII p44 — 760

SceRad3 — I Ia FeS Ib II Arch III IV V Va VI VII p44 — 778

B

```
SacXPD   -MLKLRDWQEKLKDKVIEGLRNNFLVALN------------APTGSGKTLFSLLVSLEVK 47
StoXPD   --MELRDWQKQLKDKVINAVKEGFLVALN------------SPTGSGKTLFSLLVGLNTK 46
hXPD     MKLNVDGLLVYFPYDYIYPEQFSYMRELKRTLDAKGHGVLEMPSGTGKTVSLLALIMAYQ 60
SceRad3  MKFYIDDLPVLFPYPKIYPEQYNYMCDIKKTLDVGGNSILEMPSGTGKTVSLLSLTIAYQ 60
          :  :  .        :       * :   .::   ::        *:*:***: *  :  :    :

SacXPD   -------PKVLFVVRTHNEFYPIYRDLTKIREKRNITFS--------FLVGKPSSCLYAE 92
StoXPD   -------GKVIFVVRTHNEYFPVYRDVKKIDP--SLKFS--------FVIGKPSACPFAA 89
hXPD     RAYPLEVTKLIYCSRTVPEIEKVIEELRKLLNFYEKQEGEKLPFLGLALSSRKNLCIHPE 120
SceRad3  MHYP-EHRKIIYCSRTMSEIEKALVELENLMDYRTKELGYQEDFRGLGLTSRKNLCLHPE 119
          *:::  **  *           :: ::          .       : .:  .  *  ..

SacXPD   KGAES—E---DIPCKY-------------------CELKGSIVEVKTDDS---PLSLVK 125
StoXPD   KDVDT—E---DINCKY-------------------CEIKNAIEVKIDKS----PFEKLK 126
hXPD     VTPLRFGKDVDGKCHSLTASYVRAQYQHD--TSLPHCRFYEEFDAHGREVPLPAGIYNLD 178
SceRad3  VSKERKGTVVDEKCRRMTNGQAKRKLEEDPEANVELCEYHENLYNIEVEDYLPKGVFSFE 179
              :*   *               :*  *         *.  ..

SacXPD   KLKKDGLQDKFCPYYSLLNSLYKADVIALTYPYFFIDRYREFIDIDLR-EYMIVIDEAHN 184
StoXPD   ELKENGIKEGFCPYYSLLESAKEADIITITYPYFFIEKYLSLLDINLE-EYFIVIDEAHN 18
hXPD     DLKALGRRQGWCPYFLARYSILHANVVVYSYHYLLDPKIADLVSKELARKAVVVFDEAHN 238
SceRad3  KLLKYCEEKTLCPYFIVRRMISLCNIIIYSYHYLLDPKIAERVSNEVSKDSIVIFDEAHN 239
          .*     .. ***:         .:::  :* *::  :   . :. ::  .  .:::*****
```

Figure 3.3: Alignment of XPD helicases

(A) Cartoon of the domain arrangement of *S. acidocaldarius* XPD (SacXPD), *S. tokodaii* XPD (StoXPD), human XPD (hXPD) and *S. cerevisiae* Rad3 (SceRad3). The position and sequence of conserved helicase motifs (black), and the position of motor domain 1 (cyan), motor domain 2 (pink), FeS cluster, arch domain and p44-interacting domain are indicated. Numbers to the right of each domain structure refers to the number of amino acids in each protein. (B) Amino acid sequence alignment of the FeS cluster-encoding regions of SacXPD, StoXPD, hXPD and SceRad3. The four conserved cysteine residues are highlighted in yellow, and two conserved residues (R55 and D66 in SacXPD) are shown in green. This thesis investigates the biochemical consequence of mutating residues R55 and D66 in SacXPD to alanine.

3.2 SITE-DIRECTED MUTAGENESIS OF XPD

White and coworkers (Rudolf et al., 2006) previously described the requirement of a stable FeS cluster for DNA unwinding by SacXPD. Therefore, two mutations in the *xpd* gene from *S. acidocaldarius* were generated to investigate the role of surrounding residues on FeS cluster formation. Residues R55 and D66 represent two charged and conserved amino acids located at the N-terminus of the protein in relatively close proximity to the FeS binding domain and, consequently, may affect cluster formation and/ or the role of the cluster in DNA unwinding.

Five further XPD mutants were engineered to represent human mutations associated with XP (R551Q, D861N and R683Q) or TTD (C259Y and D673G). The location of these mutations in human XPD is outlined in figure 3.1. Additional mutants associated with XP, XP/ CS or TTD were generated and characterised by Dr J. Rudolf (G602D, G675R, and R683W) and the results will be taken into account during this chapter.

Table 3.1: Mutants of XPD engineered for this study. SacXPD mutants engineered to investigate the stability of the FeS cluster or the molecular basis of disease-causing mutations in humans. The equivalent mutations of XPD from *S. tokodaii* (StoXPD) are also shown to aid interpretation of the structure of StoXPD.

Residue/ Mutation			Disease
Human	**SacXPD**	**StoXPD**	
R75	R55	R54	-
E86	D66	D65	-
C259Y	A204Y	A206	TTD
R511Q	R373Q	R365	XP
D673G	D521G	D511	TTD
D861N	D529N	D519	XP
R683Q	R531Q	R521	XP

3.3 PURIFICATION OF WT AND MUTANT XPD

The gene encoding *S. acidocaldarius* XPD was amplified from genomic DNA by Dr J. Rudolf (Rudolf et al., 2006). The gene was cloned into pET28c for expression of untagged, recombinant protein and DNA sequencing was used to confirm cloning free from base changes. The pET28c-*xpd* construct was used to generate the *xpd*

mutants by site-directed mutagenesis. WT and mutant proteins were expressed and purified (figure 3.4 and 3.5), as outlined in Materials and Methods.
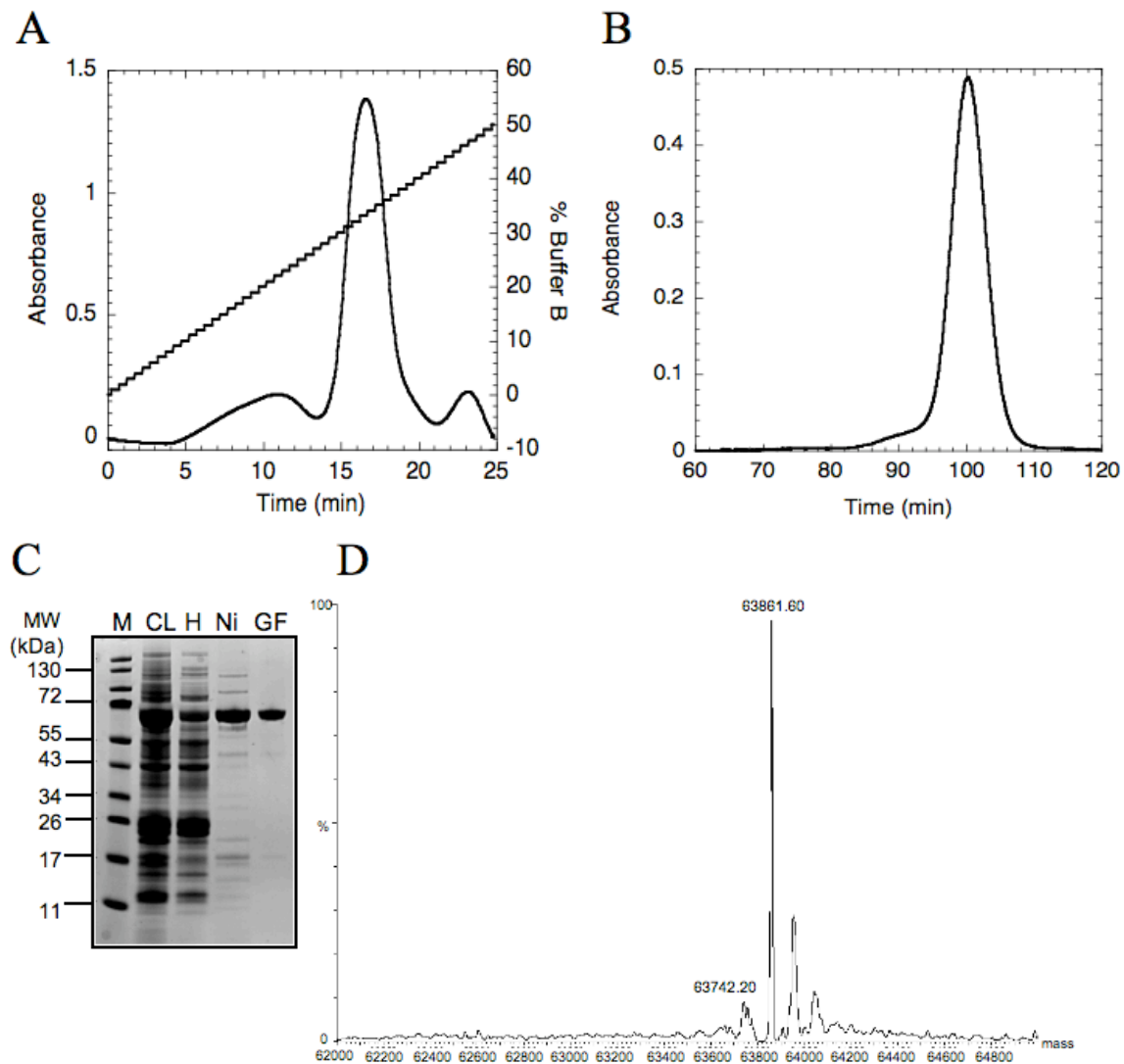
Figure 3.4: Purification of WT XPD
Chromatograms of XPD purification by cation exchange (A) and gel filtration chromatography (B). Fractions from each stage of purification were monitored by SDS-PAGE (C): M, Marker; CL, crude lysate; H, lysate after heat-treatment; Ni, XPD purified by cation exchange chromatography; GF, pure XPD after gel filtration chromatography. (D) ESI-TOF mass spectrometry of WT XPD. The major constituent had a molecular mass of 63861.6 Da (expected mass, 63837 Da).

Purified WT, R55A, D66A, R373Q, D521G and R531Q XPD were obtained at approximately 5-10 mg/ L (table 3.2). For A204Y XPD, yields approximately 5% of that achieved for WT XPD were obtained, while expression of D529N XPD under various conditions failed to produce soluble protein. This may imply that the D529N XPD mutant was not stable. Failure to purify this mutant prohibited further biochemical

analysis but it is possible that the structure of XPD may provide an explanation for its failure to express in archaea and its contribution to disease in humans.

Table 3.2: Yield of purified WT and mutant XPD. Yields of purified mutant XPD are described as a percentage of those obtained upon purification of the WT protein. D529N XPD failed to be expressed and purified.

XPD mutation	Yield (%) compared to WT
R55A	100
D66A	100
A204Y	5
R373Q	100
D521G	100
D529N	-
R531Q	100

Cation exchange chromatography involved the immobilisation of each protein to a column composed of a heparin matrix. Heparin is a negatively-charged molecular mimic of nucleic acid and should, therefore, immobilise DNA-binding proteins such as XPD. Analysis of the elution profiles from the cation exchange column revealed 3 distinct groups: WT, R55A, A204Y and R373Q (group II) eluted from the column with ~380 mM NaCl; D66A and R373Q (group III) displayed tighter binding with elution occurring at ~420 mM NaCl; R531Q (group I) bound the heparin matrix weakly and eluted from the column at ~260 mM NaCl (figure 3.5). The results suggest that the R531Q mutant would bind DNA weaker than WT. Previous findings by Dr J. Rudolf (thesis of Dr J. Rudolf) showed that the R531W mutant, which also eluted from the same cation exchange column at ~260 mM NaCl, bound to ssDNA with significantly lower affinity than WT XPD. While A204Y XPD was classed as a group II protein, it eluted from the cation exchange column with a very broad profile suggesting that its affinity for DNA may be altered in comparison to the WT protein. Mass spectrometry was employed to confirm the presence of individual mutations and to determine the intact mass of each protein (table 3.3). For each mutant, differences between expected and actual protein masses may have arisen due to the oxidation of one or more residues, e.g. methionine. Oxidation of one methionine residue would lead to a mass increase of 16 daltons.

A

B
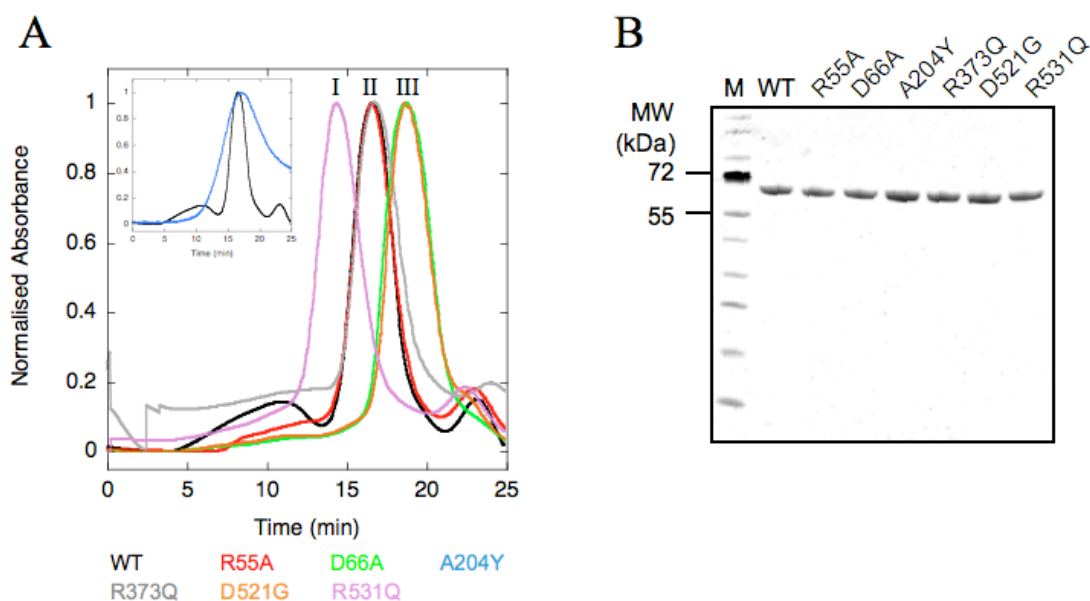
Figure 3.5: Purification of mutant XPD

(A) Chromatogram showing the elution of WT and mutant XPD from a cation exchange column composed of a heparin matrix. The elution profile of each protein has been normalised between 0 and 1. For clarity, the overlay of A204Y and WT XPD elution profiles are shown as an inset. Three profiles were observed, labelled I-III. (B) Purified WT and mutant XPD proteins were analysed by SDS-PAGE. M, Marker.

Table 3.3: Expected and actual mass of WT and mutant XPD. ESI-TOF mass spectrometry was employed to determine the intact mass of purified WT and mutant XPD proteins. Actual and expected protein masses were in good agreement.

Protein	Expected Mass (Da)	Actual Mass (Da)
WT	63837	63862
R55A	63752	63767
D66A	63793	63809
A204Y	63929	63961
R373Q	63809	63831
D521G	63779	63800
R531Q	63809	63829

3.4   SECONDARY STRUCTURE ANALYSIS OF XPD

Quantitative measurements of the secondary structure of WT and mutant XPD proteins in solution were studied by circular dichroism (CD) spectroscopy to confirm that the introduction of point mutations did not affect the structural integrity of the protein. As outlined in Materials and Methods, CD spectroscopy of WT and mutant

XPD proteins (0.5 mg/ ml) was performed in a Tris buffer containing sodium fluoride due to their low absorbance values in the far UV range (180-260 nm). Spectra were recorded between 260 nm and 180 nm and the absorbance of a buffer-only solution was also measured to enable subtraction of absorbance values not contributed by the protein.
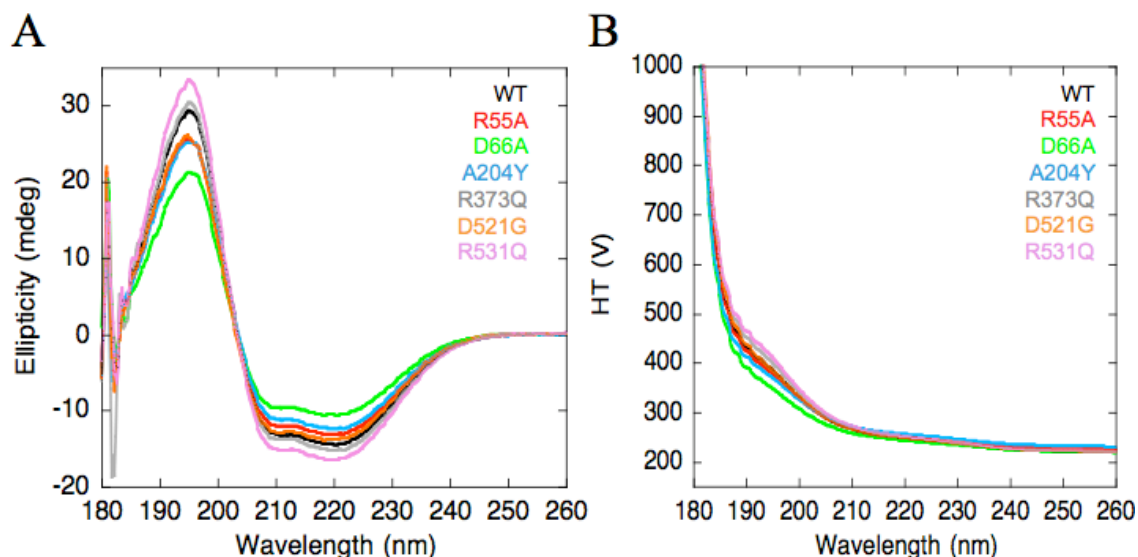
Figure 3.6: Analysis of WT and mutant XPD by CD spectroscopy

(A) CD spectra of WT and mutant XPD at far UV wavelengths of light (260-180 nm). Protein (0.5 mg/ml) was added to 300 µl of Tris buffer containing 200 mM NaF. Units of ellipticity (θ). (B) High tension (HT) voltage measurements.

The CD spectra of WT and mutant XPD in the far UV region, corrected for buffer background absorbance, indicated that the secondary structure of XPD was largely α-helical (figure 3.6 A). Confidence in the observed spectra was ensured by recording the high tension (HT) voltage (voltage applied to the photomultiplier) (figure 3.6 B). Typically, HT values (which relate to absorbance values) should remain below 1000 V (according to the JASCO J-810 Spectropolarimeter manufacturer manual).

The data suggests that all mutant XPD proteins contained regular secondary structure, as observed for the WT protein. While CD spectroscopy cannot verify the proteins were correctly folded, these results suggest that these disease-related mutations of XPD did not result in protein unfolding. Therefore, it is unlikely that the XP- and TTD-associated mutants of XPD cause disease as a result of poor protein folding or structural instability.

3.5 FeS CLUSTER IN WT AND MUTANT XPD PROTEINS

Previous work has shown that the stability of the [3Fe:4S] cluster of XPD is critical for its ability to unwind DNA (Rudolf et al., 2006). The potential for FeS cluster disruption upon generation of the XPD mutants was established by UV/ visible spectrometry and the bathophenanthroline method (Lewis, 1971; Pieroni et al., 2001), which reveals the total iron content of a protein.

3.5.1 UV/ VISIBLE ABSORBANCE SPECTRA

Purified WT and mutant XPD proteins exhibited a brown/ yellow colour, as described previously (Rudolf et al., 2006), indicating the presence of bound iron. UV/ visible spectrometry of WT and mutant XPD proteins revealed a broad shoulder between 300 nm and 500 nm (figure 3.7), a well-documented characteristic of all FeS cluster proteins (Eady et al., 1988; Coldren et al., 1997; Wollenberg et al., 2003; Rudolf et al., 2006).
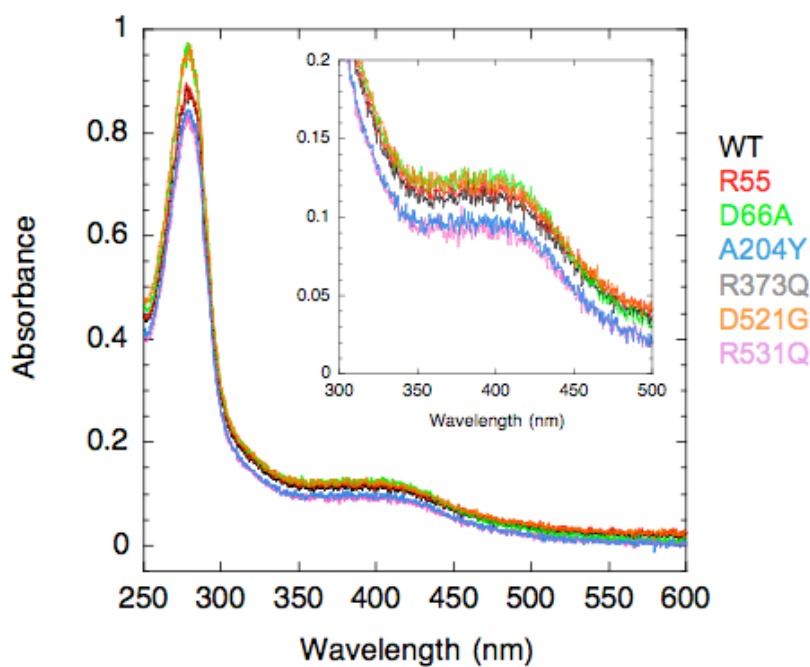
Figure 3.7: UV/ visible spectra of WT and mutant XPD

UV/ visible spectrometry revealed a broad peak centred on 400 nm, indicative of an intact iron-sulphur cluster in each XPD mutant. Inset shows zoomed-in region between 300-500 nm. Spectra coloured as indicated at the right of the plot.

3.5.2   NUMBER OF IRON IONS BOUND TO XPD MUTANTS

Further confirmation of FeS cluster stability was obtained by comparing the number of iron ions present in mutant XPD proteins to WT using the bathophenanthroline method (Pieroni et al., 2001). This colorimetric assay exploits the absorbance change that results from chelation of ferrous iron (Fe^{2+}) to 4,7-diphenyl-1,10-phenanthroline (bathophenanthroline; BA). Since the assay depends on Fe^{2+}, the experiment was performed in the presence of the strong reductant, ascorbic acid. The number of iron ions bound per XPD molecule was calculated using average absorbance values at 535 nm (based on triplicate measurements) and the molar extinction coefficient of bathophenanthroline sulfonate in complex with Fe^{2+} (22369 M^{-1} cm^{-1}). Refer to Materials and Methods for the equation used to calculate the iron content of XPD.

Table 3.4: Iron content in WT and mutant XPD. Absorbance values (at 535 nm) represent the mean of triplicate measurements (using same protein preparation but sample used for measurement was prepared three times) with standard errors shown.

Protein	Mean Absorbance	Total Iron (pmol)	Fe: XPD ratio
WT	0.1204 ± 0.001	13295	2.7
R55A	0.1268 ± 0.002	14001	2.8
D66A	0.1417 ± 0.002	15647	3.1
A204Y	0.1385 ± 0.002	15293	3.1
R373Q	0.1496 ± 0.001	16519	3.3
D521G	0.1520 ± 0.002	16784	3.4
R531Q	0.1396 ± 0.003	15415	3.1

The data reveals that WT and all mutant XPD proteins bound 3 irons (table 3.4), in agreement with previous findings (Rudolf et al., 2006).

Together, these results confirm the inclusion of a stable FeS cluster in all mutant XPD proteins studied during this investigation. Therefore, substitution of bulky, charged residues for small, uncharged residues (R55A and D66A XPD) at the N-terminus of XPD appears to have little impact on the stability of the cluster. The integrity of the FeS cluster is unlikely to contribute to the disease phenotype associated with the XPD mutants studied during this investigation.

3.6 DNA BINDING BY WT AND MUTANT XPD

The DNA binding affinity of WT and mutant XPD proteins was determined by fluorescence anisotropy. In contrast to electrophoretic mobility shift assays, fluorescence anisotropy facilitates the determination of dynamic binding equilibria of DNA-protein interactions by measuring the rotational freedom of fluorescently-labelled molecules in solution (Lundblad et al., 1996). All measurements were performed using a Cary Eclipse Fluorescence Spectrophotometer equipped with automatic polarisers under "magic angle" conditions (see Materials and Methods for details).

3.6.1 EFFECT OF TEMPERATURE, SALT CONCENTRATION AND pH ON DNA BINDING

DNA binding experiments were performed using fluorescein, a commonly used fluorophore in fluorescence anisotropy measurements. The fluorescent molecule has an absorption and emission maximum at 490 nm and 535 nm, respectively. However, the molecule has the capacity to exist in multiple ionisation states, leading to pH-dependent excitation and emission spectra around its pKa of pH 6.4. The impact of pH on XPD binding to fluorescein-labelled (Fl-) DNA was determined by measuring anisotropy values during the direct titration of XPD into a solution containing Fl-ssDNA in anisotropy buffer of various pH values at 20 ℃ (figure 3.8 B). To measure DNA binding affinity at pH 5.5, 20 mM HEPES was substituted for 20 mM MES in the anisotropy buffer. Since the rate of molecular rotation is related to absolute temperature, the effect of temperature was also tested. The experiment was performed as outlined in Materials and Methods at 20 ℃, 30 ℃ or 40 ℃ (figure 3.8 C). Finally, the impact of salt concentration on the binding of XPD to Fl-ssDNA was determined as outlined in Materials and Methods, using anisotropy buffer containing 0, 100, or 200 mM NaCl (figure 3.8 D). All experiments included 20 nM Fl-ssDNA (15-nucleotides; sequence noted in Appendix 1, table A1.7) and WT XPD.

Table 3.5 reveals that the strongest binding between Fl-ssDNA (15 nt) and XPD was achieved at 20 ℃, pH 7.06 and 100 mM NaCl. Therefore, all subsequent anisotropy measurements involving WT or mutant XPD proteins were performed under these conditions.

Table 3.5: Effect of pH, temperature and salt concentration on DNA binding by WT XPD. Apparent dissociation constant (K_D) of ssDNA binding (20 nM of a 15 nucleotide DNA) by WT XPD (titration between 0 and 10 μM) under variable pH (20 mM MES pH 5.5, 20 mM HEPES pH 7.06 or 20 mM HEPES pH 8.5), salt concentration (0, 100 mM, or 200 mM), and temperature parameters (20 ºC, 30 ºC or 40 ºC). Strongest DNA binding by XPD occurred at pH 7.06, 20 ºC, in the presence of 100 mM NaCl.

| Parameter | Experimental Variable | | | | | | | | |
| | pH | | | Salt Concentration (mM) | | | Temperature (ºC) | | |
	5.5	7.06	8.5	0	100	200	20	30	40
K_D (μM)	0.75	0.35	1.46	0.65	0.62	3.21	0.32	0.43	0.59
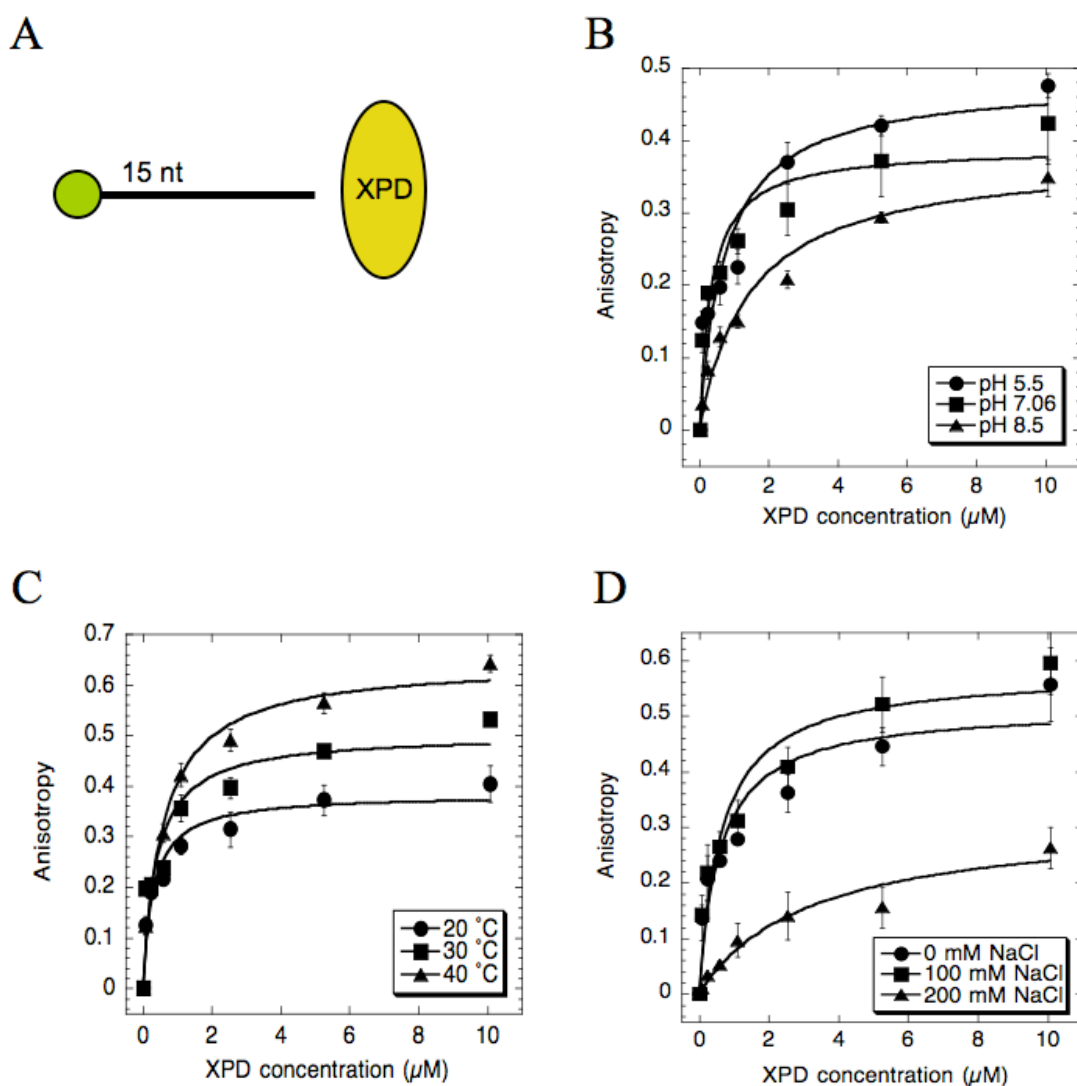
Figure 3.8: Effect of pH, temperature and salt concentration on DNA binding affinity of XPD

Fluorescence anisotropy values were recorded upon titration of WT XPD into equilibrated buffer containing 20 nM Fl-ssDNA (15 nt) (A). The experimental conditions were varied to measure the effect of buffer pH (B; at 20ºC with 100 mM NaCl), incubation temperature (C; at pH 7.06 with 100 mM NaCl), and NaCl concentration (D; at 20 ºC at pH 7.06).

3.6.2 DNA BINDING BY MUTANT XPD

Binding of XPD to Fl-ssDNA (15 nt), dsDNA (Fl-15mer annealed to unlabelled complementary 15mer) and a 5'overhang DNA (Fl-15mer annealed to unlabelled 45mer) was investigated by fluorescence anisotropy (oligonucleotide sequences noted in Appendix 1, table A1.7). Protein was titrated into an equilibrated solution containing 20 nM DNA and fluorescence anisotropy and total fluorescence

intensity values were recorded (figure 3.9 A). The data was fitted according to Equation 3 (direct titration) or Equation 4 (competitive titration) in Materials and Methods.
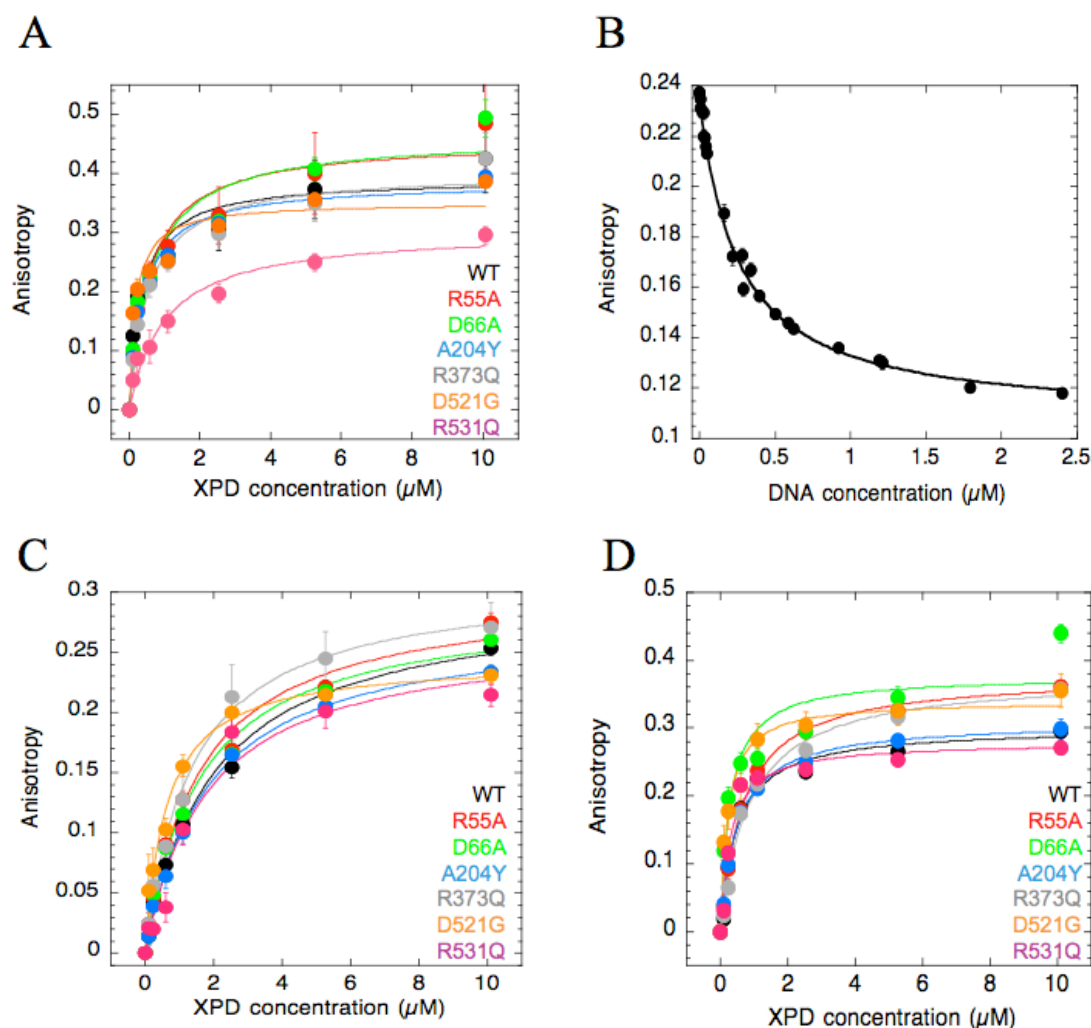
Figure 3.9: DNA binding affinity of WT and mutant XPD

Fluorescence anisotropy was recorded upon direct titration of XPD into buffer containing 20 nM Fl-ssDNA (A), dsDNA (C) and 5'overhang (D) DNA. (B) Competitive titration of unlabelled DNA into buffer containing complexes of Fl-ssDNA and WT XPD. For all plots, mean anisotropy values from triplicate measurements were plotted with standard errors shown. Curve fit equations are noted in Materials and Methods.

The results show that XPD bound to ssDNA (15 nts) and 5'overhang DNA with dissociation constants of 0.27 µM and 0.49 µM, respectively (figure 3.9 and table 3.6). Double-stranded DNA was bound approximately 4-fold weaker, with a K_D of 1.2 µM. These results are consistent with previous studies of XPD (thesis of Dr J. Rudolf). The R55A XPD mutant bound to all DNA substrates with affinities comparable to those observed for WT XPD, while the D66A XPD mutant bound ssDNA with a K_D

approximately 60% lower than the WT protein and displayed an increased affinity for a 5`-overhang substrate compared to WT XPD. Subsequent data has showed, however, that the FeS cluster was not essential for DNA binding (thesis of Dr J. Rudolf).

Interestingly, a reduced affinity for DNA was observed for the XP-associated mutants R373Q and R531Q XPD (49% and 33% of WT ssDNA affinity). The latter mutant was shown to bind weakly to the heparin matrix of the cation-exchange column during purification, in comparison to WT XPD (figure 3.5). The R531W XPD mutant also displayed reduced ssDNA binding affinity (Liu et al., 2008). In contrast, mutant D521G XPD formed a tighter association with DNA (K_D for ssDNA and dsDNA of D521G approximately 30% and 47% lower than WT, respectively). This is supported by the purification data, which shows that D521G eluted from a heparin column with a greater salt concentration compared to WT XPD (figure 3.5).

Table 3.6: DNA binding affinity of XPD. Apparent dissociation constants determined by fluorescence anisotropy upon titration of XPD into a solution of 20 nM Fl-DNA. The mean and standard errors of triplicate measurements are shown.

Protein	Equilibrium Dissociation Constants (µM)		
	ssDNA	dsDNA	5'overhang
WT	0.27 ± 0.03	1.19 ± 0.1	0.49 ± 0.04
R55A	0.27 ± 0.03	1.41 ± 0.05	0.69 ± 0.04
D66A	0.44 ± 0.04	1.03 ± 0.04	0.29 ± 0.01
A204Y	0.31 ± 0.03	1.12 ± 0.04	0.51 ± 0.02
R373Q	0.55 ± 0.02	1.37 ± 0.05	0.86 ± 0.03
D521G	0.19 ± 0.02	0.64 ± 0.02	0.22 ± 0.01
R531Q	0.82 ± 0.02	1.81 ± 0.12	0.61 ± 0.02

During all fluorescence anisotropy measurements, total fluorescence intensity was also measured. On average, the total fluorescence intensity decreased by approximately 40 % during the direct titration of XPD into Fl-ssDNA (Appendix 2). The potential for a direct interaction between the protein and fluorescein dye, as opposed to the DNA molecule, provides one possible explanation for the observed decrease in fluorescence intensity. To confirm that the K_Ds of DNA binding by XPD were not influenced by XPD/ fluorescein interactions, a competitive titration of

unlabelled DNA into complexes of WT XPD and Fl-ssDNA (of the same sequence as the unlabelled probe) was performed. Prior to the competitive titration of unlabelled DNA, the direct titration of WT XPD into a buffered solution containing Fl-ssDNA was performed to ensure adequate saturating binding conditions were achieved. Upon each addition of unlabelled DNA, fluorescence anisotropy values were recorded.

Upon titration of unlabelled DNA, anisotropy values decreased indicating that the unlabelled probe had competitively displaced Fl-ssDNA from XPD (figure 3.9 B). The results revealed an apparent K_D value of ssDNA binding to WT XPD of approximately 0.24 µM, which was in good agreement with the results of the direct titration (0.27 µM). This verifies that the anisotropy values measured by direct titration of protein into Fl-ssDNA represented a direct and specific interaction between XPD and DNA. It is possible that the observed decrease in total fluorescence intensity during direct titrations occurred as a result of local changes around the fluorescein dye, e.g. temperature or pH fluctuations.

3.7   ATPASE ACTIVITY OF WT AND MUTANT XPD

All helicases harness the free energy released upon NTP hydrolysis to translocate along DNA and catalyse strand separation. Previous studies have shown that XPD functions as a ssDNA-dependent ATPase (Rudolf et al., 2006). To monitor the ability of mutant XPD proteins to catalyse ATP hydrolysis in the presence of ss- and ds-DNA, proteins (50 nM) were incubated with 0.5 nM φX174 virion ssDNA or φX174 RFI dsDNA at 45 ºC. Samples were taken across a 5-min time course and added to malachite green, a reagent that reacts with free phosphate to produce a colour change that can be monitored by the absorbance at 650 nm. The change in absorbance over time can be directly related to the amount of free phosphate available in the sample, providing a measure of the rate of ATP hydrolysis. Results are shown in figure 3.10 and figure 3.7.

R55A and D66A XPD catalysed ssDNA-dependent ATP hydrolysis with rates comparable to those displayed by WT XPD (550 pmole phosphate released/ min/ pmol XPD). These results suggest that mutation of conserved residues located near to the FeS cluster of XPD has no impact on the ability of the enzyme to hydrolyse ATP.

While mutant A204Y catalysed ATP hydrolysis at a rate similar to that measured for WT XPD, mutant R373Q hydrolysed ATP 54% slower than WT, which

likely reflects its reduced affinity for ssDNA (table 3.6). A significant increase in the rate of ssDNA-dependent nucleotide hydrolysis was observed in the presence of D521G and R531Q XPD in comparison to WT XPD (127% and 155%, respectively). The enhanced ATPase activity of mutants D521G and R531Q may imply that the ATPase and helicase activities of these mutant proteins have become uncoupled. Alternatively, the increased ATPase activity of mutant D521G XPD may directly result from its higher affinity for DNA compared to WT XPD (figure 3.8).

In agreement with previously reported data (Rudolf *et al.*, 2006), WT and all mutant XPD proteins failed to catalyse ATP hydrolysis in the presence of dsDNA.
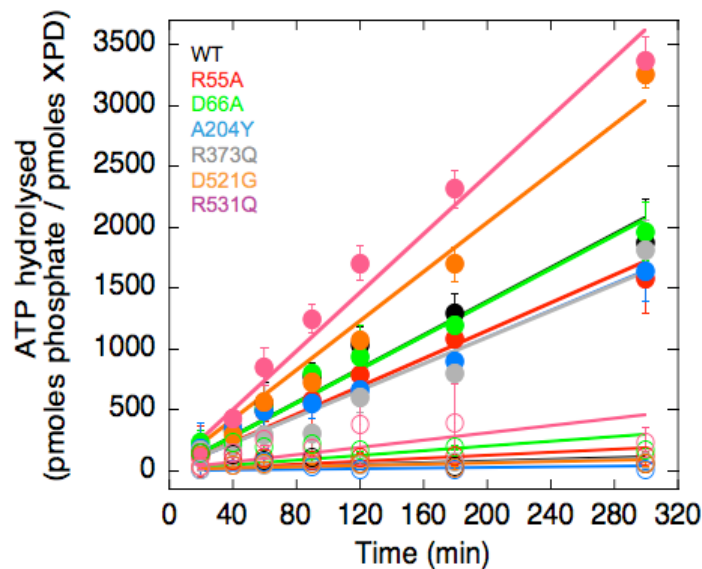
Figure 3.10: ATPase activity of WT and mutant XPD

The ATPase activity of XPD was measured by incubating 50 nM protein with 0.5 nM φX174 virion ssDNA or φX174 RFI dsDNA at 45 ºC in a final volume of 240 µl. Reactions were initiated by addition of 1 mM ATP/MgCl₂. Samples were taken at indicated time points and added to 0.3 M PCA to stop the reaction. The level of free phosphate was quantified by measuring the absorbance at 650 nm. Filled circles, ssDNA; open circles, dsDNA. Data represent the mean of triplicate experiments.

Table 3.7: Rate of ATP hydrolysis by WT and mutant XPD. Rates of ATP hydrolysis are expressed as pmole phosphate.min^{-1}.pmol XPD^{-1} and represent the mean of triplicate measurements with standard errors shown.

Protein	Rate of ATP hydrolysis (pmole phosphate / min/ pmol XPD)	
	ssDNA	dsDNA
WT	551 ± 43	82 ± 12
R55A	494 ± 37	85 ± 14
D66A	504 ± 44	98 ± 21
A204Y	488 ± 40	67 ± 10
R373Q	297 ± 36	53 ± 13
D521G	699 ± 55	57 ± 15
R531Q	855 ± 69	100 ± 12

3.8 HELICASE ACTIVITY OF MUTANT XPD PROTEINS

White and coworkers (Rudolf et al., 2006) reported that XPD from *S. acidocaldarius* unwinds DNA in an ATP-dependent manner with 5'-3' polarity. Such activity has also been shown for homologous proteins human XPD (Weber et al., 1988), human FancJ (Cantor et al., 2004), and *S. cerevisiae* Rad3 (Sung et al., 1987). This activity is dependent upon a stable FeS cluster (Rudolf et al., 2006). Sequence alignments of archaeal and eukaryal XPD proteins reveal few conserved residues located in close proximity to the FeS cluster that could influence its stability. R55 and D66 represent two of the few conserved residues located at the N-terminus of XPD that may influence cluster stability. However, the R55A and D66A mutants unwound DNA with rates comparable to the WT protein (figure 3.11) and therefore are unlikely to play a role in stabilising the cluster.

Previous studies have shown that several disease-associated mutations of human XPD inactivate the helicase activity of the enzyme. For instance, mutations located at the N-terminus (R112H and A349P) and C-terminus (G602D and R683W) of human XPD represent inactive helicases (Broughton et al., 1995; Berneburg et al., 2000; Dubaele et al., 2003; Levran et al., 2005). In this light, the helicase activity of XP- and TDD-associated mutations in SacXPD was investigated.
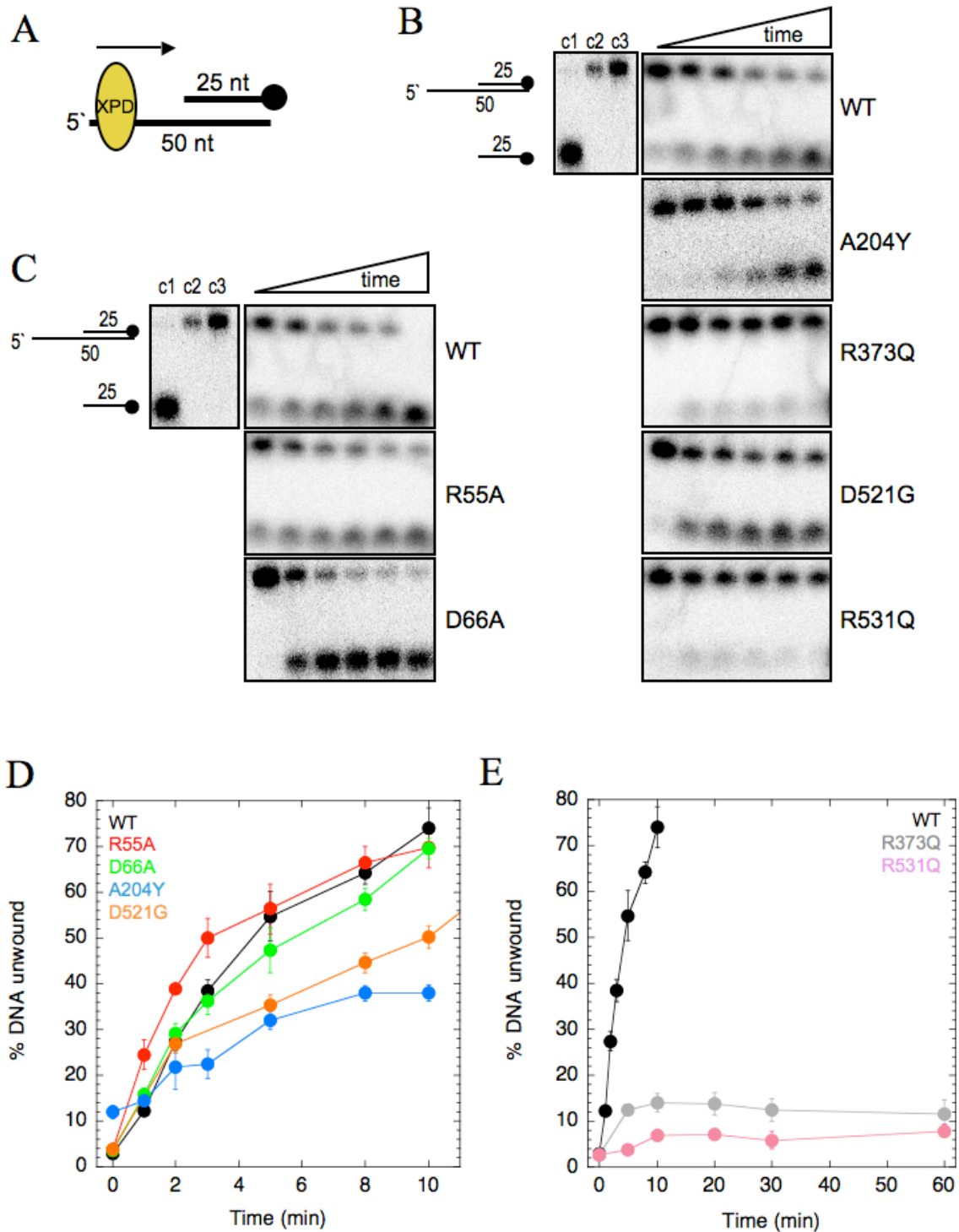
Figure 3.11: Helicase activity of WT and mutant XPD

(A) 5'overhang DNA substrate used for helicase assays. The helicase activities of disease-associated mutants (B) and R55A and D66A XPD (C) (200 nM) were determined at 45 ºC as described in Materials and Methods. For R373Q- and R531Q-XPD, samples were taken at 0, 5, 10, 20, 30, 60 min; for all remaining proteins, samples were taken at 0, 1, 2, 3, 5, 8, 10 min. c1, size marker; c2, end-point reaction without protein; c3, end-point reaction without ATP/ MgCl$_2$. ImageGauge software was used to quantify the results (D-E). All data shown are representative of a minimum of three repeat experiments using different protein preparations.

After 10 min, WT, A204Y and D521G XPD had unwound the 5' overhang DNA substrate (figure 3.11). In contrast, mutants R373Q and R531Q represent inactive helicases (no unwinding observed after 60 min). This result is consistent with the reduced affinity exhibited by both proteins for ssDNA. Mutant R531Q hydrolysed ATP at a rate 155% greater than WT XPD, which may reflect an uncoupling of ATPase and DNA unwinding activities. Previous studies have also shown that the R531W XPD mutant does not unwind DNA (Liu et al., 2008).

3.9 STRUCTURE OF XPD AND THE IMPLICATION OF DISEASE-CAUSING MUTATIONS

Biochemical and structural characterisation of eukaryal XPD faces several challenges relating to its intrinsic complexity as part of the TFIIH complex (figure 3.2). Archaeal and eukaryal XPD share 25% sequence identity, including the seven canonical helicase motifs, which renders it possible to use the simplified, functionally independent archaeal enzyme to gain greater insight into human XPD. The structure of XPD from *S. tokodaii* (StoXPD) was solved to 2.25 Å by the Scottish Structural Proteomics Facility at the University of St Andrews (Liu et al., 2008). However, the FeS cluster was not visible from the electron density. The structures of XPD from *S. acidocaldarius* (SacXPD) and *Thermoplasma acidophilum* (TacXPD) have also been solved (Fan et al., 2008; Wolski et al., 2008), and reveal clear density corresponding to the FeS cluster. The overall features of the XPD structure will be presented here, largely focusing on the *S. tokodaii* structure, to highlight key features of the protein that may explain the phenotypes associated with XP, XP/ CS and TTD.

3.9.1 STRUCTURE OF ARCHAEAL XPD

The structures of archaeal XPD reveal a four-domain protein, consisting of two motor domains, one FeS cluster binding domain, and an Arch domain (Fan et al., 2008; Liu et al., 2008; Wolski et al., 2008). The two motor domains are reminiscent of the Rad51/ RadA motor domains (Ariza et al., 2005), with a topology consisting of a central seven-stranded parallel β-sheet flanked by α-helices (figure 3.12). The canonical helicase motifs are housed within these two motor domains, with motif I (which includes the Walker A box) positioned at the interface of the two domains and represents the ATP binding site. In the StoXPD structure, a phosphate ion was observed

at this position (Liu et al., 2008). In addition to motif I, motifs II, III, V and VI surround the ATP binding site. This is consistent with previously reported roles for these motifs in nucleotide binding and hydrolysis (Tuteja and Tuteja, 2004). Motifs Ia, Ib, IV, V and Va are positioned towards the upper half of motor domains 1 and 2. Motifs Ia, IV and V have been previously implicated in ssDNA binding (Tuteja and Tuteja, 2004). To investigate this further, the crystal structure of archaeal Hel308, a SF2 DNA helicase, in complex with DNA (Buttner et al., 2007) was used to superimpose the DNA molecule onto the XPD structure. This was possible by aligning the motor domains of the two proteins, and provided an insight into the expected path taken by ssDNA through the protein (figure 3.12). As expected, the modelled oligonucleotide was positioned across the top of the two motor domains, in close proximity to motifs Ia, Ib, IV, V and Va.

The Arch domain, bridging the two motor domains, is composed of a four-stranded antiparallel β-sheet and four α-helices (Liu et al., 2008). One of the helices has an extended loop that interacts with the FeS binding domain (Fan et al., 2008). This topology failed to match any other fold in the protein databank, implying that it represents a novel fold. The position of the Arch domain, together with the significant increase in size of the domain in eukaryotic XPD species, suggests that it may be involved in linking XPD to the TFIIH complex.

The FeS binding domain was revealed by the StoXPD and TacXPD structures (Fan et al., 2008; Wolski et al., 2008) and houses the cysteine residues responsible for coordinating the FeS cluster (figure 3.12). The domain is an insertion within motor domain 1 and is composed of four α-helices connected by loops. There is some controversy in the field relating to the number of cysteine residues involved in iron coordination within the cluster: while White and coworkers report the requirement for three cysteine residues (Liu et al., 2008), Tainer and coworkers describe the involvement of four cysteine residues (Fan et al., 2008). A recent report has proposed that the FeS cluster of XPD plays a crucial role in strand displacement by interacting with the DNA at the ss/ dsDNA junction (Pugh et al., 2008). This is consistent with the structure of XPD in which the FeS cluster lines the likely path of ssDNA (figure 3.12), and with previous claims that the FeS cluster is essential for XPD helicase activity (Rudolf et al., 2006). An electrostatic representative of archaeal XPD reveals a highly basic surface lining the channel formed between the Arch domain and motor domain 1 (figure 3.13), through which ssDNA is expected to pass. An isolated, highly basic

region was also observed on the back of domain 1, providing a region on to which the displaced 3' end may dock.
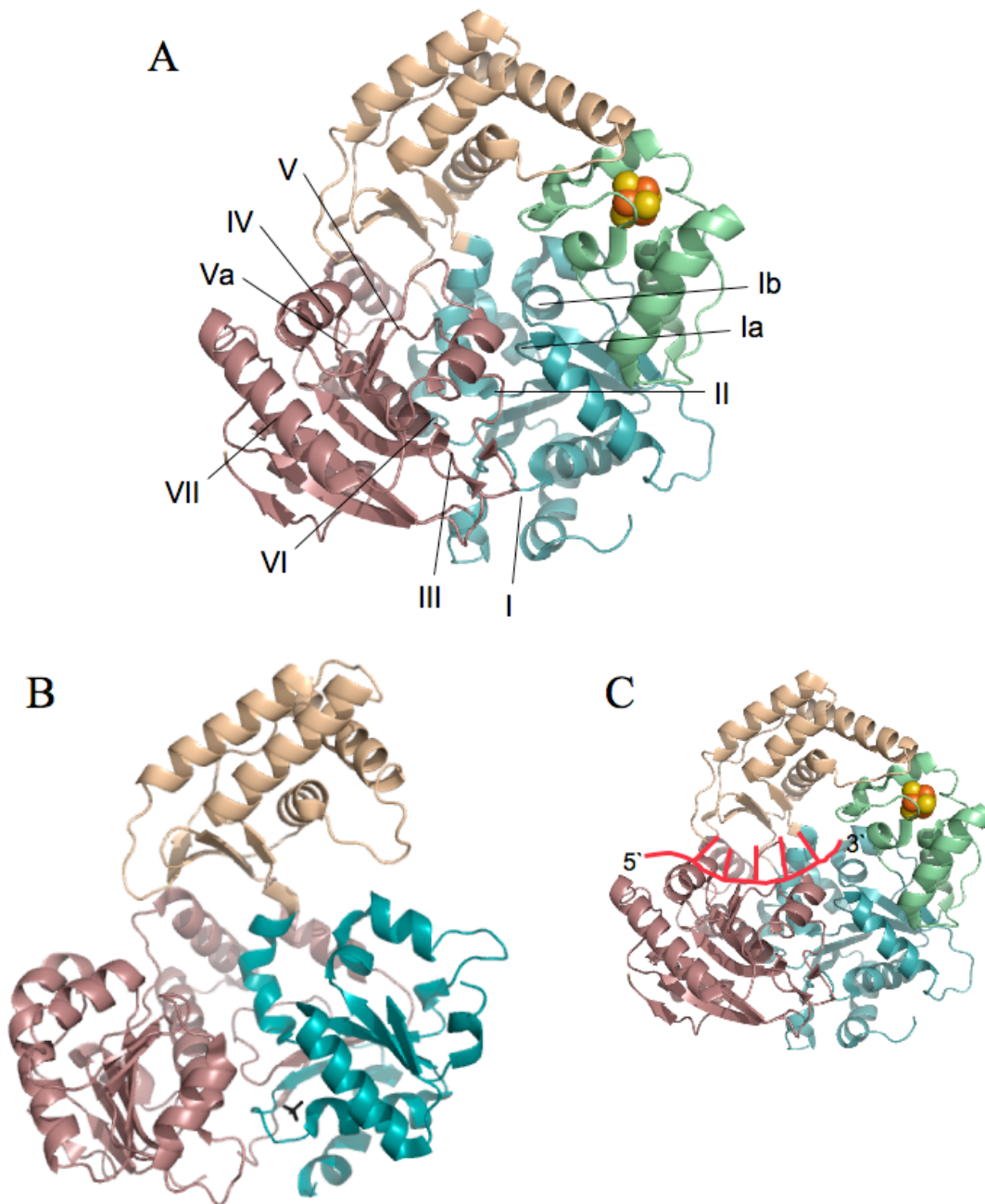
Figure 3.12: Structure of archaeal XPD

Overview of the structure of (A) SacXPD (PDB code: *3CRV*), coloured as follows: Motor domain 1, cyan; motor domain 2, salmon; Arch domain, wheat; FeS domain, green. FeS cluster shown in space-fill. The locations of the canonical helicase motifs are shown. (B) Structure of StoXPD (PDB code: 2VL7), coloured as SacXPD. The FeS domain was not visible from the electron density map. Black sticks represent a phosphate ion and likely corresponds to the position of ADP/ ATP binding. (C) SacXPD structure revealing the expected path of ssDNA. The DNA molecule was superimposed onto the XPD structure by aligning motor domains 1 and 2 of XPD with the motor domains of Hel308 in complex with DNA (PDB code: *2p6r*). Images produced using PyMOL.
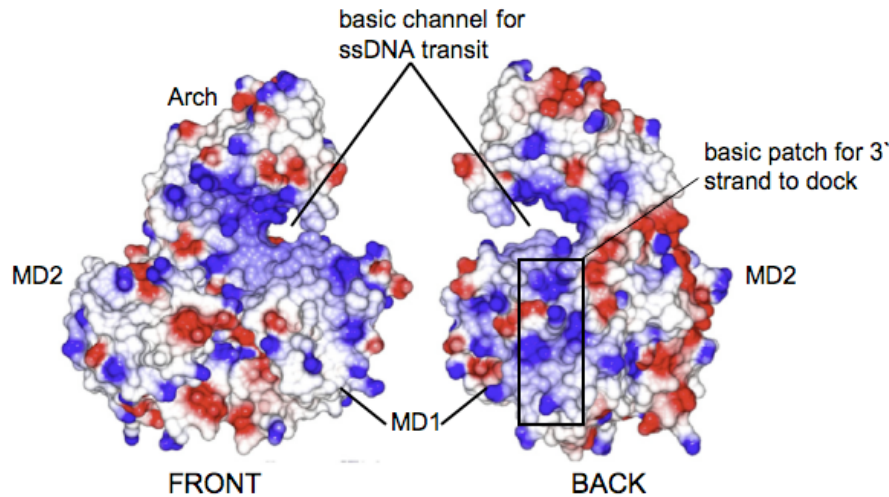
Figure 3.13: Electrostatic view of archaeal XPD

Electrostatic representation of the front and back of StoXPD with basic, acidic and neutral residues shown in blue, red and white, respectively. Motor domains 1 ('MD1') and 2 ('MD2') and the Arch domain ('Arch') are highlighted together with the two basic patches expected to bind DNA. Images produced using PyMOL.

3.9.2 STRUCTURAL BASIS FOR XP AND XP/ CS-CAUSING XPD MUTATIONS

The cartoon representation of the XPD domain structure in figure 3.14 shows that the majority of XP-related mutations arise within conserved helicase motifs. The canonical helicase motifs have been implicated in ATP and DNA binding and mutation of residues within these regions is likely to impact on the helicase activity of XPD. Residues mutated in human XPD giving rise to XP include T76A, R511Q, S541R, Y542C, D681N and R683 Q/W (residues T60, R365, S393, Y394, D519, R521 in StoXPD, respectively). The structure of archaeal XPD reveals that these residues are positioned along the top of the motor domains and, therefore, may contribute to DNA binding (figure 3.14). The electron density map for StoXPD did not permit observation of residue R601 (StoXPD residue R437), a commonly mutated residue in XP patients, but it is also predicted to lie along the top of motor domain 2.

The XP-causing mutations of XPD investigated during this study included R511Q (StoXPD, R365), D861N (StoXPD, D519), and R683Q (StoXPD, R521). The biochemical characterisation of the R511Q and R683Q XPD mutants revealed a reduced affinity for DNA and a loss of helicase activity. In the case of the R683Q mutant, an enhanced ATPase activity was also observed suggesting that the ATPase and helicase activities were uncoupled in this particular protein. The structural data supports these

findings indicating that mutation of these basic residues, which are positioned to interact with DNA, would negatively affect the ability of the enzyme to bind and, in turn, unwind DNA. The D861N human XPD mutation was reproduced in SacXPD but failed to express in *E. coli* cells. The structure of XPD reveals that this residue lies along the expected path of ssDNA and its substitution for a basic residue would likely impair DNA binding. However, the structure fails to provide an appropriate explanation for the apparent instability of this mutant in comparison to the WT protein.

Four mutations of human XPD that give rise to the combined XP/ CS phenotype are currently known, including G47R, G602D, R666W, and G675R (StoXPD residues G38, A438, R507 and V513, respectively). Although this particular study did not characterise any XP/ CS-associated XPD mutations, it is striking to note that this combined disorder often results from mutation of glycine residues. Glycine residues provide structurally flexibility and are often found within non-conserved flexible loop regions. Thus, it is unlikely that these mutations have an impact on XPD biochemistry. Intriguingly, cells from XP/ CS patients possessed low concentrations of TFIIH and exhibited an enhanced level of cleavage of undamaged DNA (van Hoffen et al., 1999). This was consistent with the results of a study involving a mouse model harbouring the G602D mutation (Andressoo et al., 2006). Together, these findings suggest that XP/ CS-associated mutations of XPD may affect the stability of XPD and, in turn, the stability of the TFIIH complex. The ultimate consequence of this would be perturbed transcriptional and DNA repair functions, typically associated with patients with the XP/ CS condition.
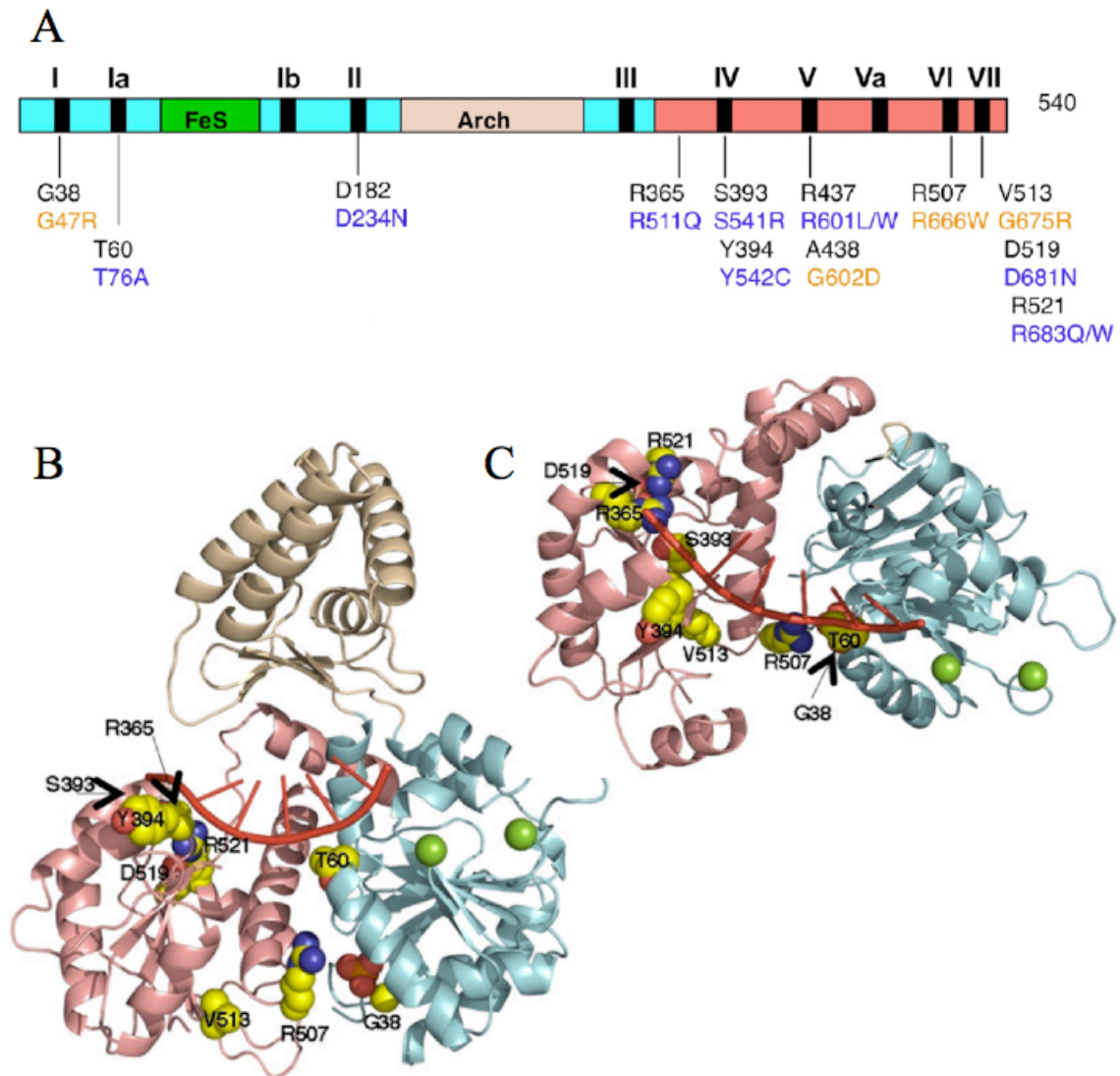
Figure 3.14: XP- and XP/ CS-causing mutations of XPD

(A) Cartoon representation of the XPD domain organisation. Motor domain 1, cyan; Arch domain, wheat; motor domain 2, salmon; helicase motifs, black. XP and XP/ CS mutations are shown in black (StoXPD), with the equivalent human mutation shown below in blue (XP) or orange (XP/ CS). (B) XP- and XP/ CS-causing mutations (space-filled spheres) are highlighted on the StoXPD crystal structure. The expected path of ssDNA (red) is indicated. The structure was rotated by 90 ºC (right) to reveal the close proximity of XP-causing mutations to ssDNA. Green spheres, Fe ions. Reproduced from (Liu et al., 2008).

3.9.3 STRUCTURAL BASIS FOR TTD-CAUSING XPD MUTATIONS

There are 12 known XPD mutations that cause TTD. However, only two of these residues are conserved in archaeal XPD. The structure, shown in figure 3.15, reveals the position of the following TTD-causing mutations: R112H (Sto, K84), C259Y (Sto, A206), R592P (Sto, G428), A549P (Sto, Y430), R616P (Sto, E454), R658 H/C (Sto, I496), C663R (Sto, T501), and D673G (Sto, D511). White and coworkers

(Rudolf et al., 2006) previously reported that mutagenesis of SacXPD to mimic the R112H human XPD mutation affected the stability of the FeS cluster leading to the loss of helicase activity.

The structure of StoXPD indicates that residue A206, along with several other residues, projects into the centre of the helical bundle that forms the Arch domain. The close association of these residues suggests that substitution of a small, uncharged alanine for a bulky tyrosine residue will compromise the stability of the Arch domain. The low yields obtained of mutant A204Y (SacXPD) obtained during purification strengthen this proposal. Furthermore, the heat-stability of WT and A204Y XPD were compared by incubating each protein for 20 min at increasing temperatures prior to assessing its ability to unwind a 5'overhang DNA substrate. The results show that the archaeal equivalent of the C259Y human mutation exhibited temperature-sensitivity (figure 3.15 C), which supports the proposal that the C259Y mutation in humans affects the intrinsic stability of the XPD structure. The potential involvement of the Arch domain in mediating protein: protein interactions with other subunits of the TFIIH complex means that the C259Y mutation may also affect the stability of the TFIIH complex.

The remaining TTD-causing mutations, including G428, Y430, E454 and D511 (StoXPD residues), cluster in the lower portion of motor domain 2 at the C-terminus of XPD. These residues are not positioned close to the Walker A or B motifs, excluding the possibility that they impinge on the ATP binding and hydrolysis functions of the enzyme, and they do not lie close to the expected path of ssDNA. The human TTD-causing D673G mutation (D521 SacXPD; D511 StoXPD) was investigated by mutation of SacXPD. This mutant displayed enhanced ATPase activity and DNA binding affinity (2.2 fold stronger dsDNA binding and 1.4 fold tighter ssDNA binding) compared to the WT enzyme.

The C-terminus of XPD is involved in interacting with p44, a subunit of the TFIIH complex, and important for modulation of the XPD helicase activity (Fan et al., 2008). Previous reports have shown that the R616P mutation negatively affects the ability of XPD to interact with p44 (Dubaele et al., 2003). Although further investigation would be required, it is possible that these XPD mutations act by destabilising the XPD interaction with p44. The absence of p44 in archaea could explain the lack of conservation of these disease-causing residues and the absence of any

discernible phenotype associated with archaeal XPD mutations that mimic those that cause TTD in humans.
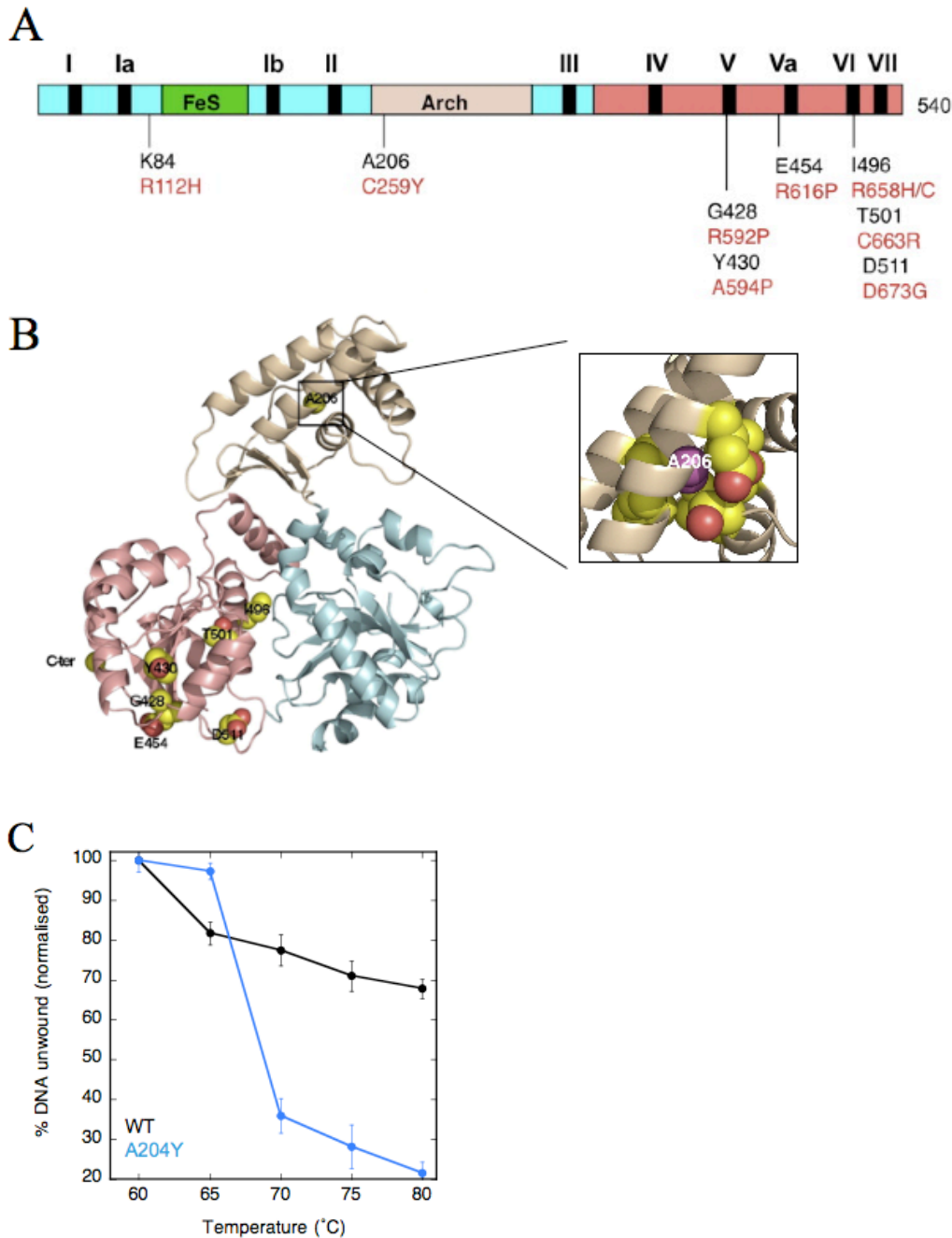
Figure 3.15: TTD-causing mutations of XPD

(A) Cartoon representation of StoXPD, with the positions of TTD-causing mutations indicated in black (equivalent human mutation shown below in red). (B) Structure of StoXPD showing positions of TTD-causing mutations (space-filled spheres). Inset, zoomed-in view of A206 (purple) and surrounding residues (coloured according to atom type). (C) Helicase activities of WT and A204Y SacXPD at various temperatures. Proteins were incubated at the specified temperature for 20 min prior to the addition of 10 nM 5'overhang DNA. Reactions were incubated for 10 min prior to separation on acrylamide gels. Plot shows the mean of triplicate results with standard errors. Reproduced from (Liu et al., 2008).

3.10   DISCUSSION AND CONCLUDING REMARKS

3.10.1   DETERMINANTS OF FeS CLUSTER STABILITY

The helicase activity of archaeal XPD and its homologues, including human FancJ, *S. cerevisiae* Rad3 and *E. coli* DinG, is dependent upon a stable FeS cluster (Rudolf et al., 2006; Pugh et al., 2008; Ren et al., 2009). Substitution of the cysteine residues involved in liganding the FeS cluster results in disruption of the cluster and, in turn, abrogation of helicase activity (Rudolf et al., 2006). Apart from the cysteine residues, few conserved amino acids are found within close proximity to the FeS cluster raising the question of how the structural integrity of the cluster is maintained. This study investigated the influence of conserved residues R55 and D66 of SacXPD on the stability of the FeS cluster by mutating them to small, uncharged alanine residues. These mutants displayed comparable activity to the WT enzyme in terms of its ATPase and helicase activities, while the D66A mutant bound DNA with a lower K_D compared to the WT protein. Together, the data suggests that the conservation of these residues across several species did not reflect a role in FeS cluster stability Table 3.8). It is important to note that these residues were selected prior to the elucidation of the XPD structure, which has subsequently revealed that the cysteine residues involved in FeS cluster coordination are not located in close proximity to residues R55 and D66.

Table 3.8: Biochemistry of R55A and D66A SacXPD. hXPD, human XPD; SacXPD, *S. acidocaldarius* XPD; StoXPD, *S. tokodaii* XPD.

Mutation/ Residue			FeS cluster	Compared to WT (%)		
hXPD	SacXPD	StoXPD		DNA affinity	ATPase	Helicase
R75	R55A	R54	Stable	100	90	~100
E86	D66A	D65	Stable	61	92	~100

The role of the FeS cluster still remains elusive. The inclusion of FeS clusters in helicases that participate in different aspects of DNA metabolism implies that this motif may not be involved in DNA damage recognition, as previously suggested (Chepanoske et al., 2000; Boal et al., 2005). It is more likely that the FeS cluster occupies a structural role within a protein, and, in the case of XPD, appears to be ideally positioned to unzip duplex DNA (figure 3.16). A role for FeS clusters in DNA unwinding is analogous to that described for the β-hairpin loop of the bacterial UvrB

(Theis et al., 1999; Skorvaga et al., 2002) and archaeal Hel308 DNA helicases (Richards et al., 2008b).
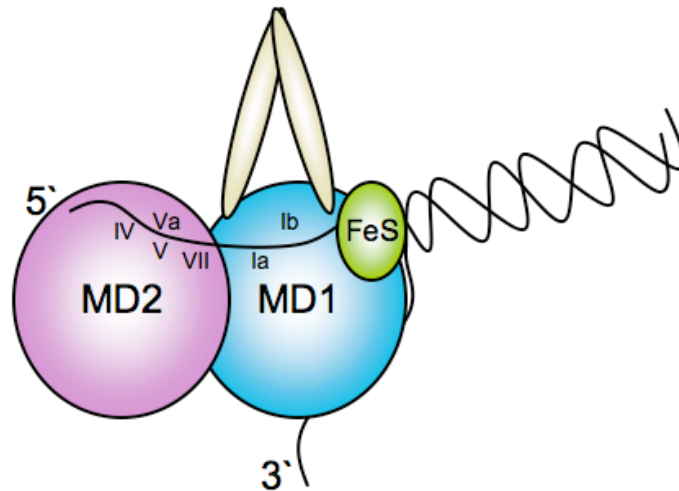
Figure 3.16: FeS cluster as a DNA "unzipper"

Duplex DNA enters XPD via motor domain 1 (MD1) where the two strands may be separated by the FeS domain. As the displaced 5' strand threads through the cleft between the Arch domain (yellow oblongs) and motor domain 2 (MD2), it contacts helicase motifs Ia, Ib, IV, V, Va and VII. The displaced 3' strand of the duplex may interact with basic residues on the surface of MD1.

3.10.2 MOLECULAR BASIS OF XPD MUTANT PHENOTYPES

It is expected that disease-causing mutations in the human protein that affect the ability of the enzyme to interact with other subunits of the TFIIH complex will be poorly conserved in the archaeal protein. Such mutations would likely generate mild phenotypes that are unrepresentative of the equivalent mutation in humans. Mutations that perturb the structure of XPD directly or those that affect the biochemistry of the protein, for instance those mutations that occur within the canonical helicase motifs, are likely to be conserved mutations that should produce a distinct phenotype.

This study reported the biochemical characterisation of two XP-causing mutations, R373Q (human R511Q) and R531Q (human R683Q). Both mutants purified in a manner comparable to the WT enzyme and CD spectroscopy data suggested that the proteins had regular secondary structure as observed for WT XPD. ATPase assays revealed that both mutants functioned as ssDNA-dependent ATPases; the R373Q and R531Q mutants hydrolysed ATP at a rate 50% and 170% compared to the WT enzyme. Crucially, both mutants bound ssDNA with reduced affinity, which provides an explanation to their inability to catalyse strand displacement under conditions

favourable for the WT protein. The structure of XPD supports these biochemical data, revealing that these mutants lie at the top of motor domain 2, along the expected path of ssDNA. The substitution of a positively charged residue for a neutral residue in a region that lines the path for ssDNA could explain the reduced DNA binding affinity and diminished helicase activity of these mutants. The biochemical and structural data (summarised in table 3.9) provide a strong basis for proposing that these mutations cause XP by inhibiting DNA unwinding.

The structural data presented in this chapter helps explain other XP-related mutations of XPD. Mutation of SacXPD (R531W) to mimic the XP-causing human mutation (R683W) resulted in impaired helicase, ATPase and DNA binding activity (Table 3.9) (Liu et al., 2008). Since archaeal XPD functions as a monomer and would not be influenced by protein-binding partners, the absence of helicase activity was explained by a diminished affinity for ssDNA which, in turn, affects the ability of the enzyme to hydrolyse ATP in a ssDNA-dependent manner. Indeed, the structure of StoXPD reveals that residue R531 is positioned along the expected path of ssDNA and it is likely that mutation of this basic residue for a bulky aromatic tryptophan would significantly perturb the interaction between XPD and DNA. The biochemistry and structural data reported for XP-causing XPD mutations are consistent with the hypothesis that the disorder results from perturbed XPD function as opposed to compromised stability of XPD and TFIIH (Lehmann, 2001). This reflects the fact that XP-causing mutations are predicted to affect the NER pathway of the cell, as opposed to transcription.

Two human XPD mutations associated with TTD were reproduced in SacXPD: A204Y (C259Y) and D521G (D673G). The purification of the A204Y mutant, together with structural data, suggests that this mutation perturbs the stability of SacXPD. The A204 residue is located within the Arch domain of XPD and, consequently, is likely to interfere with the stability of the TFIIH complex given that the Arch domain has been implicated in protein: protein interactions (Liu et al., 2008). The D521G mutation does not significantly affect the ability of the enzyme to hydrolyse ATP but does appear to bind ssDNA with increased affinity compared to the WT enzyme (summarised in table 3.9). The equivalent residue in human XPD is located in close proximity to the p44-interacting domain at the C-terminus of the protein (figure 3.1) and may, therefore, contribute to the stability of the TFIIH complex. D521G XPD was shown to bind DNA with a great affinity compared to WT XPD and this was

reflected in an increased ATPase activity associated with the mutant. However, the mutation did not appear to affect the ability of the protein to unwind DNA. The biochemical and structural data presented in this chapter are consistent with previous findings that suggest that TTD-causing XPD mutations interfere with the structure of XPD and, in turn, the stability of the TFIIH complex (Lehmann, 2001). This would perturb transcriptional activity, providing an explanation for the reduced stature and brittle hair characteristic of TTD patients.

The structural data presented here may shed light on how other mutations impact the structure and function of the protein. The PhD thesis of Dr J. Rudolf reported the biochemical consequences in archaea of the human G602D and G675R XPD mutations, both associated with the XP/ CS condition. Although CD spectroscopy data confirmed the proteins were stable and structured, neither protein could hydrolyse ATP nor unwind DNA (Table 3.9). Both these mutants target glycine residues that are typically associated with structural flexibility and found in non-conserved flexible loop regions. However, in the StoXPD structure, the archaeal equivalent of the G602 residue lies adjacent to residues of motif V, implicated in ATP and ssDNA binding. This could provide an explanation to the observed functional perturbation of the G602D mutation. Meanwhile, residue V513 (the StoXPD equivalent of the G675 residue) is situated at the base of motor domain 2 and close to a cluster of residues know to interaction with the p44 subunit of TFIIH in humans.

Table 3.9: Biochemistry of disease-related XPD mutations. hXPD, human XPD; SacXPD, *S. acidocaldarius* XPD; StoXPD, *S. tokodaii* XPD; XP, *xeroderma pigmentosum*; XP/ CS, combined *xeroderma pigmentosum* and Cockayne syndrome phenotype; TTD, trichothiodystrophy; FeS, refers to the presence or absence of a stable FeS cluster; ND, not determined; Yield, purification yield relative to WT XPD (%); N, no activity detected. Grey shading represents mutations tested in this study; remaining were investigated by Dr J. Rudolf (University of St Andrews).

Mutation/Residue			Disease	FeS	Compared to WT (%)			
hXPD	SacXPD	StoXPD			Yield	DNA affinity	ATPase	Helicase
C259Y	A204Y	A206	TTD	stable	5	87	89	~50
R511Q	R373Q	R365	XP	stable	100	49	54	N
G602D	G447D	A438	XP/ CS	stable	100	ND	6	N
D673Q	D521G	D55	TTD	stable	100	142	127	~70
G675R	C523R	V513	XP/ CS	stable	100	ND	4	N
D861N	D529N	D519	XP	ND	0	ND	ND	ND
R683Q	R531Q	R521	XP	stable	100	33	155	N
R683W	R531W	R521	XP	stable	100	529	50	-

This chapter has revealed how an archaeal system can be used to delineate the molecular basis of naturally occurring mutations that cause debilitating disease in humans.

CHAPTER 4

PURIFICATION AND CHARACTERISATION OF DING FROM *STAPHYLOCOCCUS AUREUS*

4.1   INTRODUCTION

The *dinG* gene was first identified in *E. coli* during a genetic screen for loci upregulated in response to DNA damage (Lewis et al., 1992). Subsequent bioinformatic analyses revealed that *dinG* encoded a putative helicase homologous to human XPD (hXPD), FancJ (hFancJ) and *Saccharomyces cerevisiae* Rad3 (SceRad3) (figure 4.1). These proteins have previously been shown to function as 5'-3' DNA helicases *in vitro* (Sung et al., 1987; Voloshin et al., 2003; Cantor et al., 2004; Rudolf et al., 2006), an activity that relies upon the coordination of an iron-sulphur (FeS) cluster located between the Walker A and B boxes of each protein (Rudolf et al., 2006; Pugh et al., 2008; Ren et al., 2009). Similarly, the majority of DinG proteins possess conserved helicase motifs and four cysteine residues positioned between the Walker A and B boxes, implicating them in FeS cluster coordination. Biochemical studies have confirmed that *E. coli* DinG (EcoDinG) is a FeS cluster containing protein (Voloshin and Camerini-Otero, 2007). However, some *dinG* sequences from Gram-positive and Firmicute bacteria, including DinG from *Staphylococcus aureus* (SarDinG), lack these cysteine residues implying the absence of a functional FeS cluster (figure 4.2). Given the dependence of DinG homologues on the presence of a stable FeS cluster for helicase

activity, DinG proteins lacking this cluster may have lost the ability to unwind DNA. However, these sequences retain the conserved helicases motifs (figure 4.1) observed in homologous DNA helicases and, thus, may have replaced the cluster with an alternative strand-displacement mechanism.
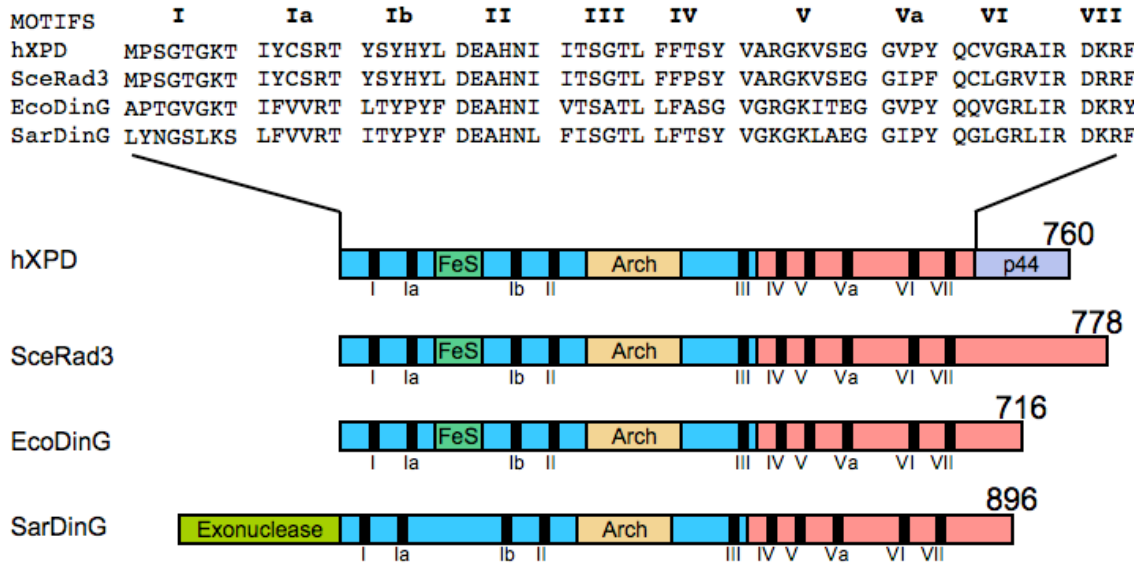
Figure 4.1: Alignment of DinG to related gene sequences

Cartoon of the domain arrangement of helicases related to DinG, including human XPD (hXPD), *S. cerevisiae* Rad3 (SceRad3), and *E. coli* DinG (EcoDinG). Approximate positions (black rectangles) and sequences of the conserved helicase motifs are indicated. Numbers to the right of each domain structure refers to the number of amino acids in each protein. Additional domains are highlighted, including motor domains 1 (cyan) and 2 (pink), the arch domain (wheat), the FeS cluster (green), the p44-interacting domain of hXPD and the exonuclease domain of SarDinG. The assignation of motor domains, arch domain and the FeS cluster in SceRad3, EcoDinG and SarDinG is based on their amino acid sequence alignment with hXPD, for which an X-ray crystal structure is known (Fan et al., 2008; Liu et al., 2008; Wolski et al., 2008).

113
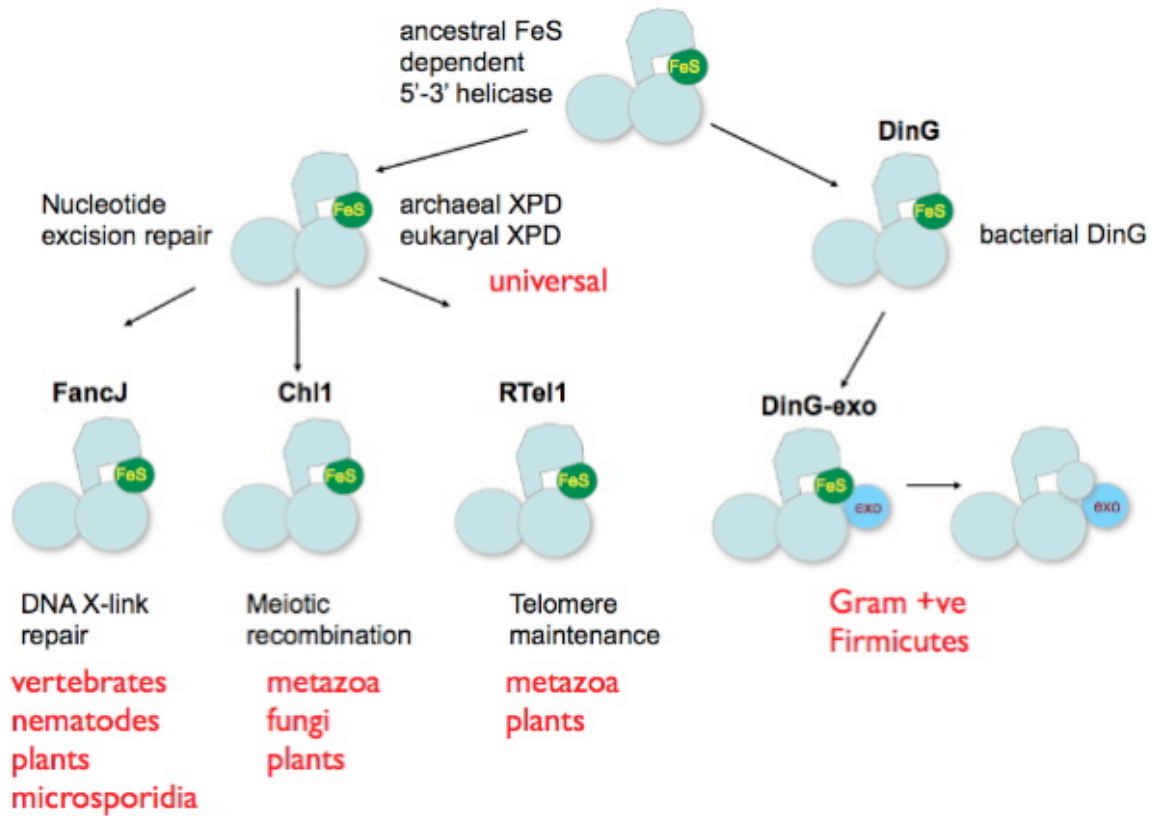
Figure 4.2: Proposed evolution of DinG-related helicases

An ancestral 5'-3' helicase with an essential iron-sulphur cluster binding domain diverged to give rise to the XPD family in archaea and eukaryotes and DinG family in bacteria. Gene duplication and specialisation of XPD produced FancJ, Chl1 and RTel1, while similar evolutionary events in bacteria resulted in the acquisition of an N-terminal putative exonuclease domain and subsequent loss of the FeS cluster in Gram-positive bacteria. Adapted from (White, 2009).

In addition, a large proportion of these non-FeS binding DinG proteins, including SarDinG, have acquired an N-terminal domain (figure 4.1) that exhibits homology to the epsilon proofreading subunit of DNA polymerase III (figure 4.3). This protein is the archetypal member of the DnaQ family of 3'-5' exonucleases; other family members include the bacterial RNaseT and RNaseD proteins, exonuclease I and the nuclease domain of the human Werner syndrome protein (WRN) (Viswanathan and Lovett, 1999). The DnaQ superfamily are characterised by an α-β structure, comprising a central core of parallel β-sheet structures. The presence of a putative exonuclease and helicase domain could endow SarDinG with the ability to displace and subsequently degrade a strand from a duplex, akin to that observed for the multifunctional, eukaryotic WRN protein (Bukowy et al., 2008).
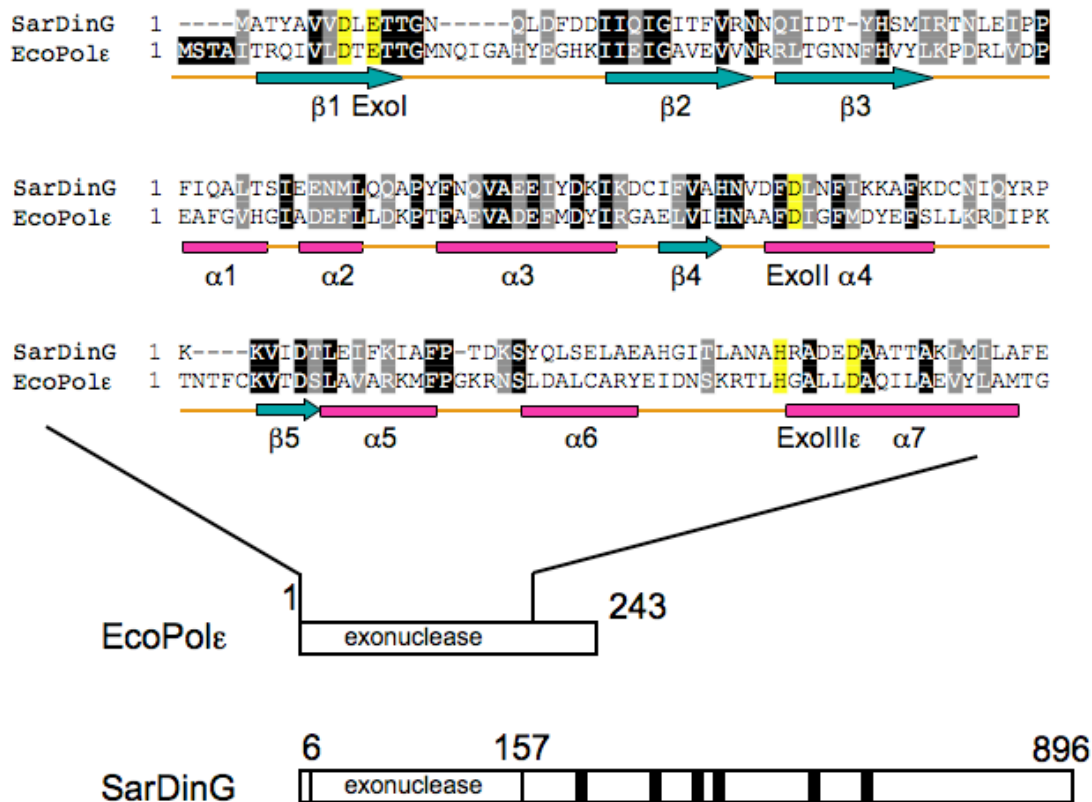
Figure 4.3: Similarity of SarDinG to the *E. coli* ε-subunit of DNA polymerase III
Cartoon representation and alignment of ε-subunit of DNA polymerase III from *E. coli* (EcoPolε) and the exonuclease domain of DinG (SarDinG). Secondary structural elements of EcoPolε are shown below the sequence: loops, orange; α-helices, pink; β-strands, teal. Active site residues in the ExoI, ExoII and ExoIIIε motifs (marked) are shaded in yellow, other identical residues are shown as white on black, and similar residues are shown as white on grey. Modified from (Kirby et al., 2006).

Previous work has revealed that EcoDinG possesses functions reminiscent of archaeal XPD (Rudolf et al., 2006), including DNA-dependent ATPase and 5'-3' helicase activities (Voloshin et al., 2003). EcoDinG appears to have specificity for unwinding substrates with 5' ssDNA, DNA-RNA duplexes and D-loops (Voloshin and Camerini-Otero, 2007). The ability of DinG to unwind synthetic D-loop substrates, which mimic intermediates of homologous recombination, suggests that the protein may play a role in recombinational repair. The *in vivo* function of bacterial DinG remains elusive since neither deletion nor overexpression of *dinG* from *E. coli* produced a discernible phenotype, suggesting that the protein is non-essential for cell viability (Voloshin et al., 2003). Rather, a mild effect on cell survival following UV irradiation and treatment with mitomycin C and nalidixic acid was observed (Van Dyk et al., 2001;

Voloshin et al., 2003). These findings do not exclude the possibility that bacterial DinG participates in DNA repair processes.

This chapter deals with the initial characterisation of wild-type (WT) DinG from *S. aureus*. The homology of DinG to hXPD and hFancJ, two proteins associated with rare genetic diseases upon mutation (Lehmann, 2001; Cantor et al., 2004), presents an invaluable opportunity to use a genetically-tractable protein to gain further insight into more complex, eukaryotic proteins and repair pathways. If not otherwise stated, DinG refers to SarDinG.

4.2 PURIFICATION OF WT DING

The Scottish Structural Proteomics Facility at the University of St Andrews PCR-amplified the gene encoding DinG from *Staphylococcus aureus* genomic DNA. The gene was cloned into the pDEST14 vector (by Gateway® cloning) to produce recombinant protein with an N-terminal polyhistidine tag. DNA sequencing confirmed that DinG had been cloned without mutation. Protein was expressed in *E. coli* BL21 Rosetta cells and purified to homogeneity using affinity (nickel-chelating) and size exclusion chromatography (figure 4.4), as described in Methods and Materials. Protein was stored at -80 ºC since prolonged storage of the protein at 4 ºC resulted in precipitation and degradation, as previously reported for *E. coli* DinG (Voloshin et al., 2003).

Protein was recovered at 5-10 mg / L culture after the two-step chromatography procedure, although a significant proportion of the protein was lost due to aggregation or multimerisation. This can be clearly observed in the gel filtration chromatogram, in which DinG eluted from the column in distinct peaks (figure 4.4 B-C). The high molecular weight DinG (fractions 10-14) may reflect protein aggregation since previous studies with the *E. coli* homologue reported the inclusion of 1M NaCl during purification to prevent non-specific protein interactions (Voloshin, Vanevski et al., 2003). However, by comparing multiple gel filtration chromatograms of DinG, it was apparent that this single high molecular weight species was consistently observed, eluting from the column in the same volume; this suggests that the high molecular weight species may represent a DinG multimer.
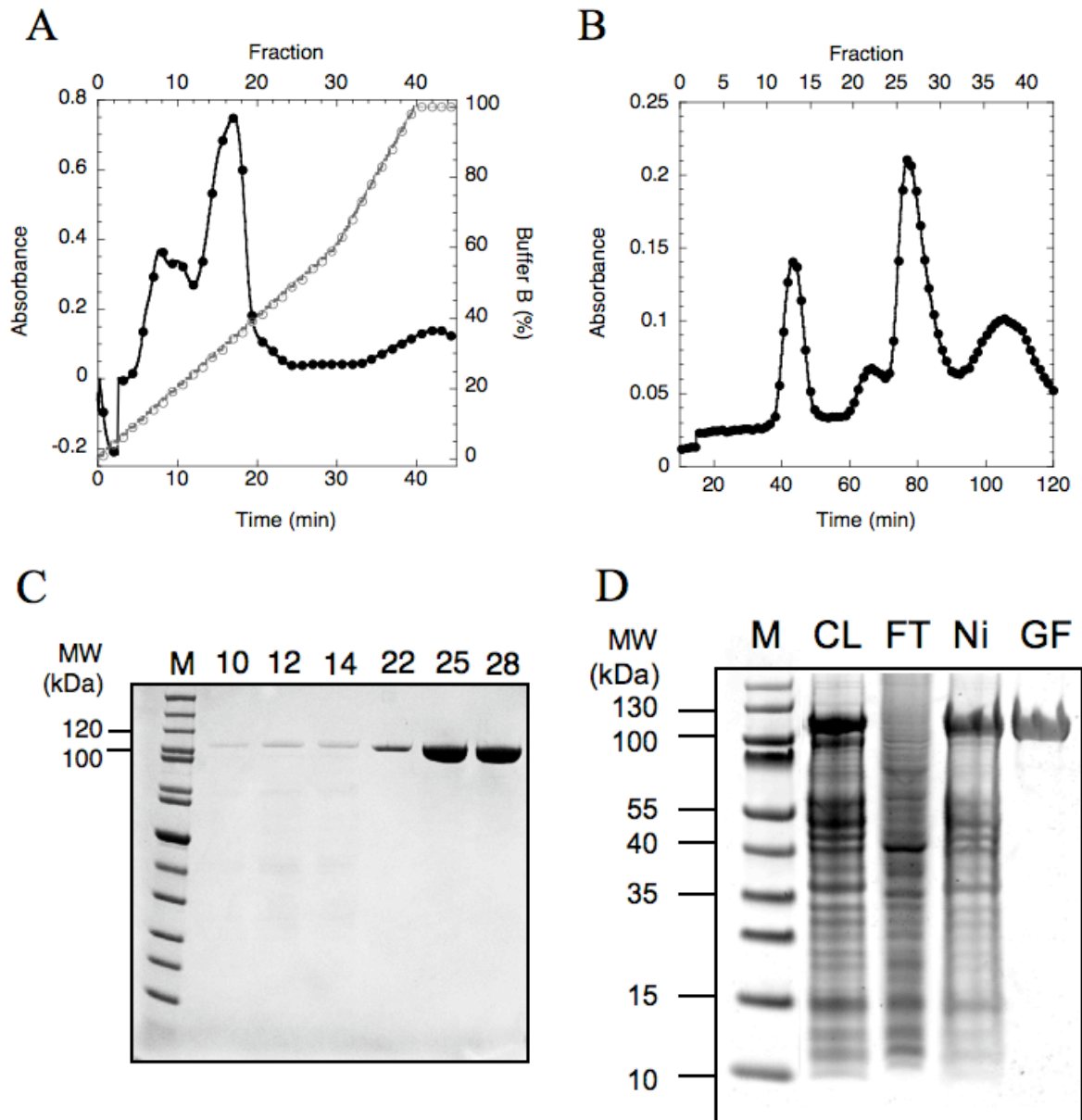
Figure 4.4: Purification of recombinant DinG

Elution chromatograms showing DinG purification by nickel-chelating (A) and size exclusion chromatography (B). Fractions from major elution peaks during gel filtration chromatography were analysed by SDS-PAGE (C). Fraction numbers are indicated above the gel image. M, marker. (D) Fractions from each stage of DinG purification were analysed by SDS-PAGE. From left to right: M, marker; *E. coli* cell lysate (CL), flow through from nickel column (FL), and DinG after nickel-chelating (Ni) and gel filtration (GF) chromatography.

MALDI-TOF and ESI-TOF mass spectrometry were employed to confirm the identity and integrity of purified DinG, respectively. While unequivocal protein identification was possible, DinG exceeded the upper size limit of ESI-TOF mass spectrometry resulting in spectra that could not be confidently interpreted to obtain an

intact mass. Determination of the approximate size of DinG was, therefore, performed using a Superdex 200 10-300 gel filtration column (GE Healthcare) calibrated with proteins of known molecular weight (Appendix 5). DinG (4 mg/ ml) was passed through the calibrated column and the majority of protein eluted after 52 ml. Using the equation of the standard curve, the mass of DinG was calculated to be 115 kDa (figure 4.5), which was in good agreement with the expected size of 107 kDa for monomeric DinG. A higher molecular weight species (elution volume = 35 ml), estimated to be 8.5 times the size of monomeric DinG, was also observed consistent with the hypothesis that DinG forms aggregates via non-specific protein interactions. Although only monomeric and aggregated DinG were observed by gel filtration analysis, it does not necessarily follow that the functional form of DinG *in vivo* is monomeric: multimerisation may occur in the presence of DNA or nucleotide cofactors.
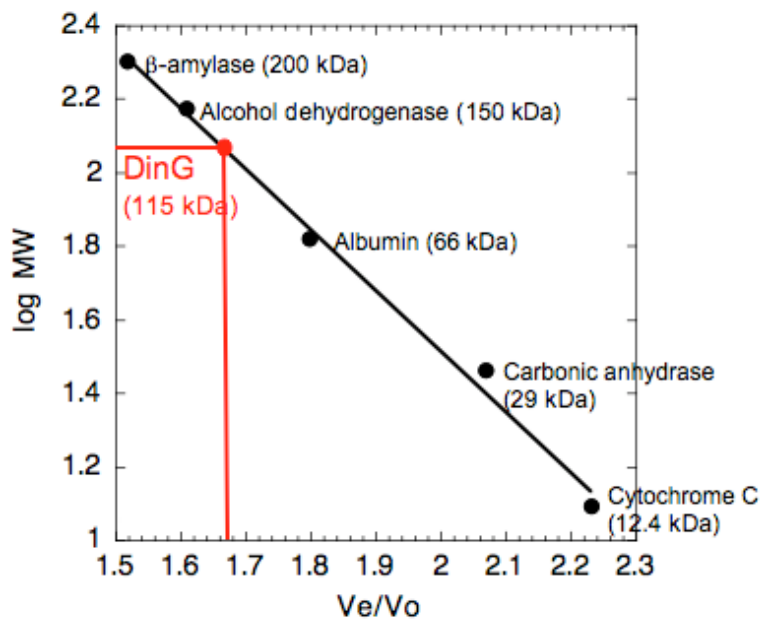
Figure 4.5: Determination of size and oligomeric state of DinG
A standard curve was generated based on the elution volume (Ve) of proteins of known molecular weight (as indicated on plot) from a Superdex 200 10-300 size exclusion column (GE Healthcare). Vo, void volume, refers to the elution volume of blue dextran (31 ml). The molecular mass of DinG was estimated using the equation of the standard curve and its elution volume (52 ml) from the size exclusion column.

4.3   DOMAIN ORGANISATION OF DinG

The InterProScan algorithm, provided by EMBL-EBI, was used to scan DinG for recognisable domain patterns. As shown in figure 4.6, DinG comprises an N-terminal exonuclease and downstream helicase domain.

To characterise DinG biochemistry, three mutants were constructed in addition to the wild-type (WT) protein. The first mutant, referred to as K304A, was produced by substitution of a highly conserved lysine residue (residue 304) to alanine, resulting in disruption of the Walker A box. The lysine residue occupies the site that will be later occupied by the magnesium ion upon interaction of the protein with NTP-Mg^{++} complexes (Tuteja and Tuteja, 2004). Upon NTP-Mg^{++} binding, the lysine residue interacts with the β- and γ-phosphoryl groups of NTP, thereby stabilising the transition state during ATP hydrolysis (Tuteja and Tuteja, 2004). Therefore, it is expected that mutation of this lysine residue will abolish ATP hydrolysis and dramatically reduce the ability of the protein to bind NTP-Mg^{++} complexes. A second mutation, known as D10A/ E12A, was generated by substitution of residues D10 and E12 to alanine. These highly conserved residues play an essential role in exonuclease activity. By mutating these residues, the helicase activity of DinG can be assessed in the absence of exonuclease activity. Cloning only the helicase domain of the protein generated a final DinG mutant, called Δnuclease. This construct would permit the functional analysis of DinG helicase activity in the absence of any steric influences imposed by the N-terminal exonuclease extension. These mutants were expressed and purified (figure 4.6 B) as outlined for WT DinG.
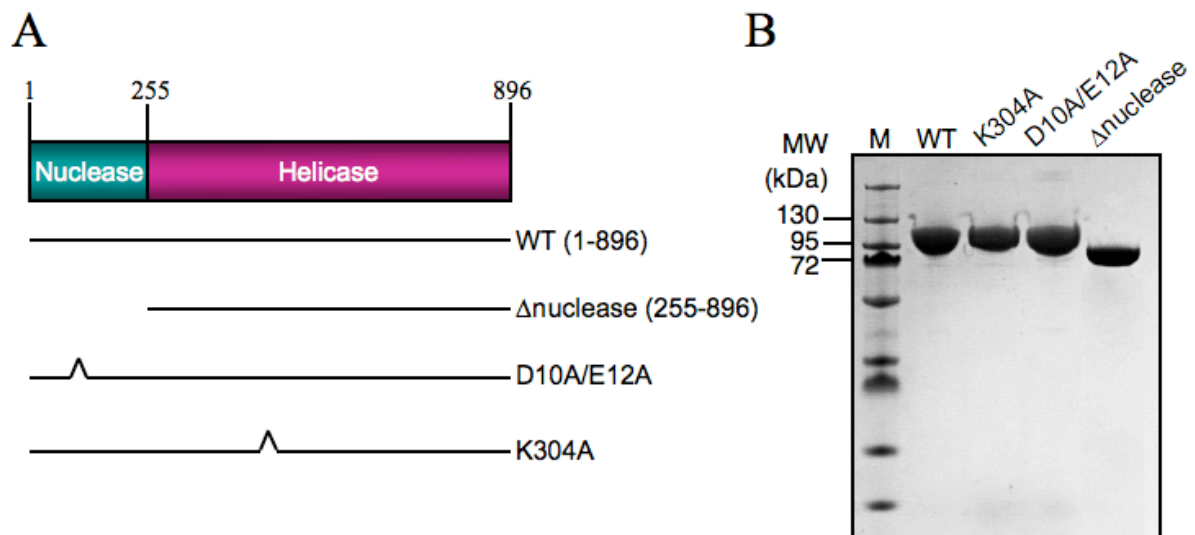
Figure 4.6: Domain organisation of DinG and purification of mutants

(A) Domain organisation of DinG and constructs used throughout this chapter. Approximate positions of point mutations are shown. Numbers refer to amino acid residues. (B) SDS-PAGE showing purified WT and mutant DinG. M, marker.

4.4   DING HAS DNA-INDEPENDENT ATPASE ACTIVITY

The sequence of DinG comprises conserved helicase motifs (figure 4.1) that endow proteins with various helicase-related functions, including the ability to catalyse ATP hydrolysis. ATPase assays were performed with DinG in the presence and absence of DNA, as outlined in Materials and Methods. Briefly, DinG (0.2 µM) was incubated with 10 nM φX174 virion (ssDNA) or RFI DNA (covalently closed circular dsDNA) at 37 ºC. Reactions were initiated by 1 mM ATP/ MgCl$_2$ and stopped, at specific time points, by the addition of chilled 0.3 M PCA. The assay relies on the interaction of malachite green with free phosphate to generate a concentration-dependent colour change that can be measured spectroscopically at 650 nm. ATPase assays were performed in the absence of protein and DNA, to ensure that free phosphate was not present in any reagents used during the assay, and in the presence of K304A DinG, which lacks the essential lysine residue required for ATP hydrolysis .
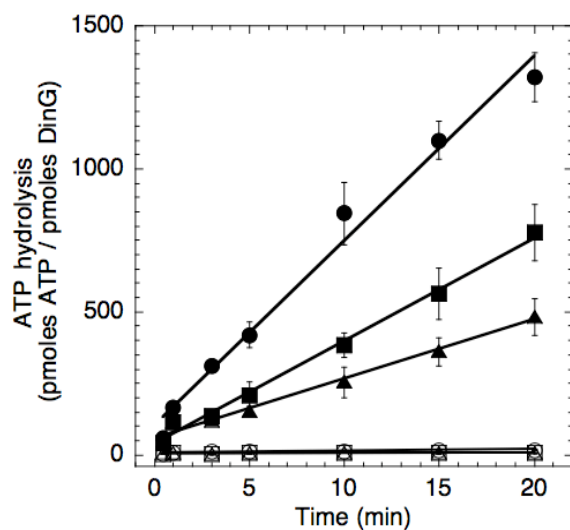
Figure 4.7: ATPase activity of DinG

10 nM DNA and 0.2 µM enzyme were incubated at 37 ºC prior to reaction initiation by 1 mM ATP/MgCl$_2$. At specific time points, reactions were stopped by addition of chilled 0.3 M PCA solution. Reactions were performed in triplicate, using different protein preparations, and standard errors are shown. ssDNA, circles; dsDNA, squares; triangles, no DNA; closed shapes, WT-DinG; open shapes, K304A-DinG.

Table 4.1: Rate of ATP hydrolysis by DinG

The rate of ATP hydrolysis by WT and K304A DinG is shown as pmoles phosphate.min^{-1}.pmoles enzyme^{-1}. Rates represent the mean of triplicate data with standard errors shown.

Substrate	Rate of ATP hydrolysis	
	WT	**K304A**
φX174 ssDNA	69 ± 6.2	2 ± 0.8
φX174 RFI dsDNA	37 ± 3.2	1 ± 0.6
No DNA	24 ± 2.8	1 ± 0.5

ATP was hydrolysed by DinG at a rate of 69 and 37 pmoles phosphate/ min/ pmole enzyme in the presence of ssDNA and dsDNA, respectively (figure 4.7 and table 4.1). This suggests that the protein may be able to translocate along both single- and double-stranded DNA, and may have implications for its substrate specificity with regards to its helicase and nuclease activities. Previous studies have shown that the homologous protein from *E. coli* (EcoDinG) catalyses ssDNA-dependent ATP hydrolysis at a similar rate to DinG from *S. aureus* (*kcat* of 24.1 ± 0.92) (Voloshin et al., 2003). However, EcoDinG exhibits minimal ATPase activity in the presence of dsDNA, implying that these two proteins may exhibit alternate substrate specificity during DNA unwinding and translocation. The ability of DinG to hydrolyse ATP was not dependent on DNA; rates of ATP hydrolysis in the absence of DNA were approximately 3-fold lower than those observed in the presence of ssDNA. As expected, the K304A DinG mutant had minimal ATPase activity in the presence or absence of DNA. These results provide useful negative controls and confirm that any ATPase activity observed was the result of catalysis by WT DinG.

4.5   LACK OF DING HELICASE ACTIVITY

The DinG protein from *E. coli* has been shown to unwind a large variety of DNA substrates *in vitro*, including 5' flap and D-loop structures (Voloshin and Camerini-Otero, 2007). The substrate specificity of DinG helicase activity was examined *in vitro*. The polarity of DNA unwinding was determined using DNA substrates possessing a 5' or 3' single-stranded tail. Helicase assays comprising a splayed duplex substrate (providing the helicase with a single-stranded loading site), or a blunt-ended duplex (containing a 7-nucleotide, uncomplementary region to mimic a typical 'bubble' substrate of nucleotide-excision repair) tested the ability of DinG to unwind dsDNA under various conditions relevant to the *in vivo* environment.

Helicase assays were performed as outlined in Materials and Methods, with reactions initiated by the addition of 1 μM DinG. Unless otherwise stated, reactions were performed at pH 6.8. The results reveal that DinG is a very poor helicase *in vitro*, unable to unwind any tested DNA substrate within 1 hr at 37 ºC (figure 4.8). By comparison, the homologous protein from *E. coli* (EcoDinG) has been reported to unwind DNA, with 5'-3' polarity, in 2.5 min when incubated at 30 ºC at a concentration of 100 nM (Voloshin et al., 2003). The striking difference in helicase activities between

DinG from *S. aureus* and *E. coli* may result from the presence of the N-terminal exonuclease domain in the former protein. Helicase assays were, therefore, repeated in the presence of D10A/ E12A DinG (1 µM), which lacks exonuclease activity. However, D10A/ E12A DinG failed to unwind a variety of DNA substrates (figure 4.8, right hand panel), implying that the functionality of the DinG exonuclease domain does not affect the helicase activity of the protein.
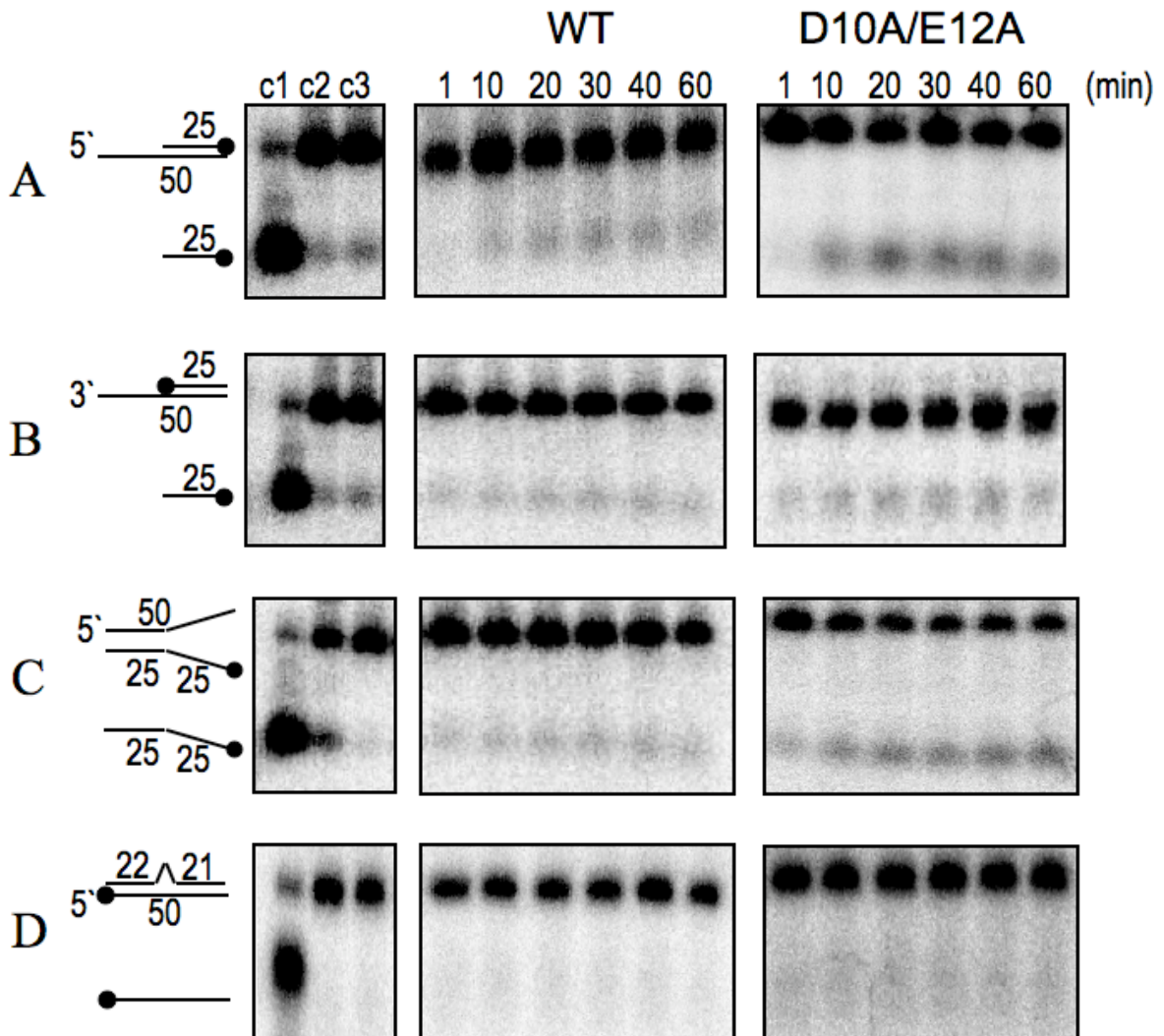
Figure 4.8: Lack of DinG helicase activity

Helicase activity was determined by incubating WT and D10A/ E12A DinG (1 µM) with DNA (black circle, 5' [^{32}P] radiolabel) at 37 °C. After specific time points (indicated above gel), reactions were stopped, deproteinised and separated on a native 12% polyacrylamide: TBE gel. c1, substrate incubated at 95 °C; c2, end-point reaction in the absence of protein; c3, end-point reaction in the absence of ATP/ MgCl$_2$. Each reaction was repeated at least three times using different protein preparations.

The D10A/ E12A DinG mutant does not exclude the possibility that the exonuclease domain physically inhibits the helicase activity of the protein, e.g. by

sterically blocking access of the downstream helicase domain to the DNA. To investigate this hypothesis, Δnuclease DinG (1 µM) was incubated with a 5'-overhang DNA substrate over 60 min at 37 ℃. The 5'-overhang DNA substrate was chosen since DinG homologues possess 5'-3' helicase activity. However, removal of the exonuclease domain did not affect the helicase activity of DinG (figure 4.9). Therefore, it is unlikely that acquisition of an N-terminal exonuclease domain by DinG has led directly to the inability of the protein to unwind DNA substrates with an efficiency comparable to homologous proteins lacking the nuclease domain, e.g. EcoDinG.
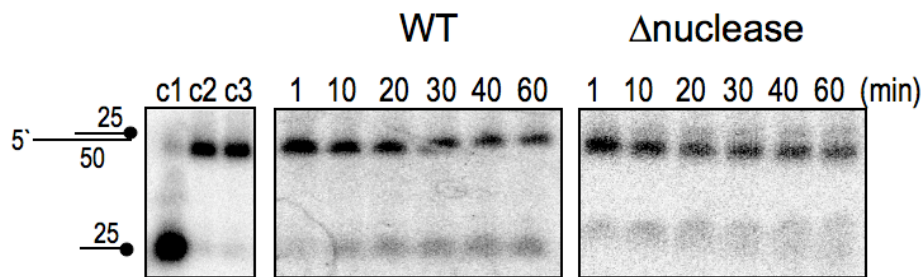
Figure 4.9: Comparison of WT and Δnuclease DinG helicase activity

Representative native 12% (v/v) polyacrylamide:TBE gel comparing the ability of WT and Δnuclease DinG (1 µM) to unwind a 5'-overhang DNA substrate (black circle, 5'-[^{32}P] radiolabelled) at 37 ℃ over 60 min (time points indicated above gels). c1, 25 nt DNA size marker; c2, end-point reaction without protein; c3, end-point reaction without ATP/ MgCl$_2$. The images are representative of triplicate data.

The helicase activity of EcoDinG protein was reported to exhibit pH-dependent characteristics, with strand displacement occurring at pH values greater than pH 7 (theoretical pI of EcoDinG is 7.6). Therefore, helicase assays were performed to assess the ability of DinG (1 µM) to unwind a 5'- and 3'-overhang DNA substrate at pH 6.5, 7.0 and 7.5 over 60 min. As revealed in figure 4.10, increasing the alkalinity of the assay conditions did not promote strand displacement by DinG.
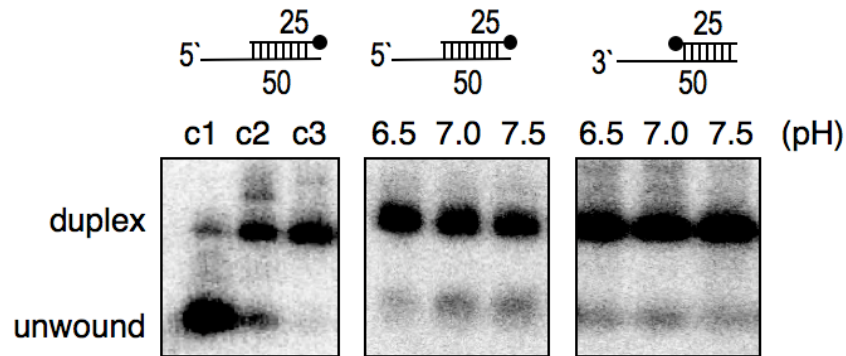
Figure 4.10: pH-dependence of DinG helicase activity

DinG (1 µM) was incubated with 10 nM DNA (5' or 3' overhang) at pH 6.5, 7.0, or 7.5 for 60 min at 37 ºC. Products were separated on 12% (v/v) native polyacrylamide: TBE gels. c1, substrate incubated at 95 ºC; c2, without protein; c3, without ATP/ $MgCl_2$. Each experiment was performed in triplicate.

Sequence analysis of homologous proteins presents an alternative explanation for the absence of DinG helicase activity. *E. coli* DinG, for instance, has been shown to possess an FeS cluster that is coordinated by four highly conserved cysteine residues located between the Walker A and B boxes (figure 4.1); mutation of any one of these coordinating residues abrogate EcoDinG helicase activity (Rudolf et al., 2006). This indicates that the ability of the helicase to unwind DNA is dependent upon the presence of a stable FeS cluster (Voloshin and Camerini-Otero, 2007; Ren et al., 2009). By contrast, DinG from *S. aureus* lacks these cysteine residues and, therefore, an FeS cluster (figure 4.11). The dependence of EcoDinG and SacXPD helicase activity on a stable FeS cluster may provide an explanation for the absence of DNA unwinding by DinG from *S. aureus*.

Despite possessing the canonical sequence motifs characteristic of superfamily II helicases (figure 4.1), evidence to suggest that DinG from *S. aureus* is a *bona fide* DNA helicase is lacking. Absence of helicase activity despite encoding helicase motifs has been observed with *S. solfataricus* XPB1 and XPB2 (Richards et al., 2008a), members of the SWI/SNF family (Carlson and Laurent, 1994), including Rad54 (Swagemakers et al., 1998), and transcription-repair coupling factors CSB/ERCC6 (Selby and Sancar, 1997) and bacterial Mfd (Selby and Sancar, 1995). In each instance, these proteins have been hypothesised to couple ATP hydrolysis to DNA translocation to perform its function *in vivo*. Therefore, the ability of DinG to translocate along ssDNA was investigated.

```
SarDinG      1 KKPTTNYNGSIKSLYSKAVDQLG--LTYRPQQLYLAETILDQLMHS-EKAMIEASLGSGK
EcoDinG      1 MALTAALKAQIAAWYKALQEQIPDFIPRAPQRQMIADVAKTLAGEEGRHLAIEAPTGVGK

SarDinG     58 SLAYLLAALMYNIETGKHVMISTNTKLLQSQLLEKDIPAVNEALN-FKINALLIK-----
EcoDinG     61 TLSYLIPGIAIAREEQKTLVVSTANVALQDQIYSKDLPLLKKIIPDLKFTAAFGRGRYVC

SarDinG    112 -------SKSDYISLGLISQILKDDTSNYEVNILKMQLLIWITETPSGD--IQELNLKGG
EcoDinG    121 PRNLTALASTEPTQQDLLAFLDDELTPNNQEEQKRCAKLKGDLDTYKWDGLRDHTDIAID

SarDinG    163 QKMYFDQKIETYVPARHDVHYYN-----FIKRNAQNIQIGITNHAHLIHSDVENSIYQLF
EcoDinG    181 DDLWRRLSTDKASCLNRNCYYRECPFFVARREIQEAEVVVANHALVMAAMESEAVLPDP

SarDinG    218 DD--CIVDEAHRLPDYALN--QVTNELSYADIKYQLGLIGKNENEKILKA-IDQLEKQRI
EcoDinG    241 KNLLLVLDEGHHLPDVARDALEMSAEITAPWYRLQLDLFTKLVATCMEQFRPKTIPPLAI
```

Figure 4.11: Alignment of DinG from *S. aureus* and *E. coli*

Sequence alignment of EcoDinG and SarDinG (residues 226-497 only) reveal that the latter lacks the four cysteine residues (highlighted in yellow) responsible for coordinating the FeS cluster present in the *E. coli* protein. Red shading, motif I (Walker A box); blue shading, motif II (Walker B box).

4.6 DING CANNOT DISPLACE STREPTAVIDIN FROM SSDNA

The inability of DinG to unwind DNA suggests that the protein may harness the energy from ATP hydrolysis to perform alternative functions in the cell. For instance, DinG may translocate along DNA and displace protein, as opposed to DNA, an activity that has been previously reported for the Bacteriophage T4 Dda helicase (Byrd and Raney, 2004). An *in vitro* streptavidin displacement assay (Byrd and Raney, 2004) can be employed to monitor this type of activity and exploits the tight interaction formed between streptavidin and biotin. A biotinylated 5' [^{32}P]-end labelled oligonucleotide (B50; Appendix 1, table A1.4) was incubated with streptavidin to achieve a higher molecular weight substrate. Providing a protein can track along the DNA and disrupt the biotin:streptavidin linkage, analysis by native acrylamide gel electrophoresis will reveal a band shift indicative of the displacement of streptavidin to produce a lower molecular weight product. The oligonucleotide was conjugated to biotin at either the 5' or 3' end to observe DNA translocation in the 3'-5' or 5'-3' direction, respectively. Assays were performed at 37 ºC and initiated by the addition of 1 µM DinG. At specific time points, reactions were stopped according to Material and Methods, prior to product separation by gel electrophoresis.
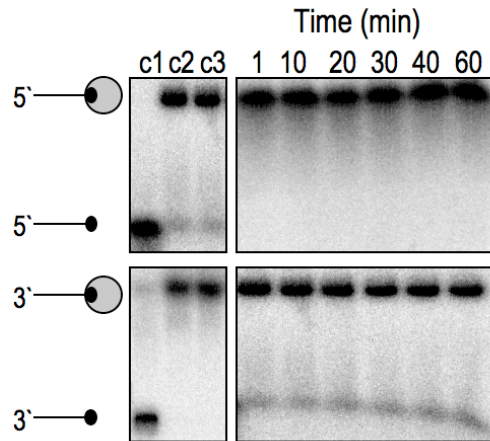
Figure 4.12: No streptavidin displacement from ssDNA by DinG

The ability of DinG (1 µM) to translocate along DNA (10 nM) in an ATP-dependent manner (1 mM ATP/ MgCl$_2$) was investigated by measuring the ability of the protein to displace streptavidin from a 5' [^{32}P]-end labelled biotinylated oligonucleotide (black circle, biotin; black line, B50 oligonucleotide; grey circle, streptavidin). c1, without streptavidin; c2, without protein after 60 min; c3, without ATP/ MgCl$_2$ after 60 min. Reactions were analysed on native 12% (v/v) polyacrylamide: TBE gels. Each experiment was performed in triplicate.

The results reveal that DinG was unable to displace streptavidin from biotinylated ssDNA by translocation in either a 5'-3' or 3'-5' direction (figure 4.12). Importantly, this does not exclude the possibility that DinG can translocate along DNA; rather, it disputes the hypothesis that DinG can displace proteins, e.g. the single-stranded DNA binding protein (SSB), during translocation. Similar findings have been reported for archaeal XPD (thesis of Dr J. Rudolf), XPB1 and XPB2 (Richards et al., 2008a). Although neither XPB1 nor XPB2 can unwind DNA (Richards et al., 2008a), archaeal XPD can melt duplex DNA despite reporting a negative result from the streptavidin displacement assay (Rudolf et al., 2006). Chapter 5 reveals that Hel308 from *S. solfataricus* PBL2025 can displace streptavidin from a biotinylated oligonucleotide, providing a positive control for this assay.

4.7 NUCLEASE ACTIVITY OF DING

In contrast to the DinG-related helicases (including EcoDinG, hXPD and hFancJ), DinG from *S. aureus* possesses an N-terminal domain (figure 4.1) homologous to the DnaQ family of 3'-5' exonucleases. To explore the catalytic activity of this domain, nuclease assays were performed using 5' [^{32}P]-end labelled DNA substrates, as described in Materials and Methods. Cleavage products of these assays were separated

on denaturing 20% (v/v) polyacrylamide: TBE gels, together with a Maxam-Gilbert A+G DNA ladder, and visualised by phosphorimaging. The caveat of using 5' end labelled DNA is that progressive DNA degradation could be analysed in a 3'-5' direction only, since cleavage in a 5'-3' direction would result in the immediate loss of the radiolabel tracer (figure 4.13).
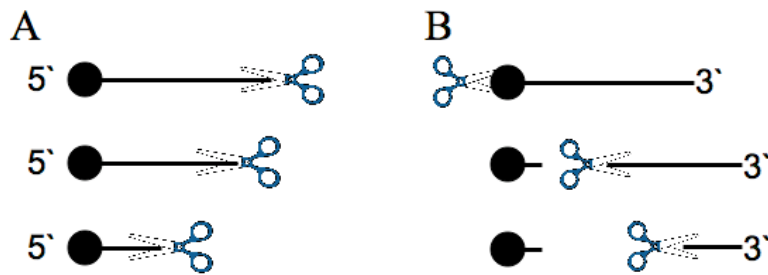
Figure 4.13: 5' [³²P]-end labelled DNA permits observation of 3'-5' nuclease activity

Nuclease assays were performed with 5' [³²P]-end labelled DNA. (A) Degradation of DNA in 3'-5' direction can be visualised by polyacrylamide gel electrophoresis because the strand retains the [³²P]-radiolabel until the end of degradation. (B) Progressive DNA cleavage in a 5'-3' direction cannot be readily observed since the [³²P]-radiolabel would be lost immediately.

4.7.1 DING FUNCTIONS AS AN EXONUCLEASE

The ability of DinG to degrade ssDNA (5' [³²P]-end labelled X26-50 oligo; Appendix 1, table A1.3) was investigated, as outlined in Materials and Methods. In the presence of 1 mM ATP, DinG (500 nM) was unable to degrade ssDNA (10 nM) within 60 min at 37 °C (figure 4.14). However, omission of ATP from the reaction permitted ssDNA cleavage in a 3'-5' direction in the same time period. This experiment does not exclude the possibility that DinG degrades ssDNA in a 5'-3' direction; however, the absence of very small fragments at early time points (indicative of immediate cleavage of the radiolabel tracer (figure 4.13)), suggests that SarDinG preferentially (if not exclusively) cleaves DNA with 3'-5' polarity.

To determine whether ATP binding or hydrolysis was responsible for the inhibition of DinG nuclease activity, the ability of the K304A mutant to digest ssDNA was investigated in the presence and absence of ATP. As outlined in section 4.3, this mutant is expected to display reduced ATP binding (compared to WT protein) and lack ATP hydrolysis activity. To reflect this, 5 mM ATP was used to test the nuclease

activity of the K304A mutant (in contrast to 1 mM ATP for WT experiments). As observed for WT DinG, ATP inhibited the ability of K304A DinG to degrade ssDNA (figure 4.14), an activity that was restored in the absence of ATP. These results imply that ATP binding inhibits DinG nuclease activity but do not exclude the possibility that ATP hydrolysis (or ADP binding) also inhibits DNA cleavage. It is important to note that ATP does not *directly* inhibit the exonuclease domain but rather as a result of its binding within the helicase domain. It may be hypothesised that ATP-driven conformational changes within the helicase domain of DinG re-positions the upstream nuclease domain such that DNA cleavage is prohibited.
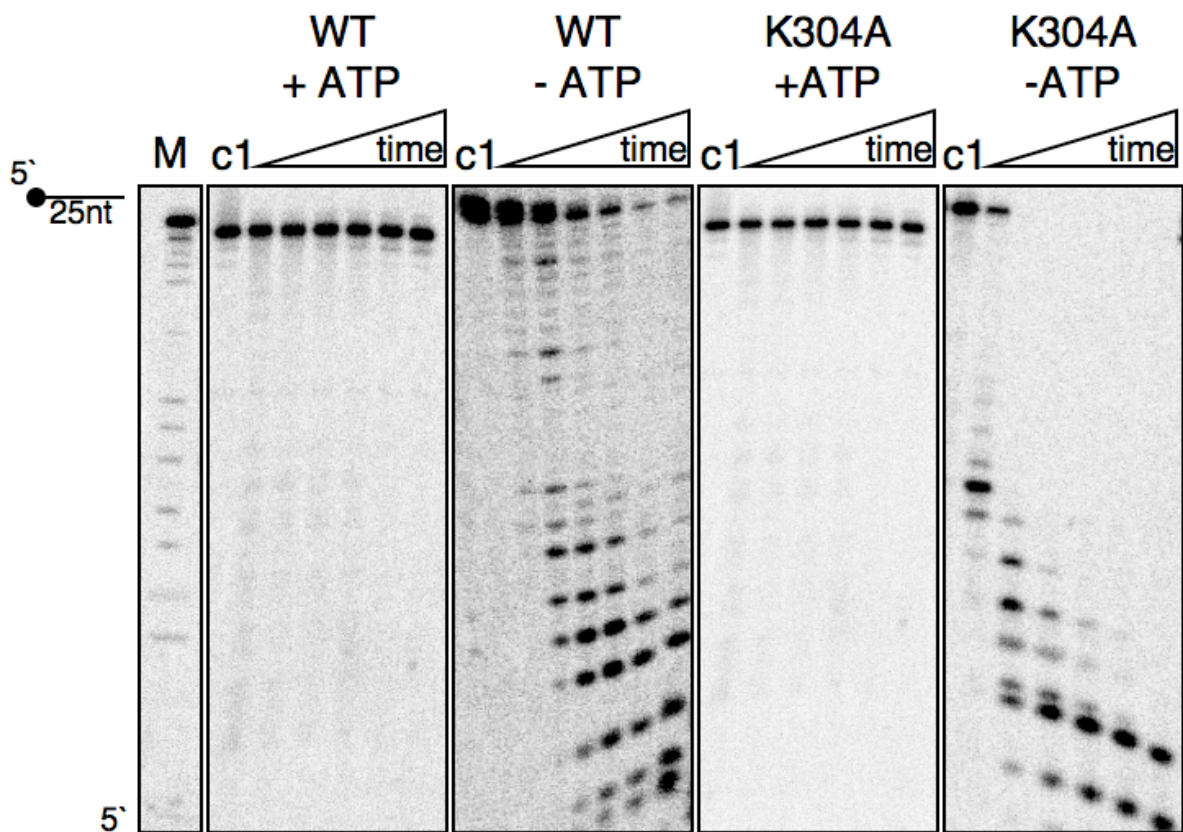
Figure 4.14: DinG cleaves ssDNA

Representative denaturing 20% (v/v) polyacrylamide: TBE gels showing cleavage of ssDNA (5' [^{32}P]-end labelled X26-50) by WT and K304A DinG (500 nM) at 37 °C. WT experiments were performed in the presence and absence of 1 mM ATP; K304A DinG experiments were performed in the presence and absence of 5 mM ATP. Time points: 1, 10, 20, 30, 40, 60 min. M, Maxam-Gilbert A+G ladder of the radiolabelled strand (black circle, [^{32}P]-radiolabel). c1, end-point reaction in the absence of protein. Each experiment was performed in triplicate.

4.7.1.1   INFLUENCE OF ATP CONCENTRATION ON DING NUCLEASE ACTIVITY

The effect of ATP concentration on DinG nuclease activity was investigated. The ability of WT DinG (500 nM) to cleave ssDNA (10 nM 5' [^{32}P]-end labelled X26-50 oligonucleotide) was inhibited at ATP concentrations above 0.5 mM (figure 4.15), representing a 1000-fold molar excess of ATP over protein. Since some inhibition was observed at ATP concentrations of 0.1 mM, all subsequent nuclease assays using WT DinG were performed in the absence of ATP. While the nuclease activity of the K304A DinG mutant was more resistant to the inhibitory effects of ATP, complete inhibition was observed at ATP concentrations of approximately 5 mM (figure 4.15).
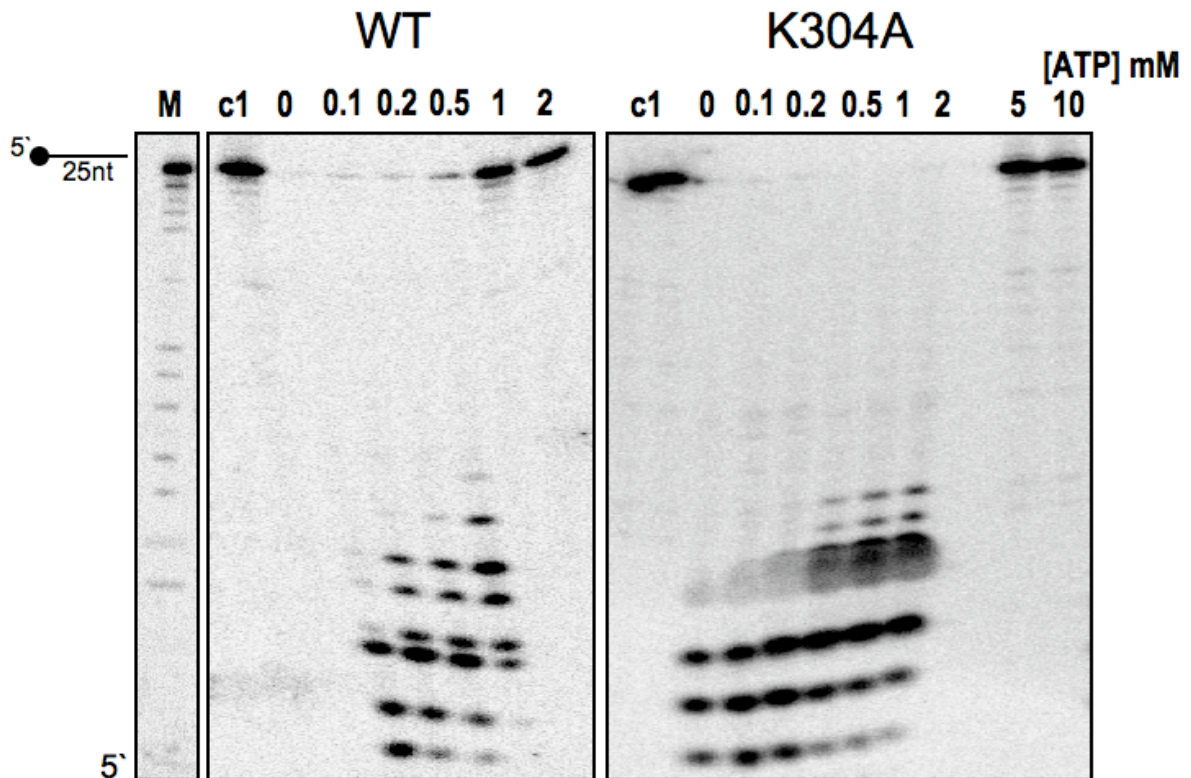
Figure 4.15: DinG nuclease activity at varying ATP concentrations

Representative denaturing 20% (v/v) acrylamide gels showing nuclease activity of WT DinG (left) and K304A DinG (right) with increasing ATP concentrations (indicated above gel). Assays were performed at 37 °C for 60 min with 10 nM ssDNA (X26-50), 500 nM protein and 5 mM MgCl$_2$. M, Maxam-Gilbert A+G ladder from radiolabelled DNA. c1, end-point assay in absence of protein. Each experiment was performed in triplicate.

4.7.1.2 INFLUENCE OF ATP HYDROLYSIS ON DING NUCLEASE ACTIVITY

The experiments performed thus far reveal that ATP binding inhibits the nuclease activity of DinG, but it remains unclear whether inhibition also occurs following ATP hydrolysis. To investigate this further, the inhibitory effect of ATP and ADP on DinG nuclease activity was assessed using WT and K304A DinG. The inability of the K304A mutant to hydrolyse ATP means that binding of ATP or ADP will trap the enzyme in the substrate or product configuration, respectively. Reactions were performed as outlined in Materials and Methods, with 1 mM ATP/ ADP included in WT DinG experiments and 5 mM ATP/ ADP included in K304A DinG experiments.
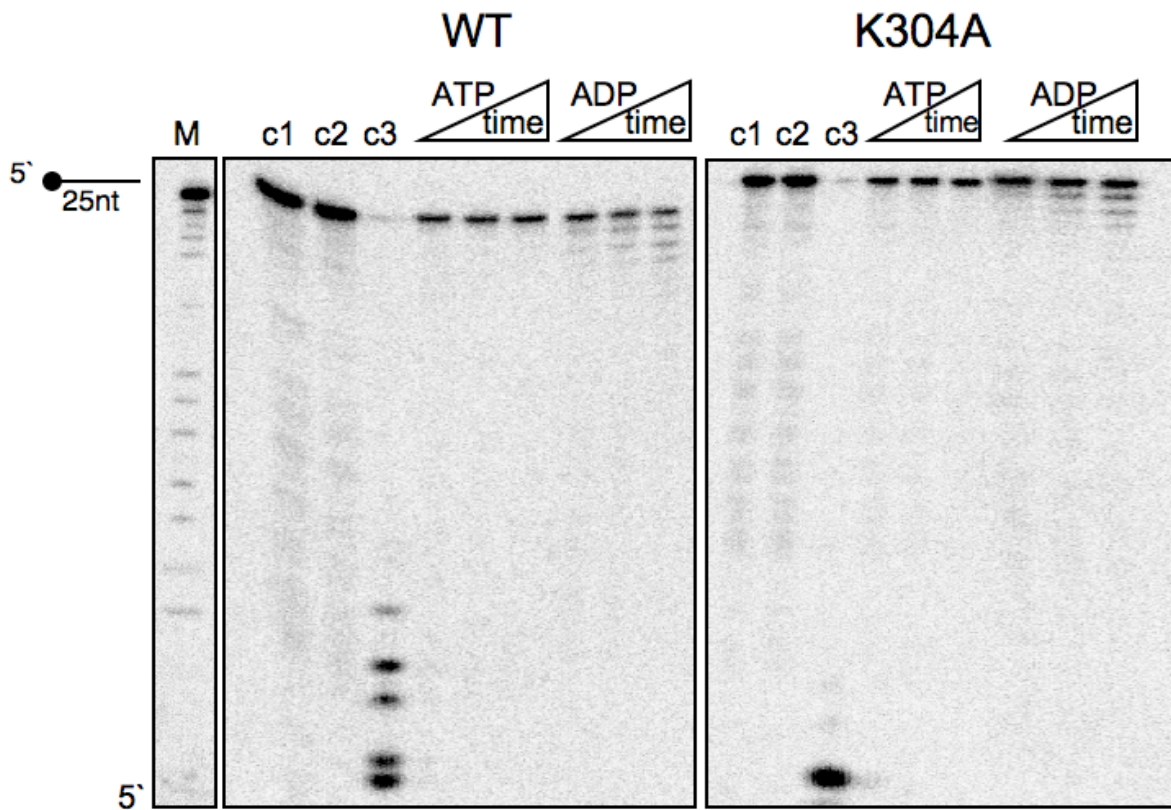
Figure 4.16: DinG nuclease activity in the presence of ATP or ADP
Representative denaturing 20% (v/v) polyacrylamide: TBE gels showing the nuclease activity of 500 nM WT DinG (left) and K304A DinG (right) at 37 °C in the presence of 10 nM DNA (X26-50), 5 mM $MgCl_2$ and ADP or ATP (for WT, 1 mM; for K304A, 5 mM). Time points: 10, 30 and 60 min. M, Maxam-Gilbert A+G ladder of X26-50 DNA. c1, without protein at 60 min; c2, reaction with 1 mM (WT) or 5 mM (K304A) ATP/ $MgCl_2$ at 60 min; c3, reaction with $MgCl_2$ only. Each experiment was performed in triplicate.

The nuclease activity of WT and K304A DinG was inhibited in the presence of ADP and ATP (figure 4.16). Since the WT protein is freely capable of hydrolysing ATP, this result suggests that either ATP binding or hydrolysis (or, rather the presence of ADP) may inhibit the ability of DinG to digest DNA. Given that K304A DinG is not capable of ATP hydrolysis, inhibition of its nuclease activity in the presence of both ADP and ATP (figure 4.16) confirms that both ATP binding and hydrolysis inhibit the ability of DinG to cleave DNA.

4.7.2 SUBSTRATE SPECIFICITY OF DING NUCLEASE ACTIVITY

The structure of DNA substrates cleaved by a nuclease and the polarity with which this occurs is an intrinsic feature that is inextricably linked to its biological role. Thus, determination of the polarity and substrate specificity of the exonuclease domain of DinG may shed light on the *in vivo* function of the protein. Experiments were performed with the WT and K304A mutant (500 nM), in the absence of ATP, as outlined in Materials and Methods. It is highly unlikely that any observed cleavage of complex DNA substrates reflects an initial unwinding event followed by cleavage of the displaced ssDNA since: 1, Helicase assays revealed minimal DNA unwinding capabilities for DinG and 2, DinG-catalysed helicase activity was not detected in the absence of ATP. Maxam-Gilbert A+G markers were generated using the 5' [^{32}P]-end labelled strand from each substrate.

4.7.2.1 POLARITY OF DING NUCLEASE ACTIVITY

The digestion pattern of a 5' [^{32}P]-end labelled ssDNA by DinG (figure 4.14) suggested that the protein possessed 3'-5' exonuclease activity. These experiments also imply that the nuclease domain of DinG digests DNA without DNA binding by the helicase domain. The polarity of DinG nuclease activity and the manner in which it binds DNA were investigated further using a 5' or 3' overhang DNA substrate (figure 4.17).
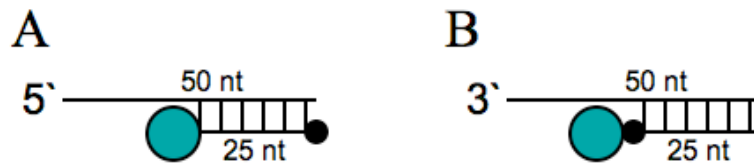
Figure 4.17: DNA substrates used to investigate DinG nuclease polarity

DNA substrates were generated by annealing a 5' [^{32}P]-end labelled strand to a second strand possessing a 25-nucleotide region of complementarity. (A) 5'-overhang comprising B50 and H25 oligonucleotides; (B) 3'-overhang composed of B50 and X26-50 oligonucleotides. Oligonucleotide sequences noted in Appendix 1 (table A1.3); Black circle, 5' [^{32}P]-radiolabel; teal circle, nuclease domain of DinG. The helicase domain of DinG has been omitted because it is unclear how the protein binds DNA.
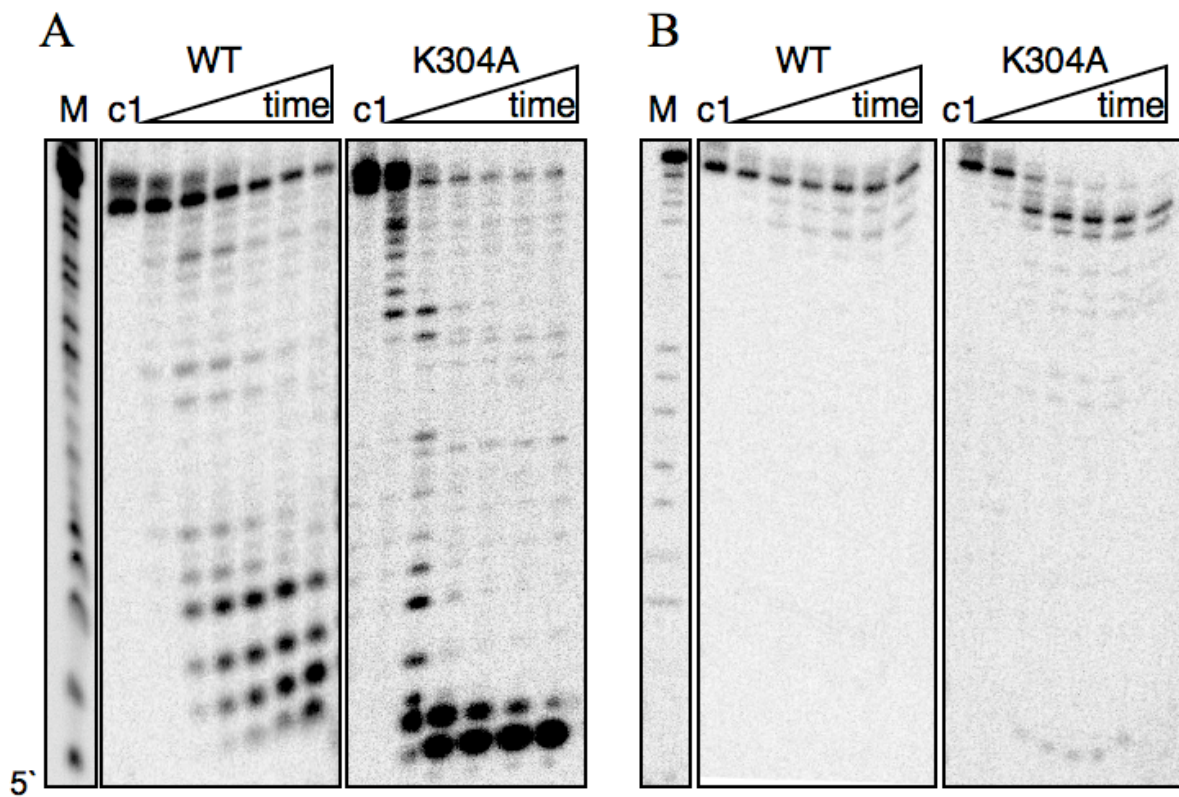
Figure 4.18: Polarity of DinG nuclease activity

Cleavage of a 5'-overhang (A) and 3'-overhang (B) DNA substrate (10 nM) by WT or K304A DinG (500 nM). M, A+G marker of radiolabelled strand (for (A), H25 oligo; for (B), X26-50 oligo). Time course: 1, 10, 20, 30, 40, 60 min. c1, without protein at 60 min. Each experiment was performed in triplicate.

In the presence of the 5'-overhang DNA substrate, DNA cleavage was catalysed by WT DinG and the K304A mutant (figure 4.18 A), indicating that the protein possesses 3'-5' exonuclease activity as predicted by sequence analysis. The K304A mutant displays a faster rate of DNA cleavage, compared to the WT enzyme.

In the presence of the 3'-overhang DNA substrate, WT DinG failed to cleave DNA in a 5'-3' direction as revealed by the absence of very low molecular weight products that would represent immediate cleavage of the 5' radiolabel tracer (figure 4.18 B). The generation of large molecular weight products in the presence of both WT and K304A DinG suggests that the radiolabeled strand was degraded to a limited extent with 3'-5' polarity. It is possible that thermal fraying of the DNA duplex (from the blunt end) permitted limited access of the nuclease domain to the 3' end of the 25 nt radiolabeled strand. Alternatively, this digestion pattern may represent cleavage by a small proportion of DinG molecules of the unlabelled 50 nt strand in a 3'-5' direction, followed by degradation of the radiolabeled strand with the same polarity. While this substrate is not ideal for measuring 5'-3' nuclease activity, these experiments suggest that DinG preferentially cleaves DNA with 3'-5' polarity, as expected for a DnaQ superfamily nuclease.

4.7.2.2   ABILITY OF DING TO CLEAVE DUPLEX DNA

The ability of DinG to cleave dsDNA was investigated using two DNA substrates: a splayed duplex, providing a single-stranded landing site for the protein to bind (figure 4.19 A), and a blunt-end duplex (figure 4.19 B). The blunt-end duplex substrate also contains a 7-nucleotide region lacking complementarity, which represents an ideal 'bubble' substrate for helicases involved in nucleotide excision repair (Evans et al., 1997). Therefore, this substrate tests the ability of DinG to cleave a fully base-paired DNA structure and the ability of DinG to digest a DNA substrate targeted by a prominent repair pathway. Assays were performed as described in Materials and Methods.
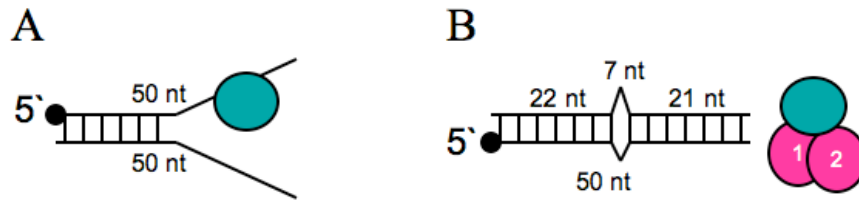
Figure 4.19: Duplex DNA substrates used for DinG nuclease assays

(A) Splayed duplex formed by annealing 5' [^{32}P]-end labelled B50 to X50 DNA; (B) Blunt-ended duplex formed by annealing 5' [^{32}P]-end labelled B50 to 'Bubble7' oligonucleotide. Oligonucleotide sequences noted in Appendix 1 (table A1.3). Black circle, [^{32}P]-radiolabel; pink circle, helicase domain of DinG; teal circle, nuclease domain of DinG. The helicase domain of DinG has been omitted from (A) as it is unclear how the protein binds DNA.
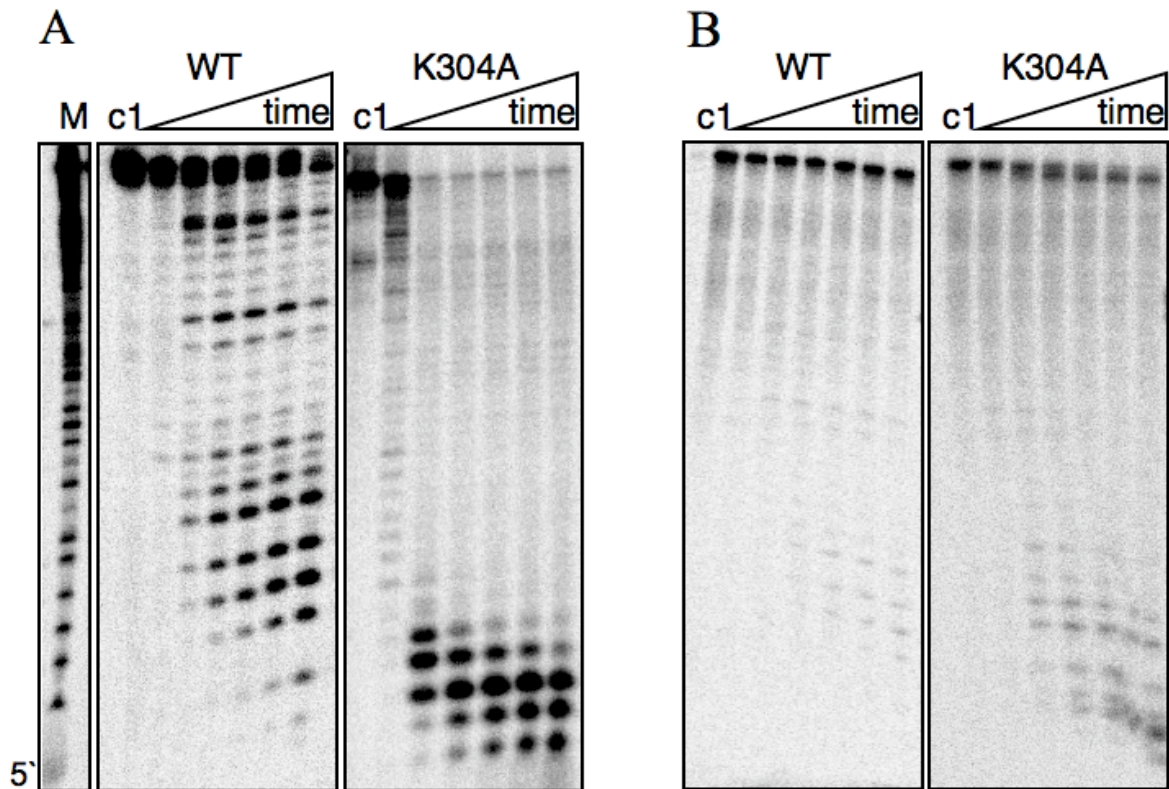
Figure 4.20: Digestion of dsDNA by DinG

Representative denaturing 20% (v/v) polyacrylamide: TBE gels showing DinG-catalysed cleavage of a (A) splayed duplex and a (B) blunt-end duplex containing a 'bubble'. Nuclease assays were performed in the presence of WT or K304A DinG (500 nM), together with 10 nM DNA and 5 mM MgCl$_2$. Time course: 1, 10, 20, 30, 40, 60 min; c1, without protein at 60 min; M, A+G marker of radiolabelled strand. Each experiment was performed in triplicate.

Ove a 60-min time course, DinG partially degraded a splayed duplex (figure 4.20 A). The degradation pattern is consistent with cleavage of the radiolabeled strand

in a 3'-5' direction and the slightly slower rate of cleavage, compared to the 25 nt ssDNA substrate (figure 4.14), may simply reflect the length of the substrate. As with all previously tested substrates, the K304A DinG mutant cleaved the splayed duplex more efficiently that the WT protein.

In contrast, minimal nuclease activity was observed in the presence of the blunt-ended duplex containing a 'bubble' (figure 4.20 B). It is unlikely that the inclusion of a 'bubble' site was responsible for the lack of observed activity, as this would have been observed as a stalling approximately mid-way along the DNA strand. Very small fragments were observed at later time points, particularly in experiments catalysed by the K304A mutant. The fragment pattern is indicative of a small proportion of protein molecules digesting the full length of the 5' [^{32}P]-end labelled strand in the 3'-5' direction.

4.7.3   QUANTIFICATION OF DING NUCLEASE ACTIVITY

The ability of DinG to cleave each DNA substrate was quantified using Image Gauge software (Fuji), which measures the intensity of photostimulated luminescence (PSL) in assigned areas, e.g. areas corresponding to uncut substrate and cleaved products, as represented in figure 4.21. Background PSL levels were subtracted appropriately. This method of quantification can decipher between cleaved and uncleaved DNA, but fails to report on the extent of DNA cleavage. For instance, quantification of the gel image in figure 4.21 concludes that approximately 85% of the DNA was cleaved after 60 min. However, inspection of the gel image reveals that only ~10% of the DNA was completely degraded by DinG. Interpretation of the data is, therefore, best achieved by examination of quantified results together with the original gel image.
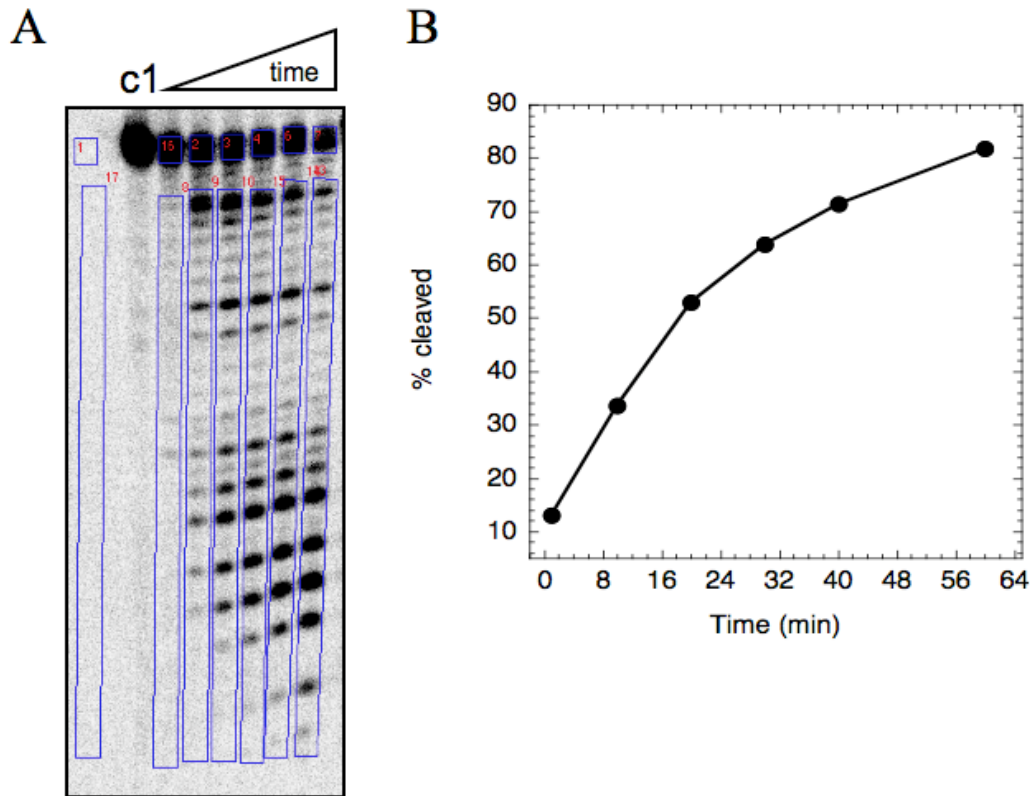
Figure 4.21: Method for quantification of DinG nuclease activity

Denaturing polyacrylamide gel showing DNA cleavage by DinG over time. c1, control. Areas of photostimulated luminescence (PSL), corresponding to uncut substrate and cleaved product, were quantified using Image Gauge software (highlighted by blue rectangles). Background PSL levels were subtracted from these values, highlighted by box 1 (for substrate reactions) and box 17 (for product reactions). (B) Plot of time against percentage cleavage was generated from quantified results.

Figure 4.22 reveals that DinG (both WT and the K304A mutant) cleaves ssDNA, 5' overhang DNA, and splayed duplex DNA substrates with 3'-5' polarity, as reported for other members of the DnaQ family of nucleases (Viswanathan and Lovett, 1999). In all cases, the K304A mutant digested DNA more efficiently than the WT enzyme. DinG exhibited a reduced ability to digest a 3' overhang DNA substrate and a fully base-paired DNA and, consequently, the ability of the enzyme to cleave a 'bubble' substrate could not be determined.
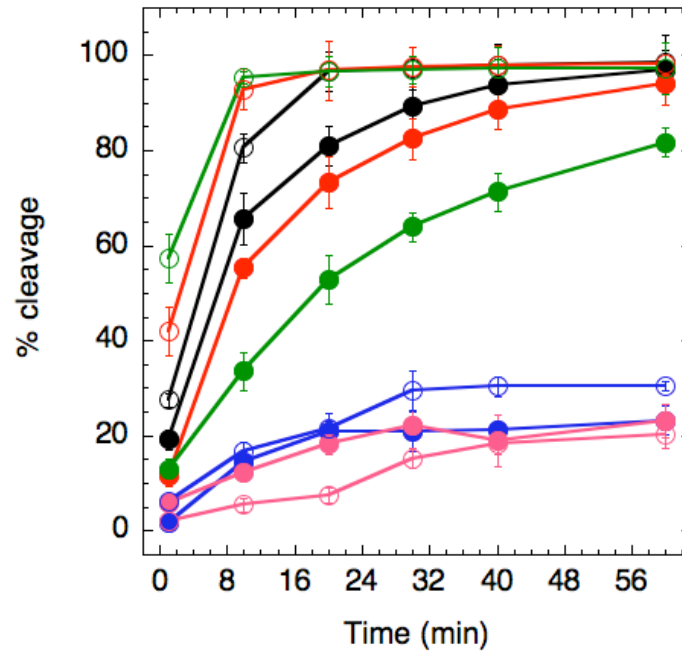
Figure 4.22: Quantification of DinG nuclease activity

Acrylamide gels from nuclease assays were analysed by phosphorimaging and quantified using Image Gauge software. Extent of DNA cleavage (%) was plotted against time (min). ssDNA, black; 5'overhang, red; 3'overhang, blue; splayed duplex, green; blunt-end duplex containing 'bubble' site, pink; closed circles, WT DinG; open circles, K304A DinG.

4.8 DNA BINDING BY DING

In an attempt to understand the basis of ATP-mediated inhibition of DinG nuclease activity, the DNA binding capabilities of the protein in the presence and absence of ATP was investigated. The ability of DinG to bind a DNA substrate containing a 7-nucleotide non-complementary region ('bubble' substrate) was also investigated.

All DNA binding studies were performed by electrophoretic mobility shift assays (EMSAs) and apparent dissociation constants were determined from the processed gel images. Apparent dissociation constants values derived from EMSAs should be regarded with caution as these studies carry a significant margin of error: a proportion of the enzyme-DNA complex may be retained in the gel wells or may migrate into the running buffer. They provide an approximation and conclusions should be drawn based on comparison as opposed to absolute values. Unless a protein-DNA interaction is particularly strong, migration through a gel matrix may cause complex dissociation and, therefore actual dissociation constants tend to be much lower than

those estimated from EMSAs. Since band smearing also occurs during protein-DNA complex formation, dissociation constants values were estimated based on the disappearance of unbound DNA. Since dissociation constants were estimated based on a visual interpretation of the gel images, the apparent dissociation constant $K_{1/2}$ is appropriate.

4.8.1   INFLUENCE OF ATP ON DNA BINDING

To examine the effect of ATP on DNA binding affinity of DinG, varying concentrations of protein were incubated with 10 nM 5' [^{32}P]-end labelled ssDNA (B50 oligonucleotide), alone or in the presence of 5 mM $MgCl_2$ only or 2 mM ATP/ 5 mM $MgCl_2$, at 37 ºC for 10 min prior to separation by native polyacrylamide gel electrophoresis. To quantify any binding influence contributed by the functionality of the nuclease or helicase domain, DNA binding affinities of D10A/ E12A and K304A DinG were also investigated. The apparent dissociation constants (table 4.2) were estimated based upon a purely visual analysis of the gel images (figure 4.23).

The presence of ATP and $MgCl_2$ had no impact on the ability of DinG to bind DNA. The K304A DinG mutant, which lacks the ability to coordinate the β- and γ-phosphoryl groups of ATP, arguably binds DNA with slightly higher affinity than WT DinG. The stronger DNA binding affinity of the K304A mutant may explain its faster rate of DNA cleavage compared to the WT enzyme. In contrast, abrogation of nuclease activity in the D10A/ E12A mutant had minimal impact on the ability of the protein to bind DNA. It should be noted that the DNA was not degraded by the exonuclease activity of DinG, as would have been predicted based on the results shown in figure 4.14 and 4.18; this may be explained by the short incubation time (10 min) of the EMSA binding experiment in comparison to the nuclease assays (60 min).

Table 4.2: Apparent dissociation constants of DinG for ssDNA
Determination of apparent $K_{1/2}$ values for DinG: ssDNA binding. DNA was used at a final concentration of 10 nM in the binding assays.

Substrate	Conditions	Apparent $K_{1/2}$ (µM)		
		WT	D10A/ E12A	K304A
	-ATP, -MgCl$_2$	2.5	1.3	0.15-0.3
50	+ATP, +MgCl$_2$	1.3-2.5	1.3	0.6
	MgCl$_2$ only	2.5	1.3	0.3-0.6
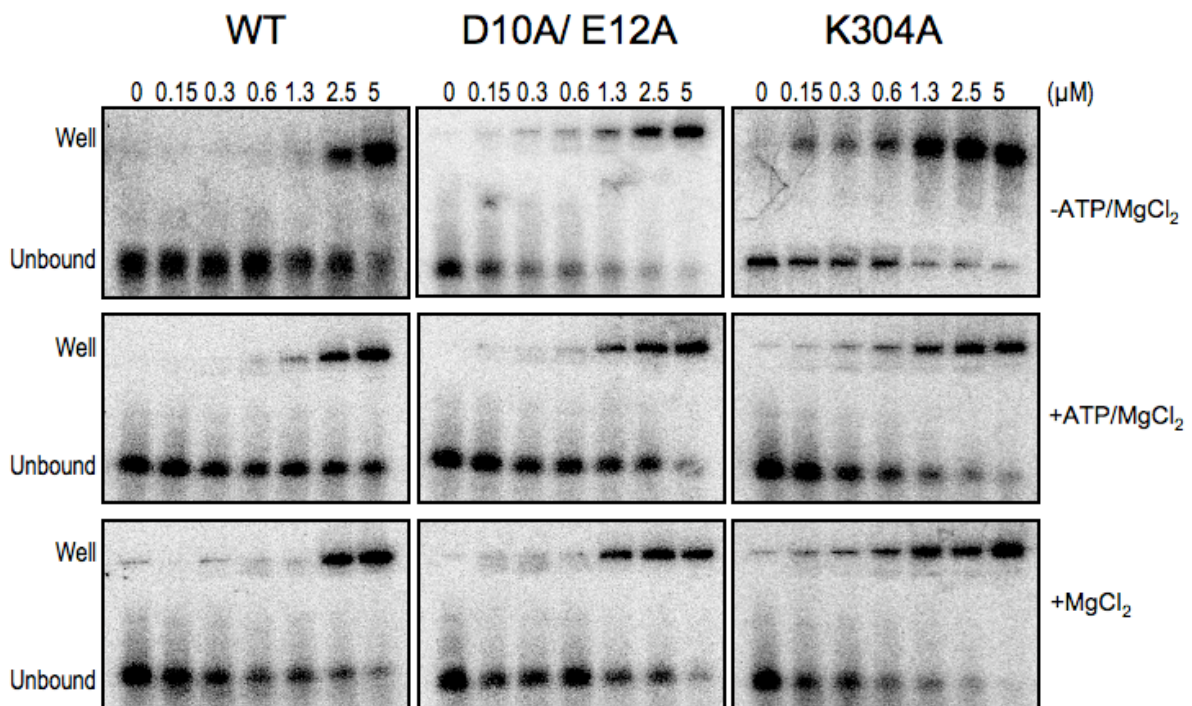
Figure 4.23: ssDNA-binding by DinG
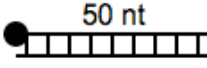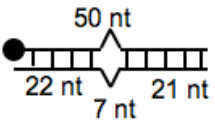
EMSAs were used to monitor the ssDNA binding affinity of DinG (WT, D10A/ E12A, K304A). The enzymes were incubated at the concentrations shown, with ssDNA (5' [^{32}P]-end labelled ssDNA), in the presence and absence of MgCl$_2$ or ATP/ MgCl$_2$, for 10 min at 37 ºC. The samples were run on native 8% (v/v) polyacrylamide: TBE gels to separate unbound substrate from protein-bound DNA. Protein-bound material remained in the wells. Each experiment was performed in triplicate.

4.8.2   ABILITY OF DING TO BIND DNA BUBBLE STRUCTURES

Following DNA damage, 'repair bubbles' form to produce single-stranded regions within duplex DNA (Evans et al., 1997). Previous research has shown that XPD recognises these single-stranded regions and, in this way, plays a fundamental role in

the recognition of DNA damage during the nucleotide excision repair pathway (Wolski et al., 2008). Given the homology observed between DinG and human XPD, it is possible that DinG participates in similar repair processes; therefore, the ability of DinG to recognise 'bubble' structures was investigated. In the following experiments, the binding affinity of DinG for a DNA 'bubble' structure was compared to its affinity for a dsDNA substrate. The 'bubble' substrate was generated by annealing a 5' [^{32}P]-end labelled B50 oligo to a second strand containing a central non-complementary region of 7 nucleotides ('bubble7' oligonucleotide; Appendix 1, table A1.3). The dsDNA substrate comprised two fully-complementary strands (5' [^{32}P]-end labelled B50 annealed to B50comp). Dissociation constants thus reflected a preference from particular structures as opposed to sequences.

Table 4.3: Apparent dissociation constants of DinG for duplex and 'bubble' DNA
DinG was incubated with a fully base-paired duplex or a 'bubble' DNA structure. Apparent dissociation constants ($K_{1/2}$) were determined from EMSA gels. DNA was used at a final concentration of 10 nM.

	Apparent $K_{1/2}$ (µM)		
Substrate	**WT**	**D10A/ E12A**	**K304A**
	2.5	0.15	0.15
	2.5-5	0.3	0.3
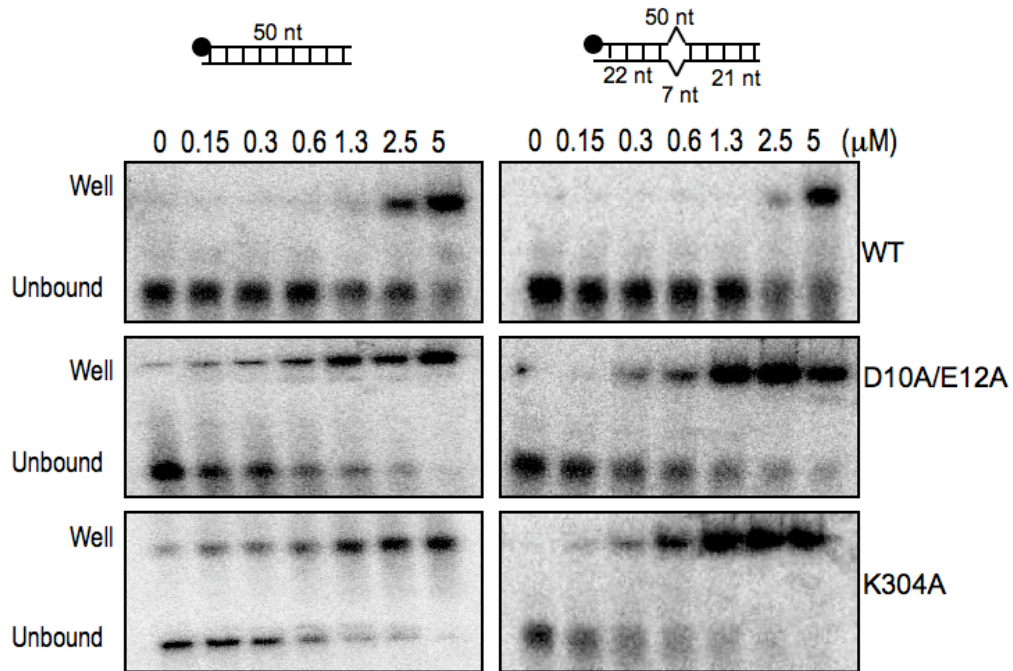
Figure 4.24: Affinity of DinG for duplex and 'bubble' DNA structures

The binding affinity of DinG (WT, D10A/ E12A and K304A) for dsDNA lacking (left) or possessing (right) a bubble region was investigated by EMSAs. Protein concentrations are shown. These data are representative gels generated from triplicate experiments using different protein preparations.

No significant difference was observed in the affinity of DinG for a dsDNA substrate containing or lacking a 'bubble' region (figure 4.24 and table 4.3). This is in contrast to XPD, which has been shown to play a role in the nucleotide excision repair pathway (thesis of Dr J. Rudolf). It has been proposed that the damage recognition role of XPD is linked to its FeS cluster (Wolski et al., 2008). While DinG from *E. coli* possesses an FeS cluster, SarDinG has lost this over the course of evolution following the acquisition of the N-terminal nuclease domain. In this light, it is likely the role of DinG in *S. aureus* has diverged from human XPD and *E. coli* DinG. The affinity of D10A/ E12A and K304A DinG for DNA was greater than WT DinG, consistent with results shown in figure 4.23.

4.9 SENSITIVITY OF *S. AUREUS* LACKING DING TO DNA DAMAGING AGENTS

It has been previously suggested that *E. coli* DinG may play a role in the bacterial SOS response and/ or DNA repair since a slight but reproducible UV

sensitivity of *E. coli* cells lacking or overexpressing *dinG* was observed (Voloshin et al., 2003). However, the *dinG* deletion did not significantly compromise UV resistance in these cells suggesting that the protein may play a redundant role in the repair of UV damage (Voloshin et al., 2003). Furthermore, previous studies in *E. coli* revealed that expression from the *dinG* promoter was up-regulated in response to DNA damage by mitomycin C (Lewis et al., 1992) and nalidixic acid (Van Dyk et al., 2001).

In an attempt to ascertain the *in vivo* function of DinG, we tested the impact of *dinG* disruption in *S. aureus* cells using the TargeTron™ Gene Knockout System (Sigma). The TargeTron™ system works by enabling the insertion of a non-coding group II intron within the gene of interest (as outlined in Materials and Methods), thereby preventing protein translation. Consequently, these cells are not genetic knockouts for DinG, but rather they cannot synthesise the DinG protein.

S. aureus cells, containing functional or disrupted *dinG*, were grown in the presence of various DNA damaging agents (table 4.4), as outlined in Materials and Methods, and the effect on cell growth was determined (figure 4.25). Unpublished work (performed by Richie Gerrard, University of St Andrews) revealed minimal effect of hydrogen peroxide (10 mM), methyl methansulphonate (15 mM) and phleomycin (0.25 mM) on the viability of *S. aureus* cells lacking DinG. To extend the investigation, the effect of UV radiation, mitomycin C (MMC), 4-nitroquinoline 1-oxide (4-NQO) and nalidixic acid on the viability of WT and DinG mutant *S. aureus* cells was tested.

Cells lacking DinG exhibited comparable survival rates to WT cells upon exposure to each DNA damaging agent (figure 4.25), rendering it difficult to elucidate the *in vivo* role of DinG. However, the absence of a discernible phenotype from cells lacking DinG does not exclude the possibility that DinG functions in DNA repair and/ or restoration of DNA replication following DNA damage. Rather, these experiments may indicate that DinG occupies a redundant role in these processes that can be compensated by other proteins. This would be consistent with observations reported for *E. coli* cells lacking DinG (Voloshin et al., 2003). To extend this investigation, it would be necessary to knockout several genes, in addition to DinG, that are expected to function during DNA repair and compensate for the loss of DinG. Redundancy among groups of functionally-related proteins, e.g. DNA helicases or DNA nucleases, is a well-documented phenomenon. For instance, redundancy exists between proteins of the base excision repair (BER) and nucleotide excision repair (NER) pathways from various organisms whereby multiple DNA glycosylases recognise a broad spectrum of DNA

lesions (Legrand et al., 2008). In addition, protein complexes from both NER and BER pathways have been shown to compete for repair of abasic sites within DNA (Torres-Ramos et al., 2000).

Table 4.4: DNA-damaging agents and their *in vivo* consequences

Agent	*In vivo* consequences
Hydrogen peroxide	Induces oxidative stress (Hagensee and Moses, 1989)
Methyl methansulphonate	Alkylating agent; It methylates bases and generates DNA mutations (Beranek, 1990)
Phleomycin	Induces DNA breaks (Sleigh, 1976)
UV radiation	Produce helix-distorting lesions, including CPDs and (6-4) photoproducts (Mitchell and Nairn, 1989; Cadet et al., 2005)
Mitomycin C	Formation of interstrand DNA cross-links (Ueda and Komano, 1984)
4-nitroquinoline 1-oxide	Induces oxidative stress (Arima et al., 2006)
Nalidixic acid	Inhibition of DNA gyrase; Upregulates several genes including *E. coli dinG* (Van Dyk et al., 2001)

A

B
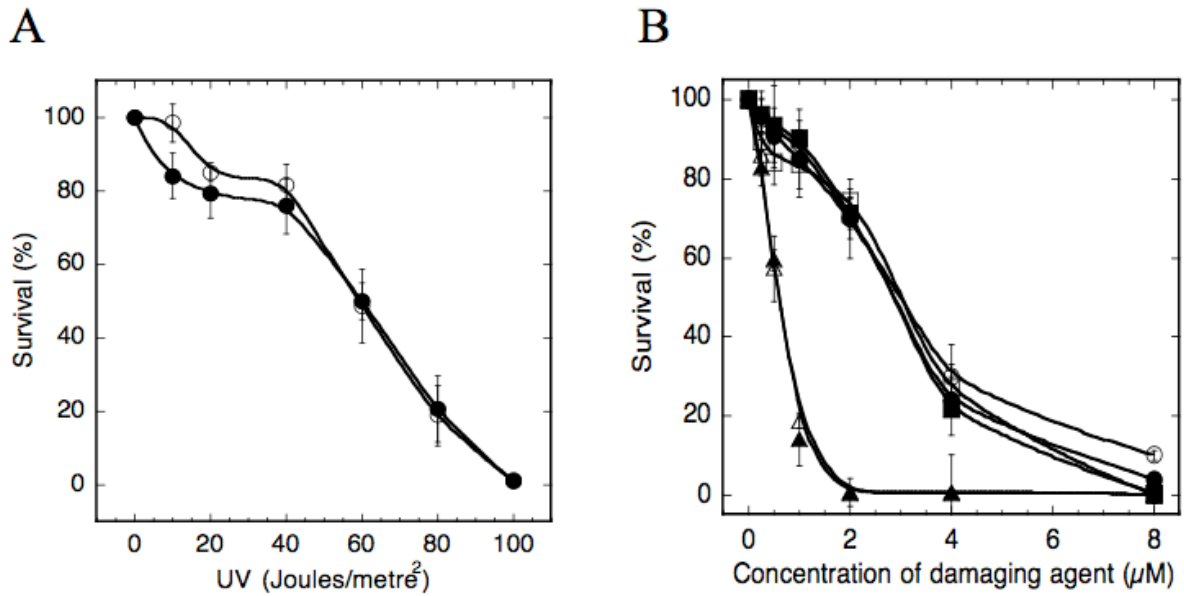
Figure 4.25: Sensitivity of *S. aureus* to DNA damaging agents in the presence or absence of DinG

S. aureus cells were grown to an OD_{600} of 0.65 prior to the addition of DNA damaging agents. Following damage treatment, cultures were grown for 10 min at 37 ºC before being diluted, plated on TSB agar and incubated overnight at 37 ºC. Colonies were counted and plotted. (A) *S. aureus* cells were subjected to 0, 10, 20, 40, 60, 80 or 100 Joules/m^2 of UV radiation. (B) *S. aureus* treatment with 0, 0.25, 0.5, 1, 2, 4, or 8 µM nalidixic acid (circles), mitomycin C (squares) or 4-NQO (triangles). Open shapes, WT *S. aureus*; closed shapes, *S. aureus* lacking DinG. All experiments were performed at least three times with different *S. aureus* cultures.

4.10   DISCUSSION AND CONCLUDING REMARKS

The expression, purification and characterisation of DinG from the Gram-positive bacterium *S. aureus* are described in this chapter. DinG is homologous to *E. coli* DinG and several 5'-3' DNA helicases in humans, including FancJ and XPD. DinG from *S. aureus* differs from each of these homologues in its absence of an FeS cluster and the inclusion of an N-terminal exonuclease domain. In this light, SarDinG appears to have undergone functional divergence from this core group of homologous proteins. The presence of a putative helicase and nuclease domain within the same open reading frame renders this protein particularly attractive for investigation. In addition, characterisation of DinG may lend potentially useful insights into clinically-important human proteins, including XPD which is associated with xeroderma-pigmentosum (XP) and trichothiodystrophy (TTD), owing to their shared ancestry.

DinG demonstrated the ability to hydrolyse ATP in the presence of single- and double-stranded DNA. ATPase activity was also observed in the absence of DNA,

at a rate approximately 4-fold less than in the presence of ssDNA. As expected, the Walker A box mutant (K304A DinG) was unable to catalyse the conversion of ATP to ADP and free phosphate. DNA-independent ATPase activities have been observed for *E. coli* DinG (Voloshin et al., 2003) and other DNA helicases, including *E. coli* RecQ (Wickner et al., 1974; Zittel and Keck, 2005) and DnaB (Wickner et al., 1974). However, in these cases, the DNA-independent rate of ATP hydrolysis was significantly lower than that observed in the presence of nucleic acids. Helicases are motor proteins that couple the energy from NTP binding and hydrolysis to DNA translocation and unwinding. The *in vivo* function of a DNA helicase is, therefore, related to its ability to bind ss- and ds-DNA, under the control of NTP binding and hydrolysis. The way in which DNA binding and ATP hydrolysis are coupled remains poorly understood, but studies in SF1 DNA helicases, including PcrA and Rep, suggest that helicase motif III plays a vital role (Dillingham et al., 1999). In PcrA, this motif contacts both ATP, via a highly conserved glutamine residue, and ssDNA, via a C-terminal aromatic-rich loop. In particular, residues Q254, W259 and R260 have been implicated in coupling ATP hydrolysis to DNA unwinding through direct contacts with both ATP and DNA (Dillingham et al., 1999). Although these residues are highly conserved among RecQ family members, examination of the sequence between motifs II and III of DinG and related proteins failed to reveal a similar grouping of residues. However, the structure of XPD from *S. acidocaldarius* shows that motif III forms a loop that may be ideally positioned for interaction with ssDNA and ATP (Liu et al., 2008). Without a structure for DinG, it is difficult to explain the poor coupling observed between DNA binding and ATP hydrolysis.

The presence of conserved helicase motifs, together with previous studies of the DinG protein from *E. coli*, prompted an investigation into the helicase functionality of SarDinG. *E. coli* DinG and other homologous DNA helicases including human XPD, have been shown to unwind a broad range of DNA substrates with 5'-3' polarity (Voloshin et al., 2003; Rudolf et al., 2006). Camerini-Otero and coworkers reported complete ATP-dependent unwinding of a substrate composed of a 55-mer oligonucleotide annealed to M13mp18 ssDNA within 2.5 min, at 30 ºC, in the presence of 100 nM *E. coli* DinG (Voloshin et al., 2003). In contrast, DinG from *S. aureus* displayed very poor (if any) helicase activity. These assays were performed at micromolar concentrations of protein, at 37 ºC, and in the presence of ATP/ $MgCl_2$. Neither mutation of DinG to abolish nuclease activity (D10A/ E12A) nor mutation to

completely remove the N-terminal nuclease domain (Δnuclease) influenced the helicase activity of DinG. The effect of pH was also investigated given the alkaline-dependent nature of *E. coli* DinG helicase activity (Voloshin et al., 2003). However, the range of pH values tested had no impact on the ability of SarDinG to unwind DNA.

One possible explanation for the inability of SarDinG to unwind DNA relates to its evolutionary divergence from homologous proteins with the concomitant loss of the FeS cluster. The FeS cluster of human XPD and FancJ, *E. coli* DinG, and *S. cerevisiae* Rad3 is essential for helicase activity (Rudolf et al., 2006; Pugh et al., 2008; Ren et al., 2009). Clinically-relevant mutations have now been linked to disruption of the FeS cluster in human XPD and FancJ (Dubaele et al., 2003; Levran et al., 2005; Rudolf et al., 2006), providing further support for the crucial role played by this domain in these proteins. It can be concluded that DinG from *S. aureus* lacks the ability to unwind DNA, despite the presence of the conserved helicase motifs characteristic of SF2 DNA helicases. Similar findings have been reported for several DNA repair proteins, including archaeal XPB1 and XPB2 (Richards et al., 2008a) and Rad54 (Swagemakers et al., 1998).

The streptavidin displacement assay, described by (Byrd and Raney, 2004), was employed to determine the ability of DinG to translocate along DNA. However, DinG was unable to disrupt the biotin: streptavidin linkage. Similar findings have been reported for several DNA repair proteins, including archaeal XPB1 and XPB2 (Richards et al., 2008a) and archaeal XPD (thesis of Dr J. Rudolf, University of St Andrews), and does not exclude the possibility that DinG can track along DNA. This experiment, however, does reveal that DinG lacks the ability to displace proteins that bind tightly to DNA, e.g. the single-stranded DNA binding protein (SSB), during DNA metabolism including replication, repair and recombination.

Figure 4.1 reveals that DinG possesses an N-terminal extension, with homology to the DnaQ family of 3'-5' exonucleases. The cleavage patterns obtained upon incubation of DinG with a variety of 5' [^{32}P]-end labelled DNA substrates are consistent with 3'-5' exonuclease activity. The ability of DinG to digest ssDNA suggests that the nuclease domain can bind and degrade DNA independent of DNA binding by the helicase domain. A caveat of this proposal arises when we consider the fact that DinG degraded the radiolabeled strand of a 5' overhang. In this instance, the data suggests that the nuclease domain binds the radiolabeled strand at the duplex DNA junction and digests the DNA in a 3'-5' direction through a fully base-paired region.

This raises the question as to why DinG cannot degrade a 3' overhang substrate to completion by firstly digesting the unlabelled 50 nt strand in a 3'-5' direction followed by digestion of the radiolabeled strand with the same polarity. Although DinG appears to lack the ability to unwind DNA, its nuclease activity achieves the same conclusion as a functional helicase that unwinds DNA with 5'-3' polarity. Further work is required to fully understand the mechanism by which DinG binds and degrades DNA.

The presence of ATP/ ADP was shown to inhibit the exonuclease activity of DinG. Given that ATP binding and hydrolysis takes place within the helicase domain of the protein, DinG nuclease activity appears to be under the control of the helicase domain. It is possible that ATP/ ADP-mediated conformational changes within the helicase domain alter the position of the nuclease domain such that DNA cleavage is prohibited. In this light, it must be presumed that the nuclease activity of DinG is inactive by default given the high cellular concentration of ATP. Upon ATP hydrolysis and subsequent release of ADP, the 'empty' DinG protein digests DNA. It may be anticipated that 1 nucleotide of DNA is digested by DinG per cycle of ATP binding and hydrolysis. The hypothesis that the helicase domain controls the nuclease domain is further supported by the fact that mutation of a residue within the helicase domain of the protein (to create K304A DinG) influenced the rate of cleavage. This mutant was shown to bind DNA with greater affinity compared to the WT protein, providing an explanation for its faster rate of DNA cleavage. The data presented in this chapter provides a glimpse into the biochemical activities of DinG and further experiments are required to delineate the precise details of how this protein digests DNA.

ATP binding and hydrolysis provides a convenient mechanism by which the cell can regulate the exonuclease activity of DinG to prevent continued degradation of DNA. ATP-mediated regulation of DinG nuclease activity is reminiscent of the χ-regulated RecBCD helicase-nuclease complex in bacteria; upon encountering χ sequences, the 3'-5' exonuclease activity of the RecBCD complex is attenuated to prevent uncontrolled degradation of DNA at double-strand breaks (Anderson and Kowalczykowski, 1997). In addition, ATP has been shown to regulate the nuclease activities of an archaeal helicase-nuclease protein, Nar71, which is thought to function during recombination-dependent replication; ATP binding switches Nar71 from a functional nuclease to a helicase, whereas release of ADP converts Nar71 to an active nuclease (Guy et al., 2004).

To link the biochemical data reported in this chapter to the *in vivo* function of DinG, a genetic approach was taken. *S. aureus* cells lacking a functional copy of DinG were generated by Dr S. Graham (University of St Andrews). The sensitivity of these cells, in comparison to WT *S. aureus* cells, to various DNA damaging agents, including UV radiation, mitomycin C, 4-nitroquinoline 1-oxide and nalidixic acid, was examined. *S. aureus* cells lacking DinG displayed a level of resistance to each agent comparable to that of WT cells, indicating that SarDinG may be dispensable for the repair of DNA lesions or restoration of DNA replication disrupted by damage. This is consistent with the redundancy of *E. coli* DinG during repair of UV-irradiated cells (Voloshin et al., 2003). These results do not exclude the possibility that DinG is involved in the repair of lesions caused by these damaging agents. Rather, these results may reflect the redundancy of *S. aureus* repair pathways. While human XPD and human FancJ (Cantor et al., 2004; Rudolf et al., 2006) have been implicated in various DNA repair pathways, it is difficult to extrapolate these roles to DinG, which has diverged in function as a result of the loss of an FeS cluster and acquisition of an N-terminal 3'-5' exonuclease domain. Further research into the biological role of DinG would necessitate the knockout of multiple repair proteins that are upregulated in the absence of DinG. Immediatley prior to the submission of this thesis, a study on the DinG protein from *E. coli* implicated DinG, together with helicases Rep and UvrD, in promoting replication through oppositely oriented highly transcribed ribosomal operons (Boubakri et al., 2009). Ribosomal operons have a tendency to generate non-translated RNA, which lead to the formation of R-loops. These R-loops block the progression of RNA polymerase, causing replication to stall. The study has shown that DinG, Rep and UvrD act together to remove R-loops and to displace stalled RNA polymerase. The ability of DinG to degrade RNA structures needs to be investigated to ascertain whether this protein performs a similar role to the homologue from *E. coli.*

CHAPTER 5

BIOCHEMICAL CHARACTERISATION OF HEL308

5.1   INTRODUCTION

DNA helicases are enzymes that couple ATP hydrolysis to DNA translocation and unwinding and are essential during DNA metabolism. More recently, DNA helicases have been implicated in the stabilisation of stalled replication forks (Bachrati and Hickson, 2008). Replication forks may stall as a result of template damage whereby a lesion, bound proteins, or secondary structures physically impede the progression of the fork during replication, or the fork may stall due to a lack of nucleotide precursors (Hishida et al., 2004; Heller and Marians, 2006a). Template damage and replication stalling are associated with genetic mutation and chromosomal rearrangements and, therefore, by facilitating the resumption of replication, DNA helicases are critical to maintaining genomic integrity (Heller and Marians, 2006a).

The RecQ helicases are believed to play key roles in replication fork repair (Hishida et al., 2004; Bachrati and Hickson, 2008), and includes RecQ (bacteria), Rqh1 and Sgs1 (yeast) and BLM, WRN, and RecQ4 (metazoans). In humans, the fundamental role played by these proteins is highlighted by the fact that mutation of RecQ helicases are often associated with rare genome instability disorders, including Rothmund-Thomson syndrome (RecQ4 helicase), Bloom's Syndrome (BLM helicase), and Werner's Syndrome (WRN helicase).

Despite their importance in humans, homologues of the RecQ family of helicases are not found in the archaea. However, a second group of DNA helicases, known as the Mus308 family, are present in archaea and metazoans (figure 5.1) and have been shown to perform many RecQ-like functions (Harris et al., 1996). The *mus308* locus was initially identified in *D. melanogaster* during an *in vivo* genetic screen for resistance to DNA crosslinking agents. The *D. melanogaster mus308* gene is composed of an N-terminal helicase and C-terminal polymerase domain (Harris et al., 1996), akin to that observed in the orthologous PolQ helicase in humans. A subset of the Mus308 helicase family, called the Hel308 helicases, was identified in both archaea and metazoans based on its homology to the helicase domain of the *D. melanogaster* Mus308 (Dme Mus308) protein (figure 5.1). Mutations in Dme Hel308 cause sensitivity to cisplatin, a potent inhibitor of replication forks, suggesting that it may also play a role in DNA repair and replication restart (Boyd et al., 1990; Guy and Bolt, 2005).

Archaeal Hel308, a SF2 DNA helicase, has significant sequence similarity to human Hel308 and mouse PolQ helicases but no homology was detected with the RecQ helicases. Biochemical studies have shown that archaeal Hel308 possesses several properties similar to that observed for human Hel308: the ATPase activity of Hel308 was stimulated by ssDNA but not dsDNA and archaeal Hel308 preferentially binds and unwinds branched substrates that arise at stalled replication forks (Guy and Bolt, 2005; Fujikane et al., 2006; Richards et al., 2008b). However, the directionality with which archaeal Hel308 translocates along and unwinds DNA is the subject of some controversy. Since the directionality of unwinding dictates whether Hel308 unwinds the leading or lagging strand of a stalled fork, this question has important implications for *in vivo* role of the protein. While the Hel308 homologue from *M. thermautotrophicus* (Mth Hel308) (Guy and Bolt, 2005) and *P. furiosus* (Pfu Hjm) (Fujikane et al., 2006) have been reported to unwind DNA with strict 3'-5' polarity, Shen and coworkers observed a bidirectional unwinding activity for the homologue from *S. tokodaii* (Sto Hjm) (Li et al., 2008). The work performed on archaeal Hel308 has culminated in the proposal that this DNA helicase is likely to play a crucial role in maintaining genome stability in archaeal species (Guy and Bolt, 2005; Richards et al., 2008b). This is particularly important given the absence of RecQ homologues in the archaea.
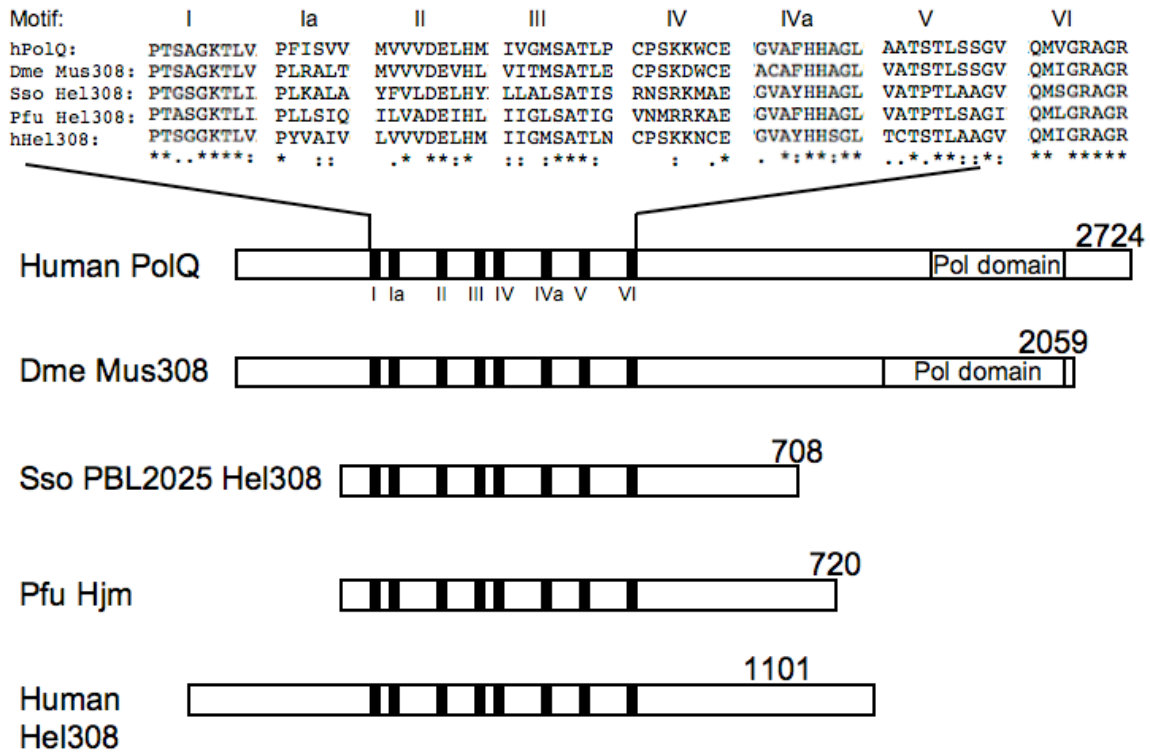
Figure 5.1: Alignment of PolQ, Mus308 and Hel308 SF2 DNA helicases

Cartoon of the domain arrangement of helicases related to Hel308 from *S. solfataricus* PBL2025 strain (Sso PBL2025 Hel308), including human PolQ (hPolQ), *D. melanogaster* Mus308 (Dme Mus308), *P. furiosus* Hjm (Pfu Hjm) and human Hel308 (hHel308). The position and sequence of conserved helicase motifs are indicated. Numbers to the right of each domain structure refers to the number of amino acids in each protein. "Pol domain" indicates the approximate position of the polymerase domain.

In this chapter, the directionality of DNA translocation and unwinding by Hel308 from *S. solfataricus* PBL2025 (Sso PBL2025 Hel308) will be investigated and, using site-directed mutagenesis, the role of the five structural domains of Hel308 will be assessed. The proposed role of Hel308 in replication fork restart raises the question of whether the protein can function in the presence of DNA binding proteins: upon stalling, single-stranded regions of the template will be bound by protecting proteins, such as the single-stranded DNA binding protein (SSB), and it will be intriguing to assess the functionality of Hel308 under biologically-relevant conditions. Since helicases often function as part of a multi-subunit complex or in association with accessory proteins (Tuteja and Tuteja, 2004), this chapter aims to elucidate potential interacting partners of Hel308. Unless otherwise stated, Hel308 refers to Sso PBL2025 Hel308.

5.2   SITE-DIRECTED MUTAGENESIS OF HEL308

Archaeal Hel308 comprises five structural domains, four of which create a channel that accommodates the 3' strand of DNA following duplex separation by a β-hairpin loop on domain 2 (figure 5.2 A) (Buttner et al., 2007). The displaced strand interacts with all five domains before finally contacting domain 5: domains 1 and 2, or the RecA-like motor domains, bind and hydrolyse ATP, and domains 3-5 are predicted to couple the free energy of ATP hydrolysis to DNA translocation and unwinding via conformational changes (Buttner et al., 2007).
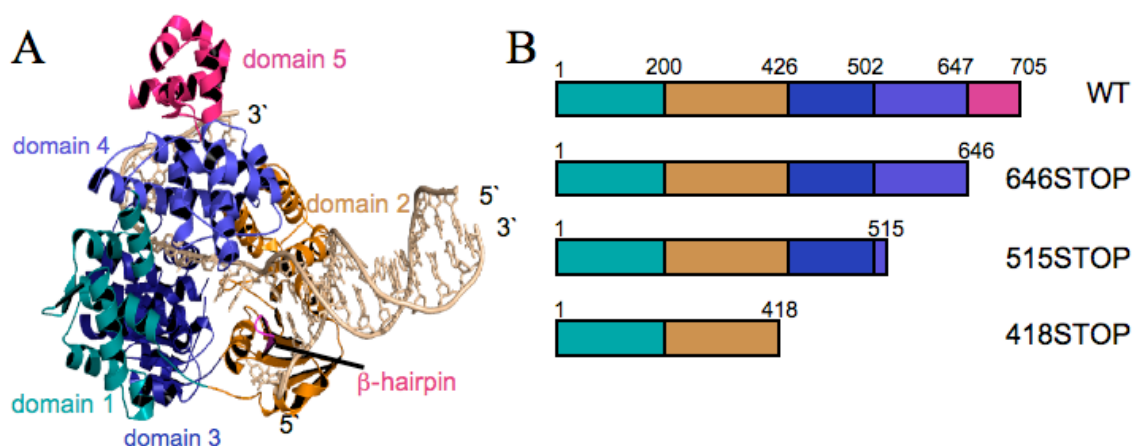
Figure 5.2: Structural domains of Archaeal Hel308

(A) Structure of Hel308 from *A. fulgidus* (Afu Hel308) in complex with DNA. (B) Constructs were generated to encode WT Hel308 and Hel308 lacking domains 5 (646STOP), domains 4-5 (515STOP) and domains 3-5 (418STOP). Domains coloured as (A). Numbers refer to amino acids.

To investigate the proposed role of each structural domain, site-directed mutagenesis was employed (using the primers shown in Appendix 1, table A1.2) to introduce stop codons in the Hel308 sequence such that the protein was expressed without domain 5 (construct '646STOP'), without domains 4 and 5 (construct '515STOP'), or without domains 3-5 (construct '418STOP') (figure 5.2 B). The 646STOP construct was generated by Dr J. Richards and has been characterised previously (Richards et al., 2008b). Thus, this chapter focuses on the biochemistry of 515STOP and 418STOP Hel308.

5.3 EXPRESSION AND PURIFICATION OF HEL308

The pDEST14 vector containing the gene encoding Sso PBL2025 Hel308 was provided by the Scottish Structural Proteomics Facility (University of St Andrews). The protein was expressed in *E. coli* BL21 expression cells and purified to homogeneity using a heat step (65 °C), followed by a two-step purification scheme consisting of nickel-chelating and size exclusion chromatography (figure 5.3 A-C). From a 6L culture, 60 mg of protein was purified. MALDI-TOF and ESI-TOF mass spectrometry were used to verify the identity and molecular mass of the purified protein, respectively (figure 5.4). The calculated molecular weight of the polyhistidine-tagged monomer was 84448.0 Da, which was in good agreement with the actual molecular weight of 84456.0 Da. The 515STOP and 418STOP Hel308 proteins were purified as described for the WT protein (figure 5.3 D).
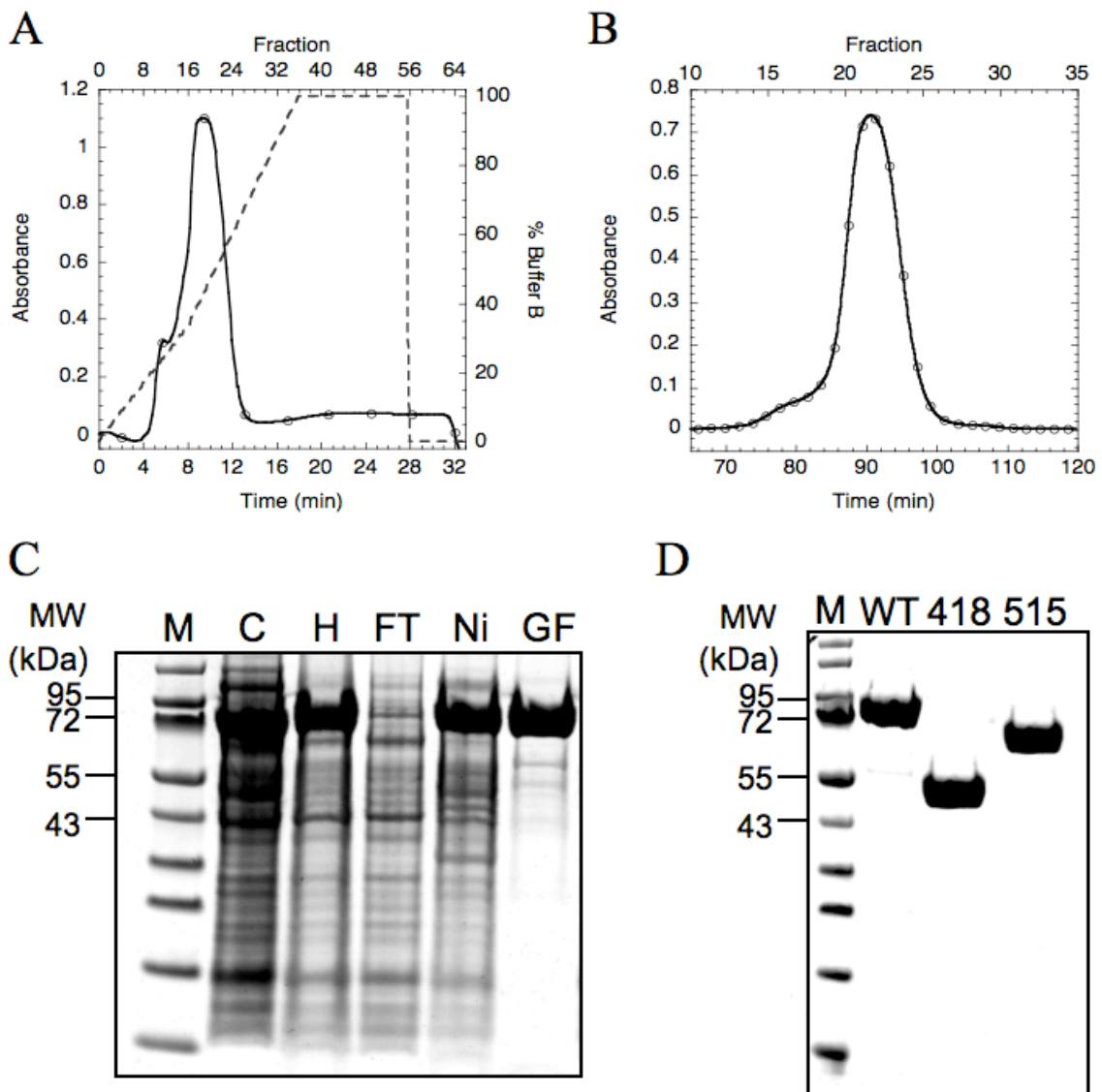
Figure 5.3: Purification of Hel308

Chromatograms from the purification of WT Hel308 by nickel-chelating (A) and gel filtration (B) chromatography are shown. Samples from each stage of the purification of WT Hel308 were analysed by SDS-PAGE (C): M, marker; C, crude lysate; H, after heat treatment at 65 ºC; FT, flow-through from nickel-chelating column; Ni, sample eluted from nickel-chelating column; GF, sample eluted from gel filtration column. (D) Purified WT, 515STOP ('515') and 418STOP ('418') Hel308 were analysed by SDS-PAGE.
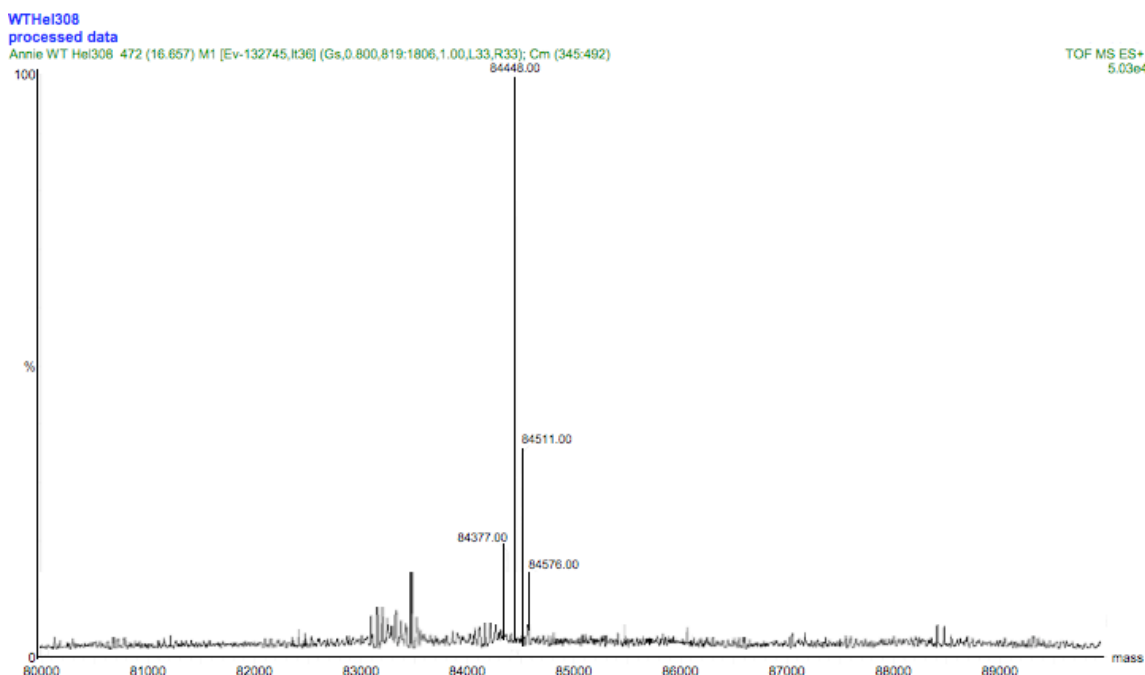
Figure 5.4: Mass spectrometry of WT Hel308

ESI-TOF mass spectrometry revealed that WT Hel308 has a molecular mass of 84448.0 Da, which was in good agreement with the theoretical mass of 84456.0 Da.

5.4 ATP HYDROLYSIS BY HEL308

Given that helicases require nucleotide binding and hydrolysis for strand separation, the ability of WT, 515STOP and 418STOP Hel308 to hydrolyse ATP was investigated using the malachite green assay described in Materials and Methods. Protein (500 nM) was incubated at 45 ℃ with 0.5 nM φX174 virion ssDNA or φX174 RFI dsDNA and reactions were initiated by the addition of 1 mM ATP/ MgCl$_2$. Incubation of the samples with malachite green provided a colorimetric measure of the amount of phosphate released during ATP hydrolysis; this was quantified by recording the absorbance at 650 nm.

Hel308 lacking domains 4-5 (515STOP) or domains 3-5 (418STOP) catalysed ssDNA-dependent ATP hydrolysis with rates 160 % and 209 % of that observed for WT Hel308 (figure 5.5 and table 5.1). Both 515STOP and 418STOP Hel308 possess motor domains 1 and 2, and therefore it is unsurprising that both hydrolyse ATP. However, the increased rate of hydrolysis in comparison to WT is reminiscent of data presented in Chapter 3, whereby mutations that abolished the helicase activity of archaeal XPD also resulted in elevated rates of ATP hydrolysis. These results may suggest that Hel308 lacking domains 3-5 cannot couple ATP

hydrolysis to DNA translocation and unwinding. The ATPase activity of Hel308 is not significantly stimulated by dsDNA, and ATP is turned over very slowly in the absence of DNA (0.8 pmol phosphate/ min/pmol WT Hel308).

Table 5.1: Rate of ATP hydrolysis by WT, 515STOP, and 418STOP Hel308
The rate of ATP hydrolysis was determined from figure 5.5, based on the amount (in pmoles) of phosphate released/ min/ pmol Hel308 in the presence or absence of ss/ ds DNA.

Protein	Rate of ATP hydrolysis (pmoles phosphate released/ min/ pmoles Hel308) ± s.e.		
	ssDNA	dsDNA	No DNA
WT Hel308	21 ± 1.2	7 ± 1	0.8 ± 0.1
515STOP Hel308	33 ± 1.3	3 ± 0.4	0.8 ± 0.1
418STOP Hel308	44 ± 1.5	3 ± 0.5	0.7 ± 0.1
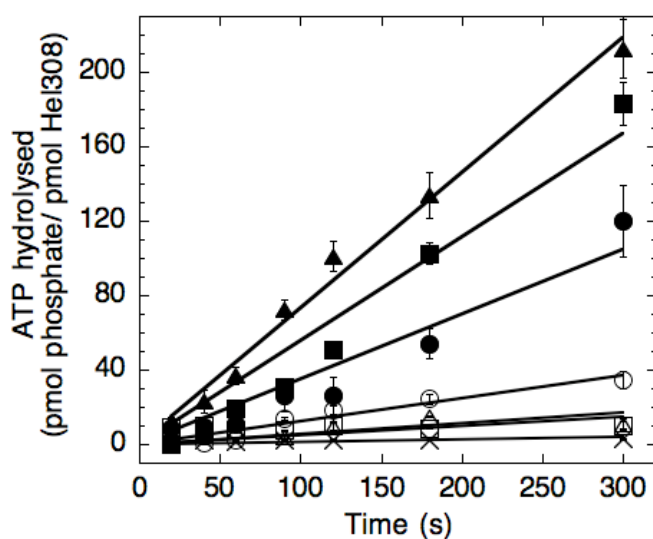
Figure 5.5: ATP hydrolysis by WT, 515STOP and 418STOP Hel308

Reactions were performed at 45 ºC in the presence of 0.5 nM φX174 virion ssDNA (filled shapes) or φX174 RFI virion dsDNA (open shapes) and 500 nM Hel308. WT, circles; 515STOP, squares; 418STOP, triangles; WT in absence of DNA, cross. Data for 515STOP and 418STOP in the absence of DNA have been omitted for clarify. Data represent the mean of triplicate measurements with standard errors shown.

5.5 POLARITY OF DNA UNWINDING BY HEL308

While the Hel308 homologue from euryarchaea *P. furiosus* (Pfu Hjm) and *M. thermautotrophicus* (Mth Hel308) unwind DNA in a 3'-5' direction (Guy and Bolt, 2005; Fujikane et al., 2006; Li et al., 2008), Sto Hjm reportedly exhibits a bidirectional helicase activity (Li et al., 2008). Shen and coworkers have suggested that these results may reflect the evolutionary divergence between Euryarchaeota (*M. thermautotrophicus* and *P. furiosus*) and Crenarchaeota (*S. tokodaii*). The directionality with which Hel308

unwinds DNA dictates whether the leading or lagging strand (or indeed both) of a replication fork is unwound *in vivo*. To resolve this controversy, the directionality of DNA unwinding by Hel308 from the crenarchaeote *S. solfataricus* (strain PBL2025) was investigated. The amino acid sequence of Sso PBL2025 Hel308 and Sto Hjm exhibit 63% identity and 81% similarity, and therefore it was anticipated that they would share fundamental properties, such as the directionality of DNA translocation and unwinding.

The role of domain 5 of Hel308 also remains controversial: one study found that domain 5 acts as an autoinhibitory domain to limit the helicase activity of Hel308 (Richards et al., 2008b), whereas a second study reported that domain 5 increased the helicase activity of Hel308, serving to couple ATP hydrolysis to DNA unwinding (Woodman et al., 2007). The structure of Afu Hel308 (figure 5.2) revealed that domain 5 bound the 3' tail of DNA (Buttner et al., 2007). The authors highlighted that this could explain how the protein preferentially binds and unwinds branched DNA substrates, such as replication forks: domain 5 could bind the single strand in front of the ss/ dsDNA junction, positioning the protein to unwind the lagging strand. The role of this domain in determining helicase directionality will also be investigated using purified 646STOP Hel308 protein (provided by Dr J. Richards).

Helicase assays were performed at 60 ºC in the presence of 500 nM Hel308, as outlined in Materials and Methods. Two DNA substrates (figure 5.6), each consisting of a 25-base pair duplex region and either a 25-nucleotide 3' extension (3' overhang) or a 25-nucleotide 5' extension (5' overhang) were constructed to monitor 3'-5' or 5'-3' unwinding, respectively.
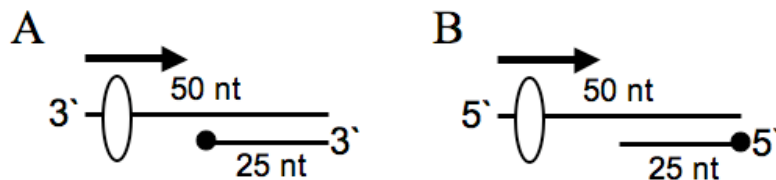
Figure 5.6: Substrate design for helicase assays

(A) 3'-overhang (oligos B50 and X26-50) and (B) 5'-overhang (oligos B50 and H25) DNA substrates were designed to observe DNA unwinding in the 3'-5' and 5'-3' direction (shown by black arrow), respectively, by Hel308 (white oval). Black circle, 5' [^{32}P]-radiolabel. For oligonucleotide sequences, see Appendix 1 (table A1.3).

Hel308 catalysed ATP-dependent DNA unwinding only in the presence of the 3'-overhang substrate (figure 5.7), confirming that Sso PBL2025 Hel308 is a unidirectional 3'-5' DNA helicase. This is in agreement with the published data on euryarchaea Mth Hel308 (Guy and Bolt, 2005) and Pfu Hjm (Fujikane et al., 2006), and supports a role for Hel308 in unwinding the lagging, and not leading, strand of stalled replication forks. These results contradict the hypothesis proposed by Shen and coworkers, which suggested that the reported bidirectional unwinding activity of StoHjm reflected the evolutionary divergence of the Euryarchaeota and Crenarchaeota (Li et al., 2008).
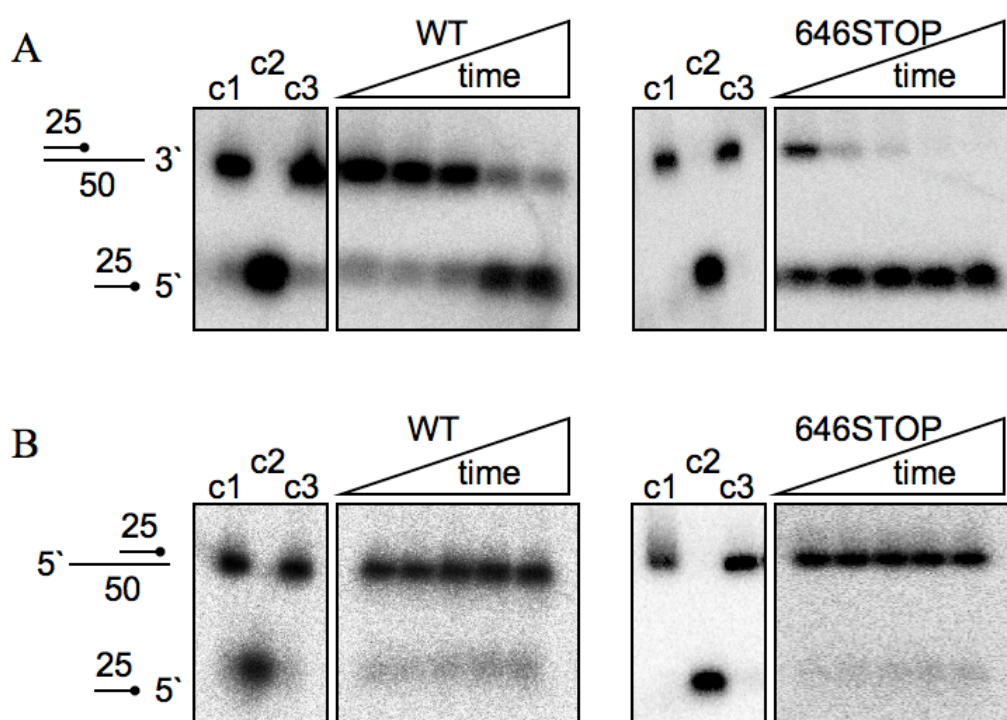
Figure 5.7: Hel308 unwinds DNA with 3'-5' polarity

WT or 646STOP Hel308 (500 nM) were incubated at 60 °C with a 3' overhang (A) or a 5' overhang (B) DNA substrate. Time course: 10, 30, 60, 120, 180 s. Black circle, [^{32}P]-radiolabel; c1, end point assay without protein; c2, size marker; c3, end point assay without ATP/ MgCl$_2$. Length of oligonucleotides (in nucleotides) are shown. Data are representative of three repeat experiments performed using different protein preparations.

The results show that domain 5 of Hel308 acts as a molecular brake to limit the processivity of the helicase (figure 5.7 B), as reported by White and coworkers (Richards et al., 2008b). Since removal of domain 5 does not lead to unwinding of the 5'-overhang DNA substrate, domain 5 does not play a role in discriminating unwinding

directionality. Given that Sso PBL2025 Hel308 unwinds branched substrates with a defined directionality, it seems unlikely that domain 5 may play a role in binding branched substrates, as proposed by Hopfner and coworkers (Buttner et al., 2007).

5.6 ROLE OF DOMAINS 3-5 DURING DNA UNWINDING

Domains 3-5 of Hel308 have been predicted to couple ATP binding and hydrolysis to DNA translocation and unwinding (Buttner et al., 2007). The elevated rates of ATP hydrolysis demonstrated by 515STOP and 418STOP, in comparison to WT Hel308 (figure 5.5), are consistent with this proposal. To determine the role of domains 3, 4, and 5 of Hel308 in DNA unwinding, the ability of WT, 515STOP, or 418STOP Hel308 (500 nM) to unwind DNA at 60 ºC was investigated, as outlined in Materials and Methods. Hel308 has been shown to preferentially unwind model replication forks with 3'-5' polarity (figure 5.7 and (Richards et al., 2008b)); therefore, a 3' overhang and a model replication fork DNA substrate (with lagging strand only) were used in the following assays (figure 5.8 A).

In agreement with White and coworkers (Richards et al., 2008b), WT Hel308 unwound the 3' overhang and the model replication fork substrate within 3 min (figure 5.8 B). Neither 515STOP nor 418STOP Hel308 exhibited helicase activity (figure 5.8 C-D), supporting a role for domains 3-5 in transducing the free energy of ATP hydrolysis into conformational changes that power DNA translocation and unwinding (Buttner et al., 2007). The helicase activity of Hel308 was ATP-dependent since no unwinding was observed in the absence of ATP/ $MgCl_2$.
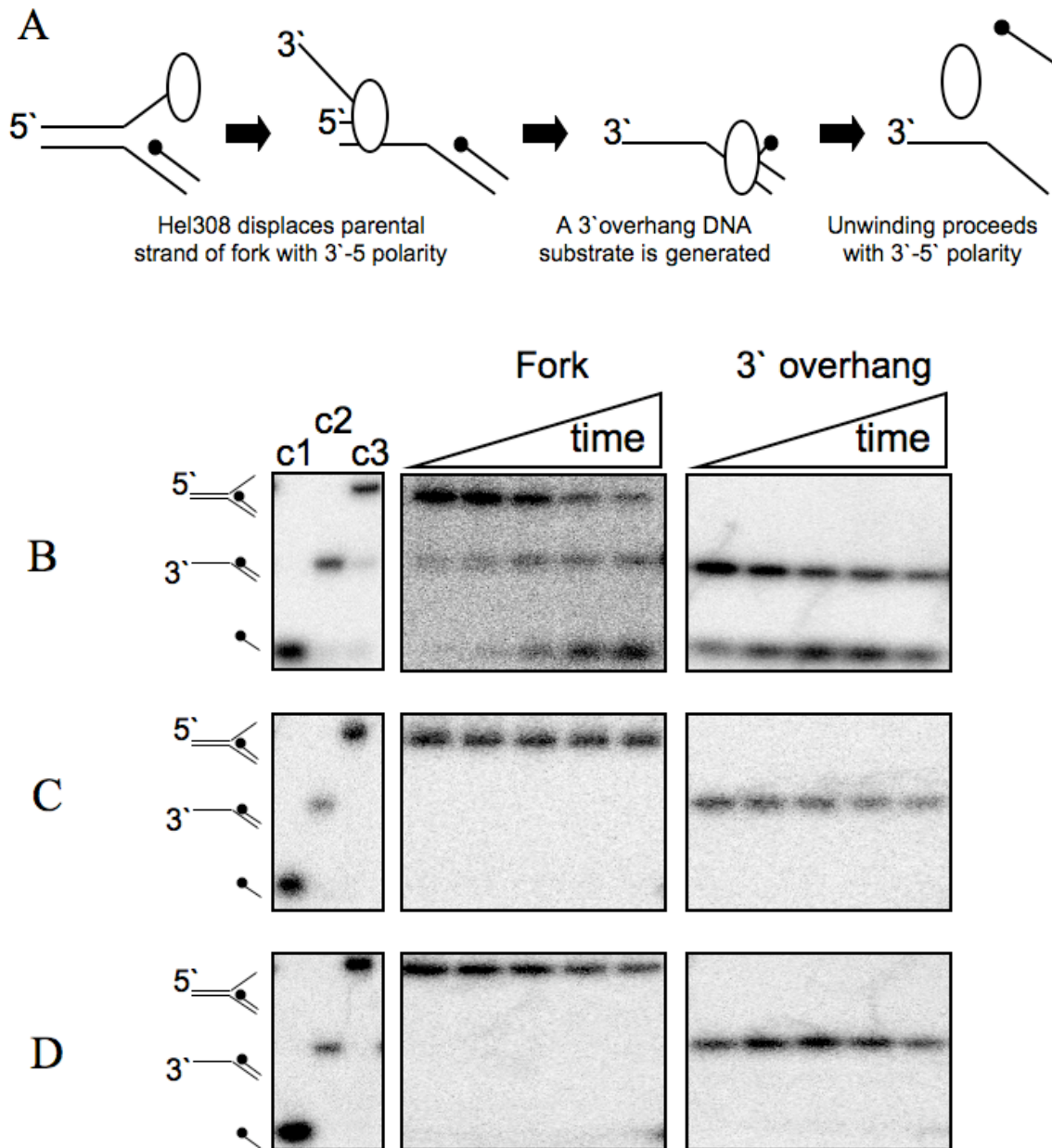
Figure 5.8: DNA unwinding by WT, 515STOP and 418STOP Hel308

(A) Schematic of reaction. Hel308 (white oval) unwinds a fork substrate via a 3'
overhang intermediate. (B-D) Representative native 12% (v/v) polyacrylamide: TBE
gels showing the ability of WT (B), 515STOP (C) and 418STOP Hel308 (D) to unwind
a fork (middle panel) and 3'-overhang DNA substrate (right panel). Samples were taken
at 10, 30, 60, 120, and 180 s. c1, size marker for displaced strand; c2, size marker for
3'overhang substrate; c3, without ATP/ $MgCl_2$ at 180 s. The figure shows representative
gels obtained from experiments performed at least three times.

5.7   POLARITY OF DNA TRANSLOCATION BY HEL308

The polarity of DNA translocation by Hel308 was investigated using the streptavidin displacement assay (Byrd and Raney, 2004), as outlined in Materials and Methods. Assays were performed at 45 ºC in the presence of 500 nM WT Hel308. Two DNA substrates were designed, each possessing a 5' [^{32}P]-radiolabel and a biotin:streptavidin complex attached to either the 3' or 5' end in order to measure 5'-3' or 3'-5' translocation, respectively (figure 5.9).
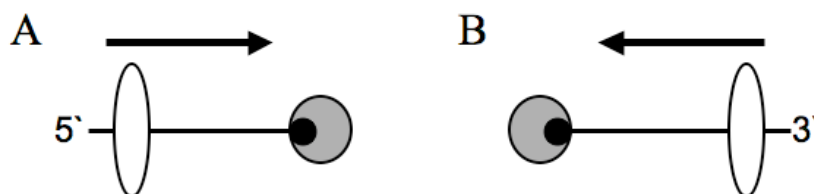
Figure 5.9: Substrate design for streptavidin displacement assays

Oligonucleotides were synthesised with biotin (black circle) conjugated to the 3' (A) or 5' (B) end (Appendix 1, table A1.4: oligonucleotides B50-3'bio and B50-5'bio, respectively). Streptavidin (grey circle) was attached to biotin by incubation at 45 ºC for 5 min. These substrates allow detection of 5'-3' (A) or 3'-5' (B) translocation by Hel308 (white oval), as indicated by the black arrow. Both substrates were 5' [^{32}P]-end labelled.

The results show that WT Hel308 can displace streptavidin from biotin attached to the 3' and 5' end of DNA, suggesting that the protein can translocate along DNA with 5'-3' and 3'-5' polarity (figure 5.10). Arguably, DNA translocation occurs at a faster rate in the 3'-5' direction. These results are consistent with preliminary, unpublished data presented in the thesis of Dr J. Richards (University of St Andrews) and single-molecule data presented in chapter 6.
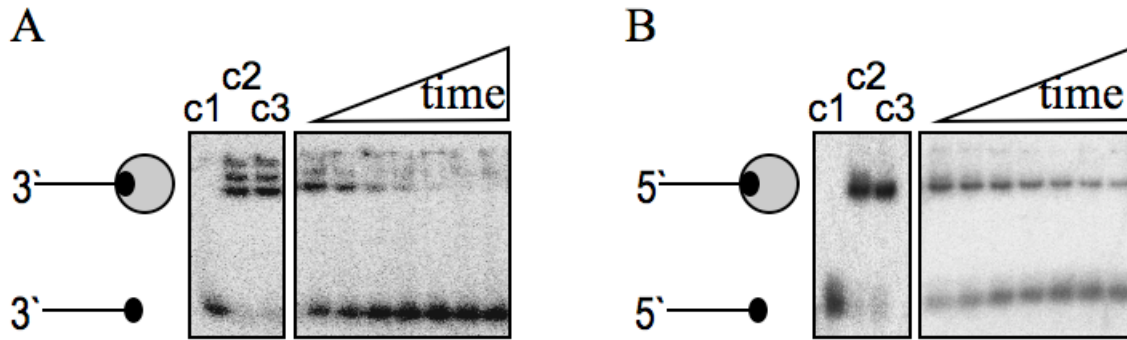
Figure 5.10: Hel308 can translocate along DNA bidirectionally

WT Hel308 (500 nM) was incubated at 45 ºC with DNA (substrates as shown in figure 5.9) and reactions were stopped after 0.5, 1, 2, 3, 5, 10, 15 min. The representative native 12% (v/v) polyacrylamide: TBE gels reveal that Hel308 can translocate along DNA in a 3'-5' (A) and 5'-3' (B) direction. c1, without streptavidin; c2, without protein at 15 min; c3, without ATP/ MgCl$_2$ at 15 min. The figure shows representative gels obtained from experiments performed at least three times.

5.8   ROLE OF DOMAINS 3-5 DURING TRANSLOCATION

Hel308 lacking domains 3-5 hydrolyses ATP but cannot catalyse strand displacement. The ability of these mutants to translocate along DNA was tested using the streptavidin displacement assay (Byrd and Raney, 2004), as outlined in Materials and Methods. WT, 515STOP and 418STOP Hel308 (500 nM) were incubated with a 5' [^{32}P]-end labelled DNA with biotin attached to the 5' end. DNA substrates were incubated at 45 ºC with streptavidin to promote its interaction with biotin (figure 5.9 B).

In contrast to WT Hel308, neither 515STOP nor 418STOP Hel308 were able to disrupt the tight biotin-streptavidin linkage (figure 5.11). This does not necessarily imply that domains 3-5 are necessary for DNA translocation, but rather that they are required for streptavidin displacement. It is possible that 515STOP and 418STOP Hel308 move along DNA but stall upon encountering the bulky streptavidin molecule. Alternatively, reduced DNA translocation capabilities may provide a convenient explanation for the lack of Hel308 helicase activity in the absence of domains 3-5. This further supports the hypothesis that domains 3-5 are essential for coupling ATP hydrolysis (performed by domains 1 and 2) to DNA translocation and unwinding.
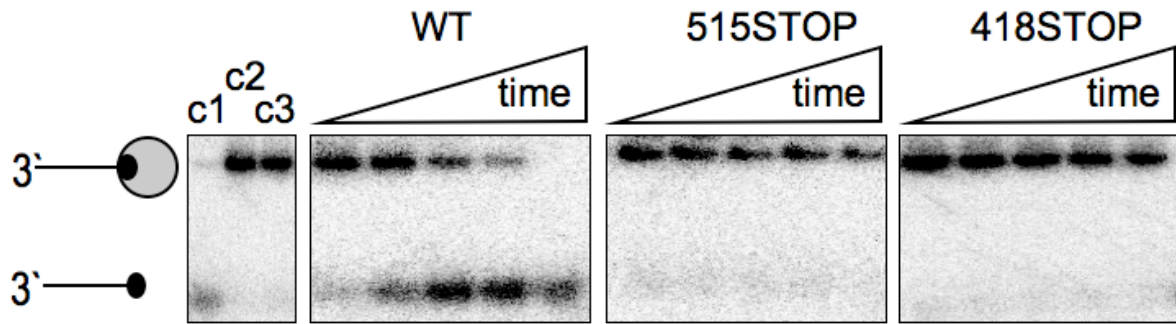
Figure 5.11: DNA translocation by WT, 515STOP and 418STOP Hel308

Representative 12% (v/v) native polyacrylamide: TBE gels comparing the 3'-5' translocation activity of WT, 515STOP, and 418STOP Hel308 (500 nM) at 45 ºC. Time points: 0.5, 1, 2, 5, 10 min. c1, without streptavidin; c2, without protein after 10 min; c3, WT Hel308 without ATP/ MgCl$_2$ after 10 min. Black circle, biotin; grey circle, streptavidin. The figure shows representative gels obtained from experiments performed at least three times.

5.9   PROTEIN DISPLACEMENT BY HEL308

The streptavidin displacement assay relies on the disruption of the biotin:streptavidin linkage; this is the strongest known natural protein-ligand interaction, with a dissociation constant of approximately 10^{-15} mol/ L. The ability of Hel308 to disrupt this interaction suggests that Hel308 may be able to disrupt tight protein interactions *in vivo*, such as those formed between DNA and DNA-binding proteins. In a cellular environment, DNA binding proteins play a vital role in protecting ssDNA from nuclease degradation. This is particularly relevant to the proposed role of Hel308 in the restoration of stalled replication forks: upon stalling, exposed regions of ssDNA are rapidly bound by the single-stranded DNA binding protein (SSB). However, SSB must be removed to allow DNA binding by DNA recombinases, such as RadA (the archaeal Rad51 family recombinase), to initiate homologous recombination (HR) and restart replication. Therefore, the potential for Hel308 to displace DNA binding proteins is of considerable interest. Based on their role in DNA metabolism, SSB, RadA, Sso2452 (a paralogue of RadA; see chapter 7), and Alba1 (the major chromatin protein that binds both ss- and ds-DNA) were chosen as the DNA-binding proteins used in the following assays. The affinity of these proteins for ss- and ds-DNA is noted in table 5.2.

Table 5.2: Dissociation constants of DNA binding proteins. K_D values for SSB (Kernchen and Lipps, 2006), Alba1 (Jelinska et al., 2005), RadA and Sso2452 (McRobbie et al., 2009). These data are not directly comparable since DNA substrates of different lengths and sequences were used.

	K_D (nM)	
PROTEIN	**ssDNA**	**dsDNA**
SSB	10 – 30	-
Sso2452	120	2420
Alba1	270	150
RadA	2640	5330
Hel308	140	5300

5.9.1   PROTEIN DISPLACEMENT DURING DNA UNWINDING

To mimic *in vivo* conditions, a helicase assay was performed using a 3' overhang substrate (figure 5.6 A) pre-incubated with excess DNA-binding proteins (Sso2452, RadA, SSB and Alba1). Reactions containing 10 nM 5' [^{32}P]-end labelled DNA, 10 µM Sso2452/ SSB/ RadA/ Alba1 and 1 mM ATP/ MgCl$_2$ were pre-incubated at 60 ℃ prior to initiation by the addition of 500 nM WT Hel308. Experiments were also performed using 646STOP Hel308 to determine the role of domain 5 in protein displacement.

DNA unwinding by Hel308 was not inhibited by Alba1, whereas reduced activity was observed in the presence of RadA and Sso2452 (figure 5.12 and table 5.3). Owing to the inherent ability of SSB to melt duplex DNA, the ability of Hel308 to displace SSB during DNA unwinding could not be investigated (figure 5.12). The dependence of the reaction on ATP suggests that Hel308 actively removed Alba1, RadA and Sso2452 from DNA, as opposed to passively displacing them by competitive DNA binding. The absence of domain 5 did not significantly affect the ability of Hel308 to displace DNA binding proteins during DNA unwinding. These data are consistent with the proposed role for Hel308 in remodelling stalled replication forks by a combination of DNA and protein displacement.
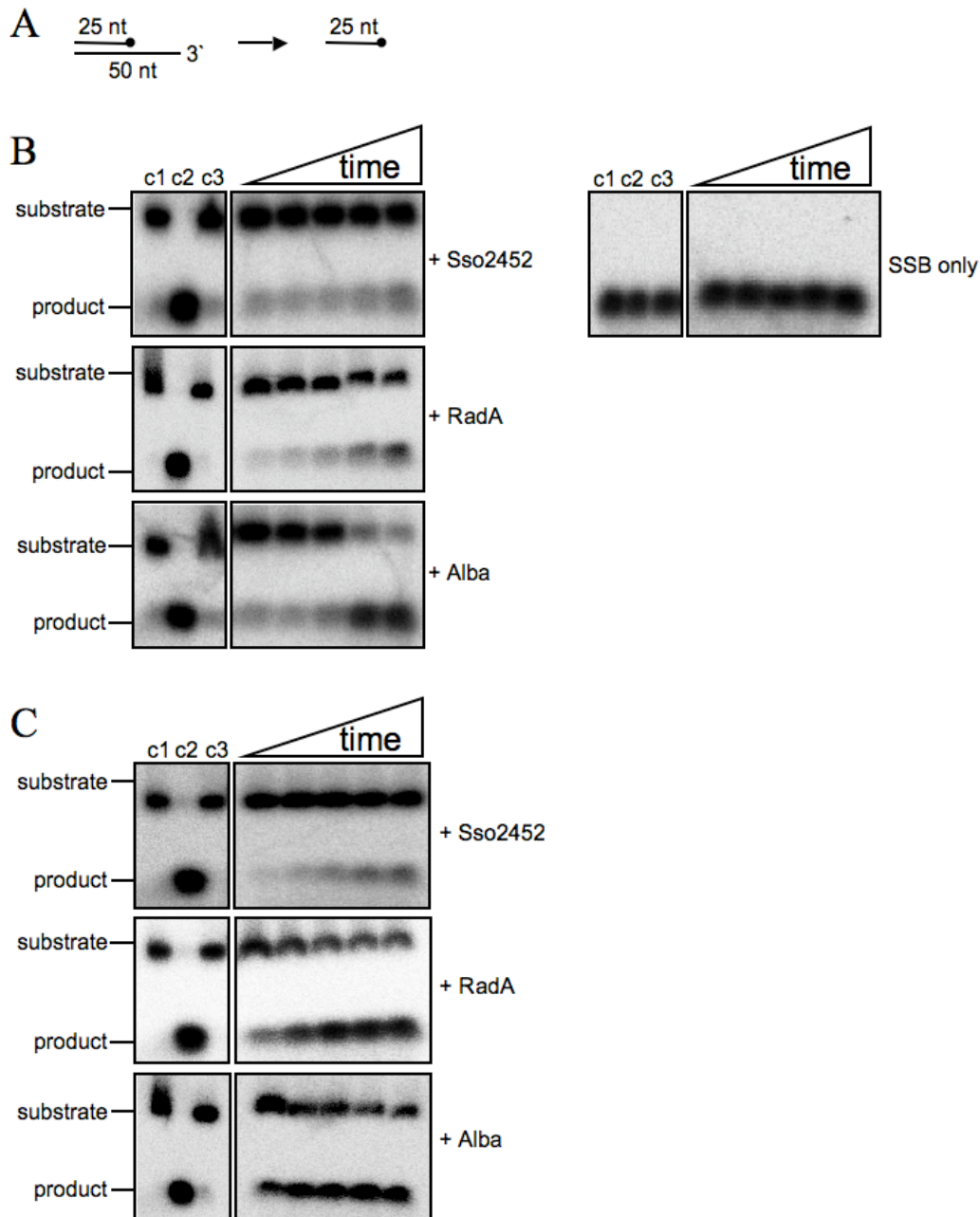
Figure 5.12: DNA unwinding by Hel308 in the presence of DNA-binding proteins

(A) Schematic of reaction. (B-C) 3'-5' helicase activity of WT (B) and 646STOP (C) Hel308 in the presence of DNA-binding proteins. DNA (10 nM) was pre-incubated with ATP/ MgCl$_2$ and 10 μM Sso2452/ RadA/ Alba1 prior to the addition of 500 nM Hel308. Time course: 10, 30, 60, 120 and 180 s. The DNA-binding protein included in each assay is indicated to the right of each gel panel. 'SSB only', reaction containing 10 μM SSB and no Hel308. c1, reaction without Hel308 after 180 s; c2, size marker; c3, reaction without ATP/ MgCl$_2$ after 180 s. The figure shows representative gels obtained from experiments performed at least three times.

Table 5.3: DNA unwinding by Hel308 in the presence and absence of DNA-binding proteins. Strand displacement (%) of a 3'-overhang substrate (10 nM) after 180 s by WT and 646STOP Hel308 (500 nM), in the presence of DNA-binding protein (10 μM). Results reflect the mean of triplicate measurements with standard errors shown.

Protein	% strand displacement after 180 s	
	WT	646STOP
Hel308 only	78 ± 4	97 ± 4
+ Sso2452	19 ± 2	17 ± 2
+ RadA	46 ± 3	58 ± 4
+ Alba1	81 ± 4	72 ± 7

5.9.2 PROTEIN DISPLACEMENT DURING DNA TRANSLOCATION

The bidirectional translocation activity of Hel308 in the presence of DNA-binding proteins was investigated using the streptavidin-displacement assay, as outlined in Materials and Methods. DNA (substrates shown in figure 5.9) was pre-incubated with 10 μM SSB/ RadA/ Alba1 at 45 ºC for 5 min prior to reaction initiation by the addition of Hel308 (500 nM).

Table 5.4: DNA translocation by Hel308 in the presence and absence of DNA binding proteins. Image Gauge software was used to quantify the percentage of streptavidin displaced by Hel308 (500 nM) in the presence and absence of DNA-binding proteins (10 μM) after 15 min at 45 ºC. Data represents the mean of triplicate measurements with standard errors shown.

Protein	% streptavidin displacement after 15 min	
	3'-5' translocation	5'-3' translocation
Hel308 only	91 ± 3	79 ± 3
+SSB	53 ± 2	5 ± 1
+RadA	89 ± 4	81 ± 3
+Alba1	72 ± 5	49 ± 2

Hel308 can translocate along DNA in the 3'-5' direction in the presence of all DNA-binding proteins tested; only SSB slightly inhibited the rate of translocation (figure 5.13 and table 5.4). In contrast, the presence of SSB and Alba1 perturbed Hel308 translocation in the 5'-3' direction. This may suggest that, despite being capable

of bidirectional translocation, Hel308 preferentially moves along DNA with 3'-5' polarity.
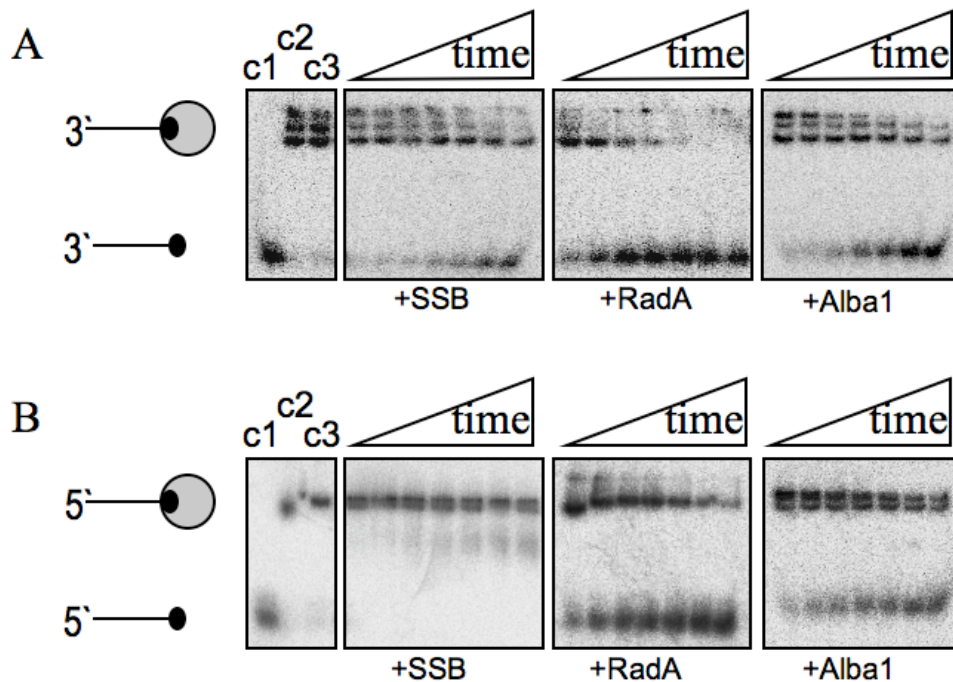
Figure 5.13: DNA translocation by Hel308 in the presence of DNA-binding proteins

Translocation of Hel308 (500 nM) along DNA in a 3'-5' (A) and 5'-3' (B) direction, in the presence of 10 µM SSB, RadA, or Alba1 was assessed using the streptavidin displacement assay. Reactions were stopped after 0.5, 1, 2, 3, 5, 10 and 15 min and analysed on native 12% (v/v) polyacrylamide: TBE gels. c1, without streptavidin; c2, without protein at 15 min; c3, without ATP/ MgCl$_2$ at 15min. The figure shows representative gels obtained from experiments performed at least three times.

5.10   DISPLACEMENT OF SSB BY HEL308 ENABLES RADA-CATALYSED HOMOLOGOUS RECOMBINATION

Upon replication fork stalling, a helicase is required to produce ssDNA as a substrate for the strand exchange reaction, catalysed by RecA (bacteria)/ Rad51 (eukaryotes)/ RadA (archaea), of homologous recombination. In bacteria, the 3'-5' helicase activities of PriA and Rep carry out this function (Heller and Marians, 2006b). Immediately following strand displacement, however, SSB binds to the ssDNA to protect it from degradation by nucleases. Therefore, for recombination to proceed, SSB must be displaced to allow RecA/ Rad51/ RadA to bind. In bacterial species, the RecFOR complex appears to play a role in loading RecA onto DNA prior to strand

exchange (Morimatsu and Kowalczykowski, 2003), while, in eukaryotes, Rad52 has been implicated in facilitating the displacement of RPA (eukaryotic homologue of SSB) by Rad51 (Sugiyama and Kowalczykowski, 2002). Given the ability of Hel308 to unwind replication fork structures (Richards et al., 2008b) and displace DNA-binding proteins from DNA, this helicase may represent an ideal candidate to displace SSB prior to RadA-catalysed strand exchange in the archaea.

D-loop assays were performed as outlined in Materials and Methods, and involved the formation of a nucleoprotein filament by RadA prior to its invasion into a dsDNA species to generate a recombination intermediate, known as a D-loop (figure 5.14 A).
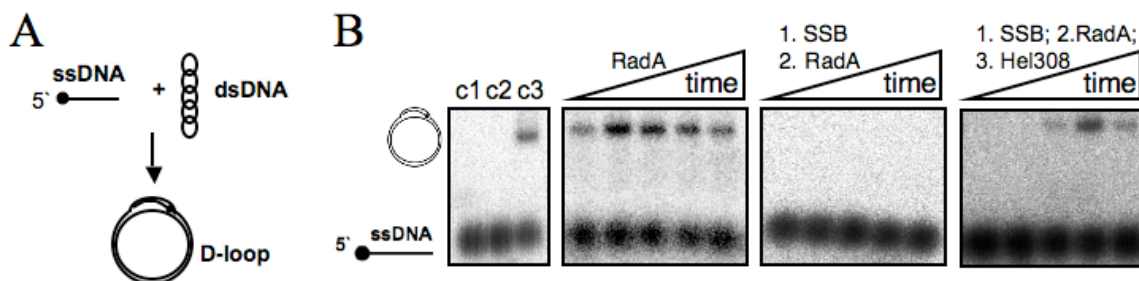
Figure 5.14: Hel308 relieves SSB-mediated inhibition of RadA D-loop formation

(A) 5' [^{32}P]-end labelled 80mer (Appendix 1, table A1.3) was incubated with a dsDNA plasmid (pUC19) in the presence of RecA/ RadA to form a D-loop. (B) Representative 0.8% (w/v) agarose gels showing D-loop formation by RadA at 60 ºC. For order of addition experiments, protein 1 was pre-incubated with ssDNA for 5 min and, where indicated, protein 2 was incubated with ssDNA-protein 1 complexes for a further 5 min prior to the addition of protein 3. Reactions were initiated by dsDNA. Time course: 1, 5, 10, 20, 30 min. c1, reaction without RadA at 30 min; c2, reaction without ATP/ MgCl$_2$ after 30 min; c3, reaction catalysed by RecA (5 µM) after 1 min at 37 ºC. Protein concentrations: RadA, 5 µM; SSB, 1 µM; Hel308, 1 µM. The figure shows representative gels obtained from experiments performed at least three times.

RadA catalysed the formation of D-loops; however, in the presence of SSB-bound ssDNA, the recombinase activity of RadA was abolished. The addition of Hel308 relieved this inhibition (figure 5.14), which implies that Hel308 may be involved in the displacement of SSB from DNA to facilitate RadA-catalysed strand exchange during homologous recombination.

5.11   HEL308 PROTEIN-PROTEIN INTERACTIONS

Displacement of SSB by Hel308 to promote RadA-catalysed recombination may rely on the physical interaction of all three proteins. To investigate the ability of Hel308 to bind SSB and RadA, polyhistidine-tagged Hel308 was incubated with either untagged RadA or untagged SSB (both *S. solfataricus*) for 1 h at room temperature, in the presence of ethidium bromide, as outlined in Materials and Methods (figure 2.10). Since Hel308, SSB and RadA bind DNA, the inclusion of ethidium bromide ensures that indirect protein interactions via DNA bridging are not permitted. The protein mixture was loaded onto a 1 ml nickel-chelating affinity column and the flow-through was collected. After substantial washing at low imidazole concentrations, bound proteins were eluted from the column with 500 mM imidazole. Both RadA and SSB were found in the column flow-through, indicating that neither protein interacts directly with Hel308 (figure 5.15).
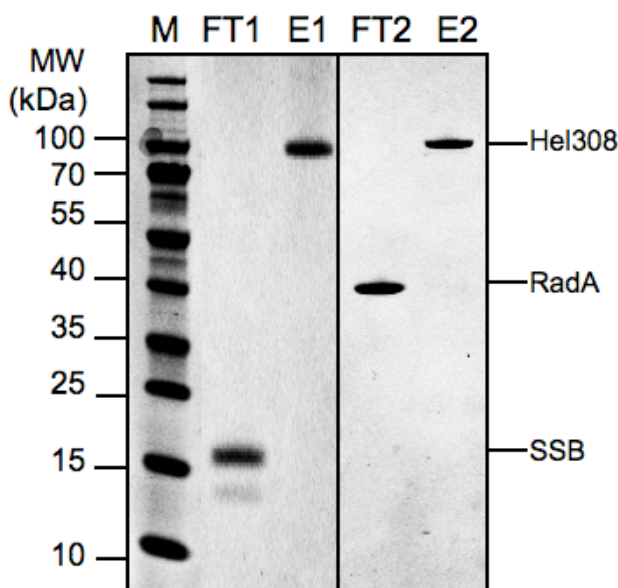
Figure 5.15: Hel308 does not interact with SSB or RadA

Polyhistidine-tagged Hel308 was incubated with untagged SSB for 1 h at room temperature. The protein mixture was loaded onto a 1 mL nickel-chelating column and the flow-through was collected (FT1). Bound materials were eluted (E1). Tagged Hel308 was then incubated with untagged RadA and the experiment was repeated. Unbound (FT2) and bound (E2) materials were collected. FT1, FT2, E1 and E2 were analysed by SDS-PAGE. The experiment was repeated three times with fresh protein.

The ability of Hel308 to unwind model replication forks and displace proteins, including SSB, from DNA substrates is consistent with the proposed role for Hel308 in replication restart following fork stalling (Guy and Bolt, 2005; Richards et al., 2008b). To establish whether Hel308 had any further role during this repair pathway, the protein-protein interaction study above was repeated using polyhistidine-tagged Hel308 and untagged Hjc, the Holliday junction-resolving enzyme from *S. solfataricus*. Western blottting with an anti-Hjc antibody revealed that Hjc co-eluted

with Hel308 from the nickel-chelating column (figure 5.16). Ethidium bromide was included in the experiment, as outlined in Materials and Methods, excluding the possibility that the interaction occurred indirectly via DNA. The interacting surface of each protein was crudely mapped by repeating the experiment with various truncations of the proteins (figure 5.16). In the absence of domains 4 and 5, Hel308 (515STOP) continued to interact with Hjc. However, deletion of domain 3 (418STOP) abrogated this interaction suggesting that residues of this domain may be responsible for engaging with the Hjc protein. Analysis of the Hjc sequence revealed a C-terminal PCNA-interacting motif (PIP motif). The importance of this site in binding Hel308 was investigated by removing the seven most C-terminal residues of Hjc ('HjcΔ7'). The co-elution of Hel308 and HjcΔ7 indicated that the PIP motif was not necessary for Hel308 binding (figure 5.16). For the Hel308/ Hjc experiments, two control experiments were performed: 1, Purified Hel308 was incubated with the anti-Hjc antibody to verify antibody specificity; 2, Purified, untagged Hjc was loaded onto the nickel-chelating column in the absence of tagged Hel308. Eluted material was incubated with the anti-Hjc antibody to verify that Hjc did not interact with the nickel-chelating column.

Subsequent to this investigation, Shen and coworkers reported that Hel308 from *P. furiosus* (Pfu Hjm) physically interacted with Pfu Hjc (Li et al., 2008). The physical association of these two proteins inhibited the helicase activity of Pfu Hjm. Thus, the authors concluded that Hjm and Hjc from *P. furiosus* physically and functionally interact.
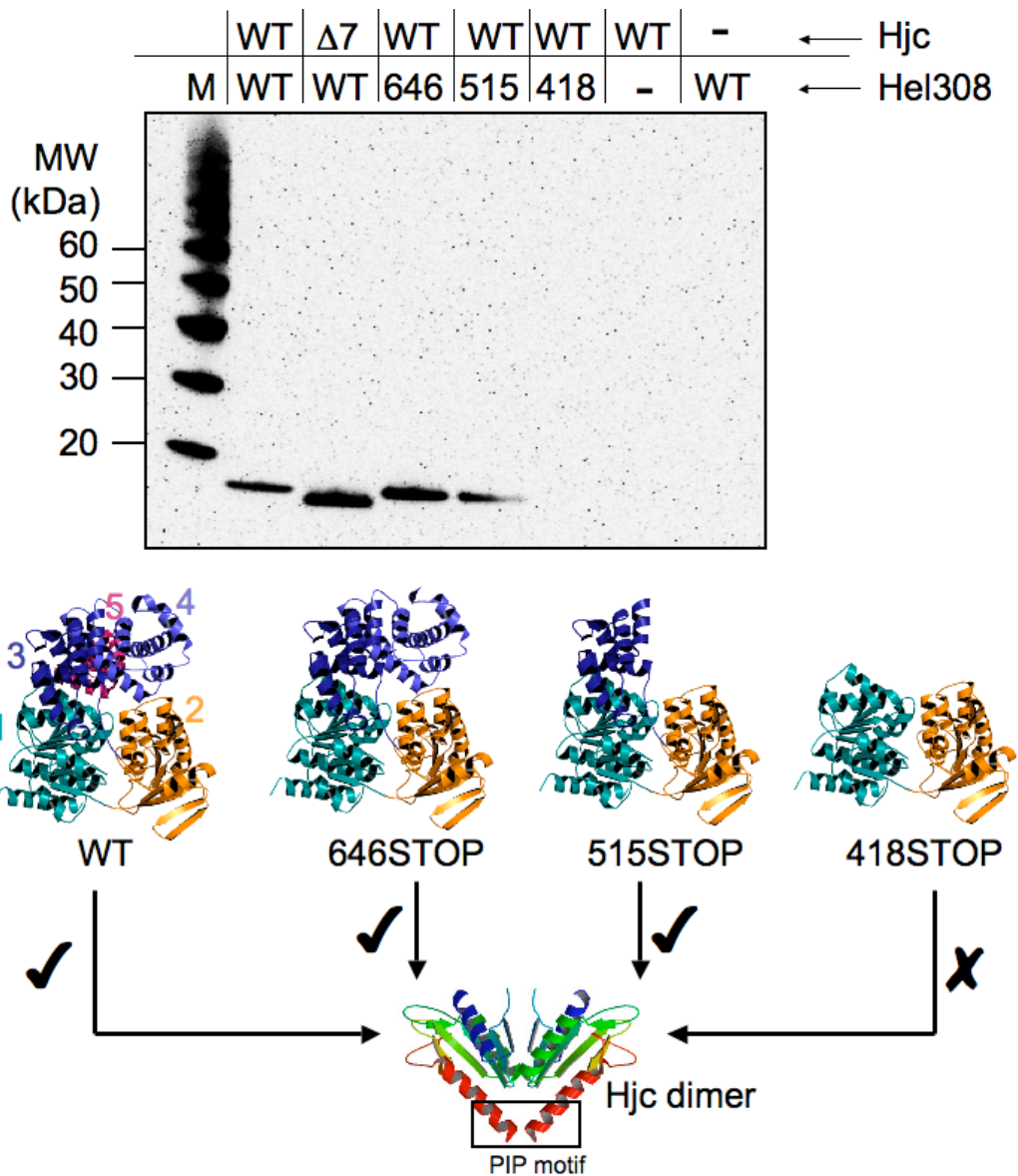
Figure 5.16: Hel308 interacts with Hjc from *S. solfataricus*

WT, 646STOP, 515STOP or 418STOP Hel308 was incubated with either WT Hjc or HjcΔ7 ('Δ7'; Hjc lacking the PIP motif) for 1 h at room temperature. Samples were loaded onto a 1 ml nickel-chelating column. Flow-through was collected and bound material was eluted. Eluted fractions were analysed by Western blotting using an anti-Hjc antibody. The table above the gel indicates the interacting pair; gel bands correspond to the presence of Hjc in the eluted fraction, indicative of its interaction with Hel308. Hjc did not co-elute with 418STOP Hel308, which suggests that domain 3 of Hel308 may be involved in the pairwise interaction. Controls: Hel308 alone, to verify the specificity of the Hjc antibody; Hjc alone, to verify that Hjc does not interact with the nickel column. M, marker. The figure shows a representative gel obtained from experiments performed at least three times.

5.12   DISCUSSION AND CONCLUDING REMARKS

This chapter reported that Hel308 from *S. solfataricus* PBL2025 catalyses ATP-dependent, bidirectional DNA translocation and unidirectional 3'-5' DNA unwinding. The observation that Hel308 unwinds DNA with 3'-5' polarity is consistent with previous claims that Hel308 unwinds the lagging strand of stalled replication forks (Guy and Bolt, 2005; Richards et al., 2008b) and contradicts proposals that the bidirectional unwinding activity of the Hel308 homologue, Hjm, from *S. tokodaii* reflects the evolutionary divergence of Crenarchaeota and Euryarcheota (Li et al., 2008).

The ability of Hel308, a SF2 DNA helicase, to translocate along DNA without defined polarity is a very unusual observation. In particular, it is difficult to consolidate its bidirectional translocation and unidirectional unwinding activity: while several DExD/ H RNA helicases, including eukaryotic prp43 (Tanaka and Schwer, 2006) and Ded1 (Yang and Jankowsky, 2006), reportedly translocate bidirectionally along RNA, this is generally accompanied by bidirectional helicase activity. The data presented in this chapter suggests that binding of Hel308 to the lagging or leading strand template permits 3'-5' translocation and concomitant unwinding of the lagging strand or parental duplex, respectively. Although Hel308 can translocate along the lagging or leading strand template in a 5'-3' direction, this is not accompanied by DNA displacement (figure 5.17). These observations may reflect an ability of Hel308 to bind to DNA in two distinct orientations, effectively uncoupling translocation from DNA unwinding in one orientation due to the incorrect placement of the β-hairpin loop of motor domain 2 at the ss-/ ds-DNA junction. The translocation and unwinding data presented in this chapter are consistent with preliminary results reported in the thesis of Dr J. Richards (University of St Andrews) and with initial translocation studies of Hel308 using single molecule fluorescence resonance energy transfer (chapter 6). Nevertheless, further studies are required to understand the molecular basis of DNA translocation and unwinding by Hel308.
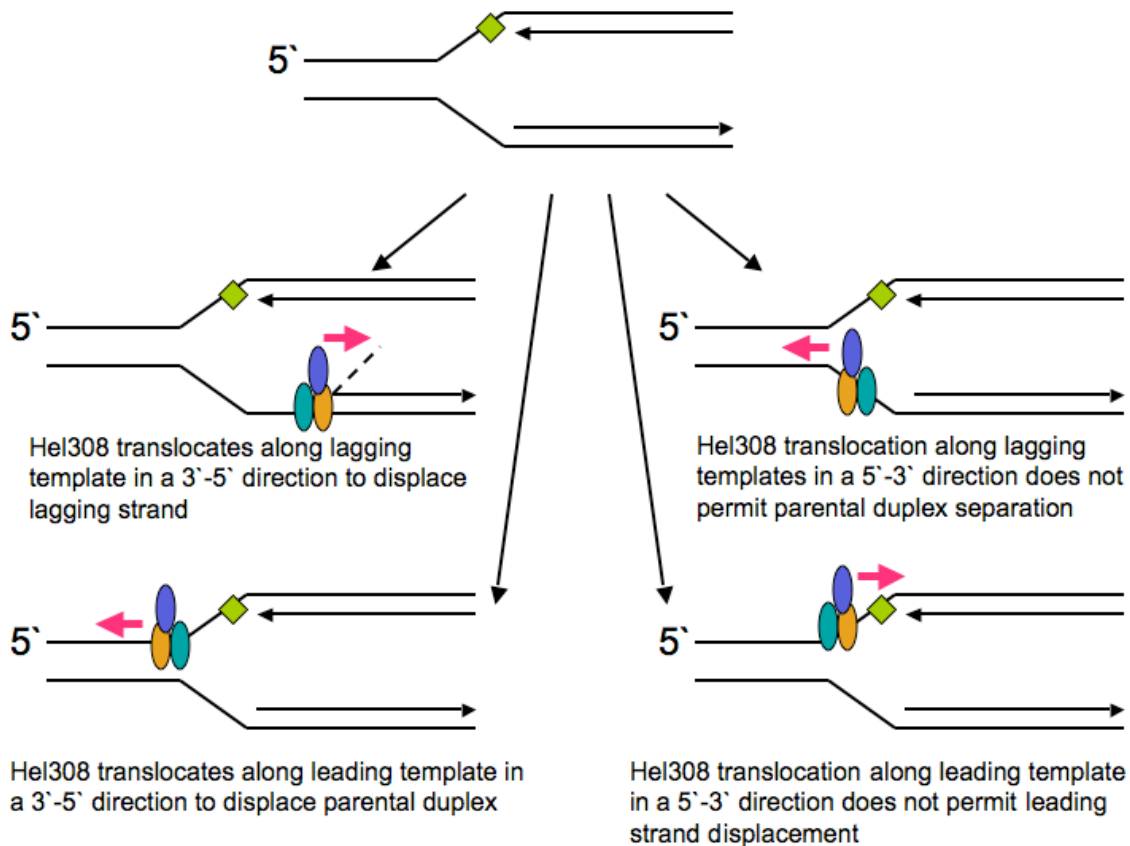
Figure 5.17: DNA translocation and unwinding by Hel308

Cartoon summarising the bidirectional translocation and unidirectional DNA unwinding activity of Hel308. Motor domain 1, teal; motor domain 2, orange; domain 4, blue; green diamond, lesion on leading strand. The β-hairpin loop, implicated in strand-displacement, is located on domain 2.

DNA translocation and unwinding by Hel308 lacking domain 5 (646STOP) was more efficient compared with the WT protein, consistent with the proposal that domain 5 is an autoinhibitory domain (Richards et al., 2008b). The substrate preference of Hel308 was not affected by the removal of domain 5, suggesting that it does not play a role in determining the polarity of DNA translocation and unwinding. This contradicts the proposal that domain 5 recognises branched nucleic acids, including replication forks, positioning the protein to unwind the lagging strand (Buttner et al., 2007).

Hopfner and coworkers postulated that domains 3-5 were essential for coupling ATP hydrolysis to DNA translocation and unwinding (Buttner et al., 2007). Site-directed mutagenesis was used to clone Hel308 lacking domains 3-5 (418STOP Hel308) and domains 4-5 (515STOP Hel308) to investigate this hypothesis. Both these mutants hydrolysed ATP at a rate slightly greater than that of WT Hel308, but failed to

translocate along DNA or catalyse strand displacement. This was consistent with an uncoupling between ATP hydrolysis and DNA unwinding/ translocation.

The data presented in this chapter reveals that Hel308 can disrupt protein-nucleic acid complexes, which may be relevant *in vivo*. DNA translocation and unwinding by Hel308 in the presence of 20-fold molar excess of RadA, SSB or Alba1 revealed significant residual activity (Richards et al., 2008b). Given that Hel308 has been shown to preferentially unwind stalled replication forks (Guy and Bolt, 2005; Richards et al., 2008b), this data is consistent with a role for Hel308 in the displacement of DNA-binding proteins from the site of a stalled replication fork. This would permit binding of DNA recombinases, such as RadA, to initiate homologous recombination, or the binding of other proteins necessary for replication restart. In support of this proposal, our data showed that Hel308 could overcome the inhibitory effect of SSB on RadA-mediated recombination. The bacterial RecFOR complex has been implicated in an analogous role, responsible for the displacement of SSB from DNA to promote RecA-catalysed recombination (Umezu et al., 1993; Morimatsu and Kowalczykowski, 2003). Since Hel308 can also displace RadA from DNA, the data may alternatively imply that Hel308 functions to limit recombinational activity, similar to that described for bacterial UvrD that has been shown to disrupt RecA nucleoprotein filaments (Veaute et al., 2005). We have presented data supporting a physical interaction between Hel308 and the Holliday junction-resolving enzyme, Hjc, which functions during the latter stages of homologous recombination to cleave the branch-points of Holliday junctions. Shen and coworkers similarly reported the direct association between the Hel308 (StoHjm) and Hjc (StoHjc) homologues from *S. tokodaii*, which resulted in the inhibition of Hjm helicase activity (Li et al., 2008). StoHjm also exhibited structure-specific single-strand DNA annealing and fork regression activities, leading the authors to support proposals that Hjm/ Hel308 proteins are involved in processing stalled replication forks, an activity that is regulated by Hjc (Li et al., 2008; Oyama et al., 2009). Negative regulation of the helicase by Hjc may act to limit recombination events within the cell. These results suggest that the Hel308 DNA helicase occupies a similar role in the cell to RecQ helicases, such as BLM and WRN, for which there are no known archaeal homologues.

Based on the data presented in this chapter, a model has been proposed for Hel308 cellular function (figure 5.18). Upon replication fork stalling, the SF2 DNA helicase displaces DNA and DNA-binding proteins to permit RadA-catalysed strand

exchange. Upon formation of the RadA nucleoprotein filament, homologous recombination (via a Holliday junction intermediate) or NER (via parental duplex reannealing) occurs to restore the replication fork and restart replication.
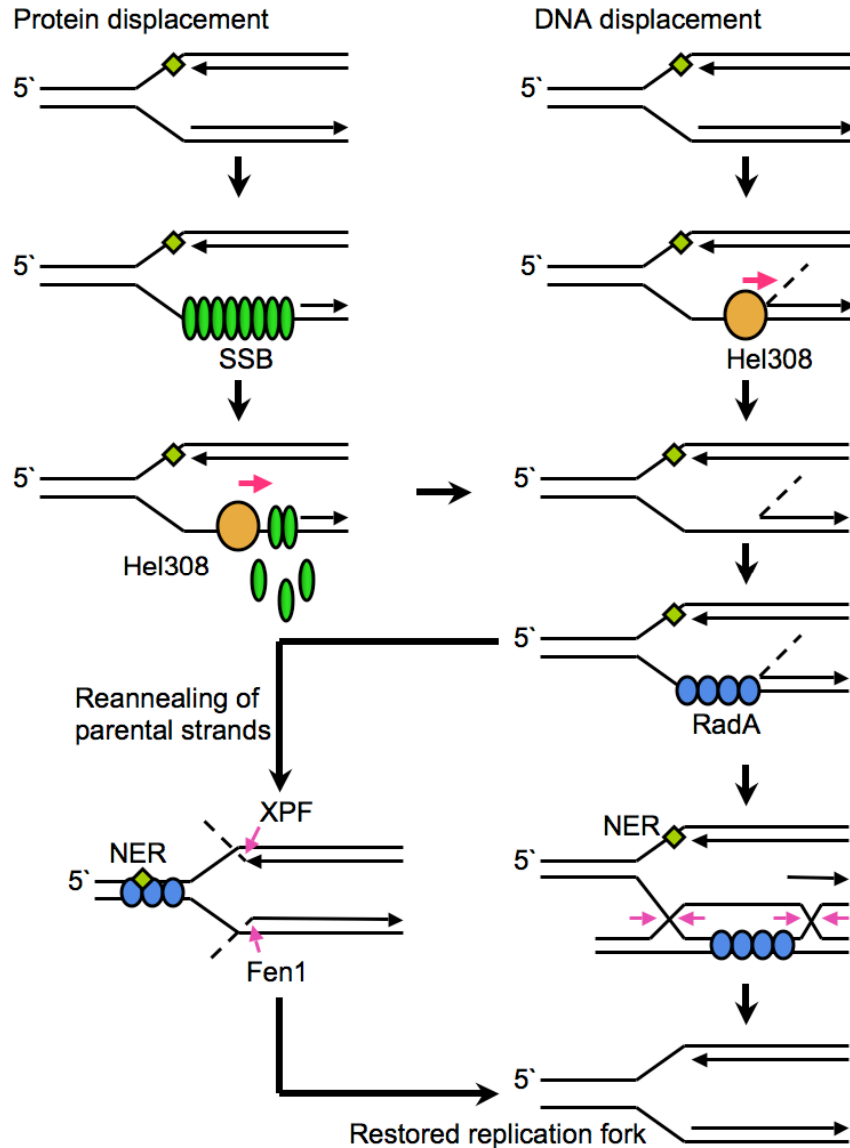
Figure 5.18: Model for Hel308 remodelling of stalled replication forks

Upon encountering a lesion on the leading strand (green diamond), the replication fork stalls. Hel308 (orange circle) displaces the lagging strand, which is immediately bound by SSB (green oval). Prior to RadA-mediated recombination, Hel308 displaces SSB to enable RadA (blue circle) binding. Resumption of replication occurs following Holliday junction resolution by Hjc (cleavage indicated by pink arrows). Alternatively, RadA could promote reannealing of the parental duplex for NER to restart replication following end processing by XPF and Fen1 (cleavage indicated by pink arrows).

CHAPTER 6

A BIOPHYSICAL STUDY OF HEL308

6.1   INTRODUCTION

DNA helicases are molecular motors that couple the free energy of ATP binding and hydrolysis to translocation and unwinding of duplex DNA (Tuteja and Tuteja, 2004). Despite the existence of several models that aim to explain DNA unwinding, a detailed understanding of how SF2 helicases unwind DNA duplexes has yet to be established. Plausible mechanisms describe the requirement for modulation of nucleic acid affinity, mediated by ATP binding and hydrolysis, with concomitant changes in protein conformation to explain directional movement coupled to DNA unwinding (Pyle, 2008).

The 'active rolling' model and the 'inchworm' model comprise the two main hypotheses offered to explain the mechanism of DNA unwinding and are described fully in chapter 1. Briefly, these two mechanisms have distinct differences: firstly, the active rolling model demands (at least) a dimeric protein, whereas the inchworm model is compatible with proteins in any oligomeric state; secondly, the active rolling model dictates that each subunit of the dimeric protein must contact only ssDNA or dsDNA at any one time, while the inchworm model relies on the protein binding both ss- and ds-DNA simultaneously; finally, the inchworm mechanism is compatible with any step size, including 1, but the active rolling model necessitates a step size (at least) equal to the footprint of each individual binding site (Soultanas and Wigley, 2000).

Work on the crystal structure of the SF1 DNA helicase, PcrA, from *Bacillus stearothermophilus*, has put forward evidence supporting the adoption of the inchworm model as a general mechanism to explain DNA unwinding and translocation by DNA helicases (Velankar et al., 1999). PcrA was crystallised alone, in the presence of DNA (duplex DNA with a 3' single-stranded tail), and in the presence of DNA with AMP-PNP, a nonhydrolysable analog of ATP (represents the 'substrate' configuration) or a sulphate ion (which binds in a position equivalent to the phosphate ion produced as a result of ATP hydrolysis, and therefore represents the 'product' configuration). The structures revealed two domains (1 and 2), each comprising two subdomains (1A and 1B; 2A and 2B). Subdomains 1A and 2A house the seven conserved helicase motifs and consist of a large central β-sheet flanked by α-helices, which has significant similarity to the motor domains of RecA (Velankar et al., 1999). Therefore, subdomains 1A and 2A are referred to as the 'RecA-like motor domains' and are found in all SF1 and SF2 helicases.

The 'substrate' and 'product' configuration of PcrA showed that nucleotide binding takes place within a cleft formed between the motor domains (Velankar et al., 1999). Furthermore, it was evident that dsDNA bound along the side of domain 2B, while ssDNA binding occurred within a cleft formed across the top of the motor domains (Velankar et al., 1999). The interaction with ssDNA was stabilised by stacking interactions present between DNA bases and aromatic residues of the PcrA motor domains (figure 6.1 B). A comparison between the 'substrate' and 'product' configurations of PcrA revealed that significant conformational changes occurred during the enzyme reaction cycle (figure 6.1 A). Firstly, the cleft between subdomains 1A and 2A in which ssDNA binds was open in the 'substrate' configuration, but closed in the 'product' form. Secondly, large and rigid movements of subdomains 1B and 2B were observed between the 'substrate' and 'product' forms of PcrA (figure 6.1 A). Furthermore, the affinity of PcrA for dsDNA altered between the 'substrate' and 'product' configuration: following ATP hydrolysis, dsDNA was bound loosely to subdomain 2B, but upon binding of AMP-PNP, a stronger interaction with dsDNA was observed. It has been concluded that the movements of subdomains 1B and 2B upon ATP binding act to change the surface of the protein such that its shape and charge becomes compatible with dsDNA binding (Velankar et al., 1999).
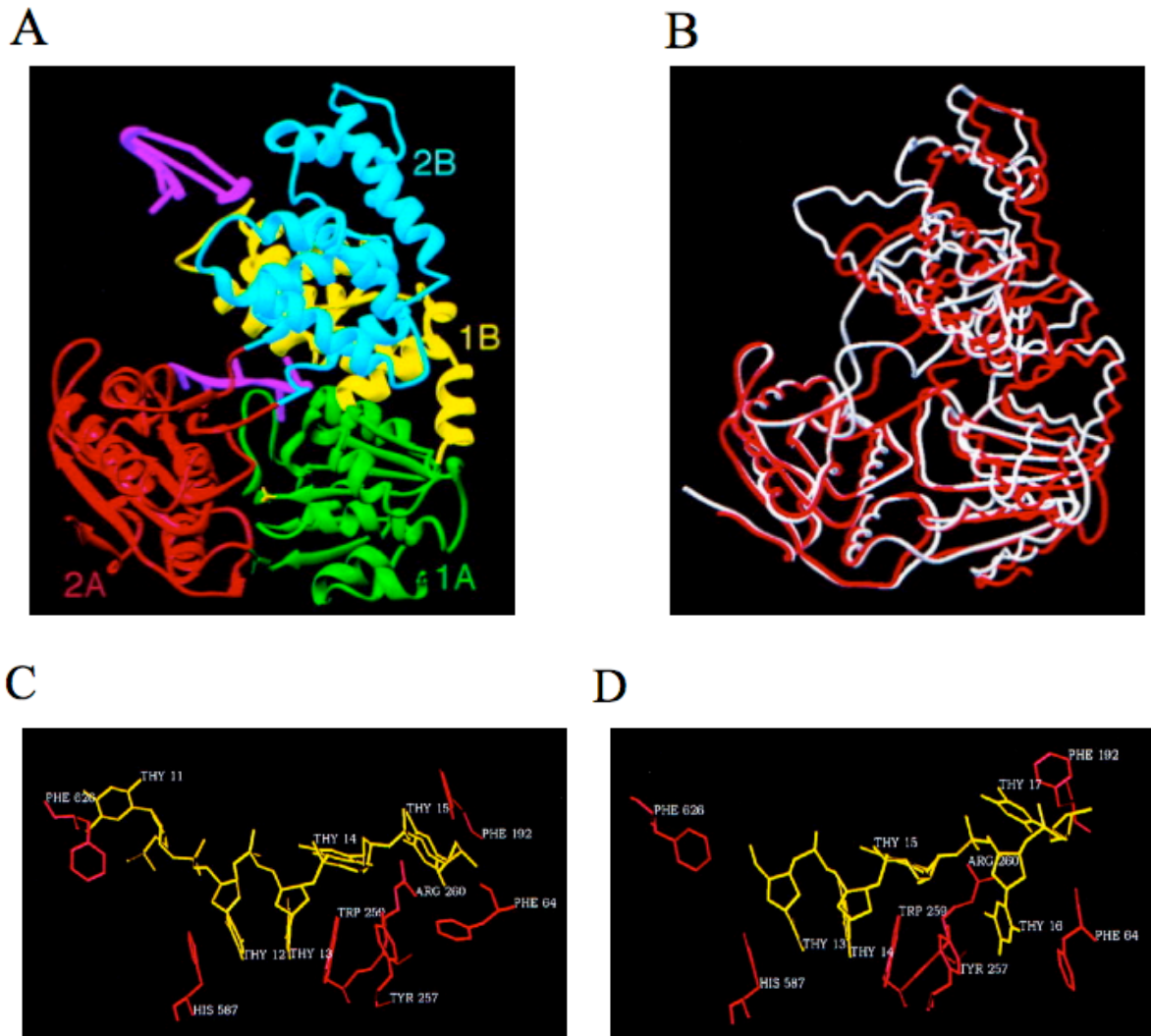
Figure 6.1: PcrA orientation in the 'substrate' and 'product' configuration

(A) Domain structure of PcrA, coloured according to subdomain. (B) Superposition of the 'substrate' (white) and 'product' (red) configurations of PcrA, highlighting the movements of domains 1B and 2B. The binding site for ssDNA in the substrate (C) and product (D) configuration of PcrA, showing the alteration in stacking interactions between aromatic residues of PcrA and DNA bases of the ssDNA chain. Adapted from (Velankar et al., 1999).

The structures also revealed that, upon ATP binding and hydrolysis, the motor domains alternate their relative affinities for ssDNA resulting in the ssDNA tail moving back and forth across the domains during the ATP cycle. By comparing the stacking interactions present in the 'substrate' and 'product' configurations, a plausible mechanism for nucleic acid translocation was determined: in the substrate configuration of PcrA, four DNA bases were shown to stack with amino acids between residues H587 and F192, while in the substrate configuration this increased to five DNA bases (figure 6.1). This change was attributed to the change in position of two phenylalanine residues,

F626 and F64, which alter the interaction between PcrA and the ss-/ ds-DNA junction or the 3' tail of ssDNA, respectively, in the substrate and product configurations. Therefore, it appears that ATP binding and hydrolysis may be coupled to enzyme activity by way of its ability to modulate the enzymes relative affinity for ss- and ds-DNA (Velankar et al., 1999). Accordingly, it was hypothesised that the helicase may wriggle along the ssDNA strand in a unidirectional manner (Soultanas and Wigley, 2000). The data derived from the crystal structures of PcrA suggested that the protein was monomeric and had a step size of 1 during DNA translocation and unwinding (i.e. PcrA unwinds 1 bp of DNA per ATP hydrolysed), consistent with an 'inchworm' mechanism.

The crystal structure of the SF1 DNA helicase, Rep, from *E. coli* provided further insight into the mechanism by which helicases unwind DNA (Korolev et al., 1997). The helicase was crystallised in complex with DNA in the presence or absence of ADP. The Rep structure resembles *B. stearothermophilus* PcrA, with each monomer comprising 2 domains, each further divided into two subdomains (1A, 1B, 2A and 2B). Like PcrA, the Rep helicase was observed in an 'open' and 'closed' conformation, referring to the accessibility of the ssDNA binding groove present across the top of the motor domains: in the closed conformation, ssDNA cannot bind, whereas in the open conformation, ssDNA can bind (figure 6.2 A-B). In the 'open' form of Rep, the relative positions and orientations of the four subdomains are virtually identically to that observed in PcrA (Velankar et al., 1999). Although the overall fold of each subdomain remained largely the same in the 'closed' configuration of Rep, a dramatic 130 ° rotation of subdomain 2B occurred (figure 6.2 B) (Korolev et al., 1997). This was enabled through the movement of a hinge region between subdomains 2A and 2B. This 'domain swivelling' results in the closure of the ssDNA binding cleft, preventing the 3' tail from docking, i.e. the 'closed' conformation. Proteolysis of Rep has revealed that the hinge region becomes accessible to digestion by trypsin upon DNA binding, suggesting that this large domain movement results following DNA binding and may have a functional role to play in Rep activity (Korolev et al., 1997). However, the ssDNA end becomes completely buried within the protein following the movement of subdomain 2B, rendering it unlikely that this conformation yields a functional helicase.
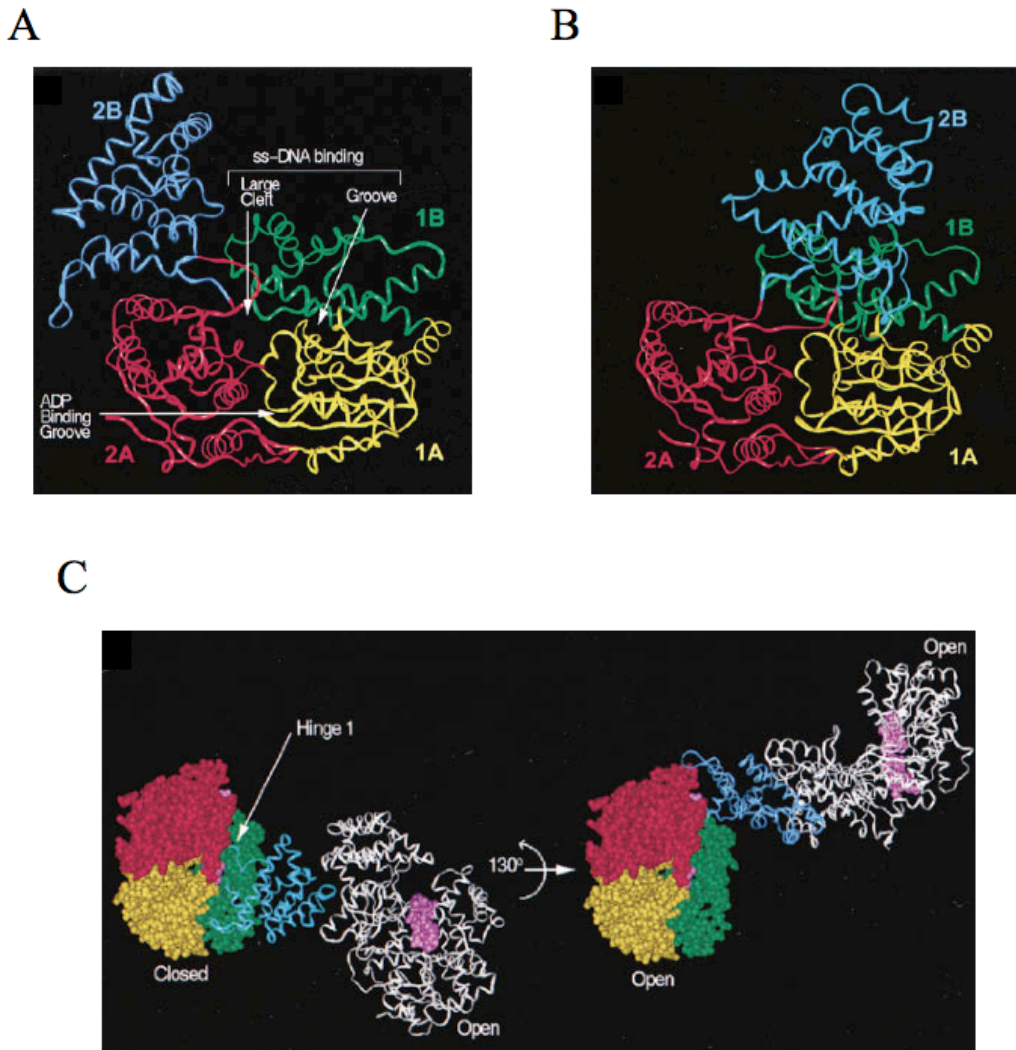
Figure 6.2: Open and closed conformations of Rep helicase

Structure of the Rep helicase in the open (A) and closed (B) conformation. The position of each subdomain, the ADP binding cleft, and the two sites for ssDNA binding are indicated in (A). Upon DNA binding, subdomain 2B rotates by 130 °, resulting in closure of the large ssDNA binding cleft (B). DNA has been omitted for clarity. (C) Hypothesised model of Rep dimerisation following a 130 ° rotation of subdomain 2B (blue ribbon) about hinge 1.This movement results in the closure of the ssDNA binding site. DNA is shown in pink space-fill. Adapted from (Korolev et al., 1997).

Waksman and coworkers addressed this point by suggesting that the large swivel motion of subdomain 2B may have a role to play in Rep dimerisation (Korolev et al., 1997) (figure 6.2 C). Indeed, such dimerisation may be linked to DNA translocation: perhaps the Rep dimer moves along DNA by a subunit-switching mechanism whereby the affinity for DNA alternates between each subunit of the dimer such that the protein creeps along the nucleic acid strand in a 'hand-over-hand'

mechanism (Korolev et al., 1997), as described for kinesin (Howard, 1996). This argues in favour of an active-rolling mechanism for the Rep helicase.

Apart from the large 130 º rotation of subdomain 2B, very little structural modulations of Rep were detected upon DNA or ADP binding. This makes it difficult to ascertain how ATP binding and hydrolysis may be coupled to DNA translocation and unwinding. However, the authors report that the close spatial arrangement of the nucleotide and DNA binding sites may allow the protein to couple these processes using subtle methods (Korolev et al., 1997). It may also be argued that the conformational changes required for the Rep helicase to transduce ATP binding and hydrolysis into DNA translocation and unwinding are too small to accurately determine through examination of structural data.

In contrast to SF1 helicases, DNA unwinding by members of SF2 is much less understood. This family is much larger and more diverse than the SF1 helicases, and includes the replication fork reversal protein RecG from *E. coli*, RecQ family DNA repair helicases, Mus308/ Hel308/ Polθ helicases, as well as a large variety of RNA helicases. To elucidate the details of DNA translocation and unwinding by SF2 DNA helicases, the repair helicase Hel308 from *S. solfataricus* strain PBL2025 (Sso PBL2025 Hel308) was chosen as a model protein. The chapter describes the use of two biophysical techniques, namely electron paramagnetic resonance (EPR) spectroscopy and single-molecule fluorescence resonance energy transfer, combined with current X-ray crystallography data, to understand the mechanistic properties of Hel308 activity.

6.2   SUMMARY OF THE HEL308 STRUCTURE

The high-resolution crystal structures of Sso PBL2025 Hel308, *Archaeoglobus fulgidus* Hel308 (Afu Hel308), and the Hel308 homolog Hjm from *Pyrococcus furiosus* (Pfu Hjm) have been solved in apo form, in complex with DNA (Afu Hel308) and in complex with nucleotide cofactors (Pfu Hjm) (Buttner et al., 2007; Richards et al., 2008b; Oyama et al., 2009). The structural data reveals that Hel308 is a monomeric, five-domain protein. Domains 1 and 2 are the classic RecA-like motor domains and comprise the conserved helicase motifs (figure 6.3 A). Nucleotide binding occurs within the cleft between the interface of domains 1 and 2. Domains 1 – 4 form the central core of the protein and are arranged such that a pore runs through the middle. The displaced 3' tail of the ssDNA weaves through the protein via the central pore,

interacting with each domain before finally contacting domain 5 (figure 6.3 B). Domain 5 forms a classic helix-hairpin-helix fold typically associated with ssDNA binding.

The structure of Afu Hel308 (Buttner et al., 2007) revealed that nucleic acid enters the protein between domains 2 and 4, with strand separation occurring on domain 2 by way of a projecting loop composed of a two-stranded β-sheet (figure 6.3 C), akin to that previously observed in the structure of the hepatitis C virus (HCV) NS3 helicase (Kim et al., 1998). The central core of the protein is not covalently closed suggesting that the structure could undergo conformational changes during ATP hydrolysis and/ or DNA binding, similar to the domain rearrangements observed in *B. stearothermophilus* PcrA and *E. coli* Rep helicase (Korolev et al., 1997; Velankar et al., 1999). The Hel308 DNA helicase has been implicated in the repair of stalled replication forks (Guy and Bolt, 2005; Richards et al., 2008b). To accommodate such bulky nucleic acids, the gap between domains 2 and 4 is predicted to open to enable entry of the DNA molecule (Buttner et al., 2007). The structural data may be consistent with this hypothesis since the majority of interactions found between domains 2 and 4 consisted of weak salt bridges and polar contacts (Buttner et al., 2007).

Upon strand separation, bases of the 3' tail stack with residues of motifs Ia and Ib on domain 1 and motifs IV, IVa and IVb on domain 2. These dual DNA binding sites suggest that Hel308 may unwind and translocate along DNA via an inchworm mechanism (Buttner et al., 2007). Stacking interactions also occur between bases of the 3' tail and residues of a central helix of domain 4. These interactions occur at positions that oppose those contributed by domains 1 and 2, suggesting that this helix may act as a 'ratchet' to power translocation: DNA translocation and unwinding may occur by way of ATP-dependent conformational changes that shift the relative positions of domain 1, domain 2 and the 'ratchet' helix of domain 4 (Buttner et al., 2007).
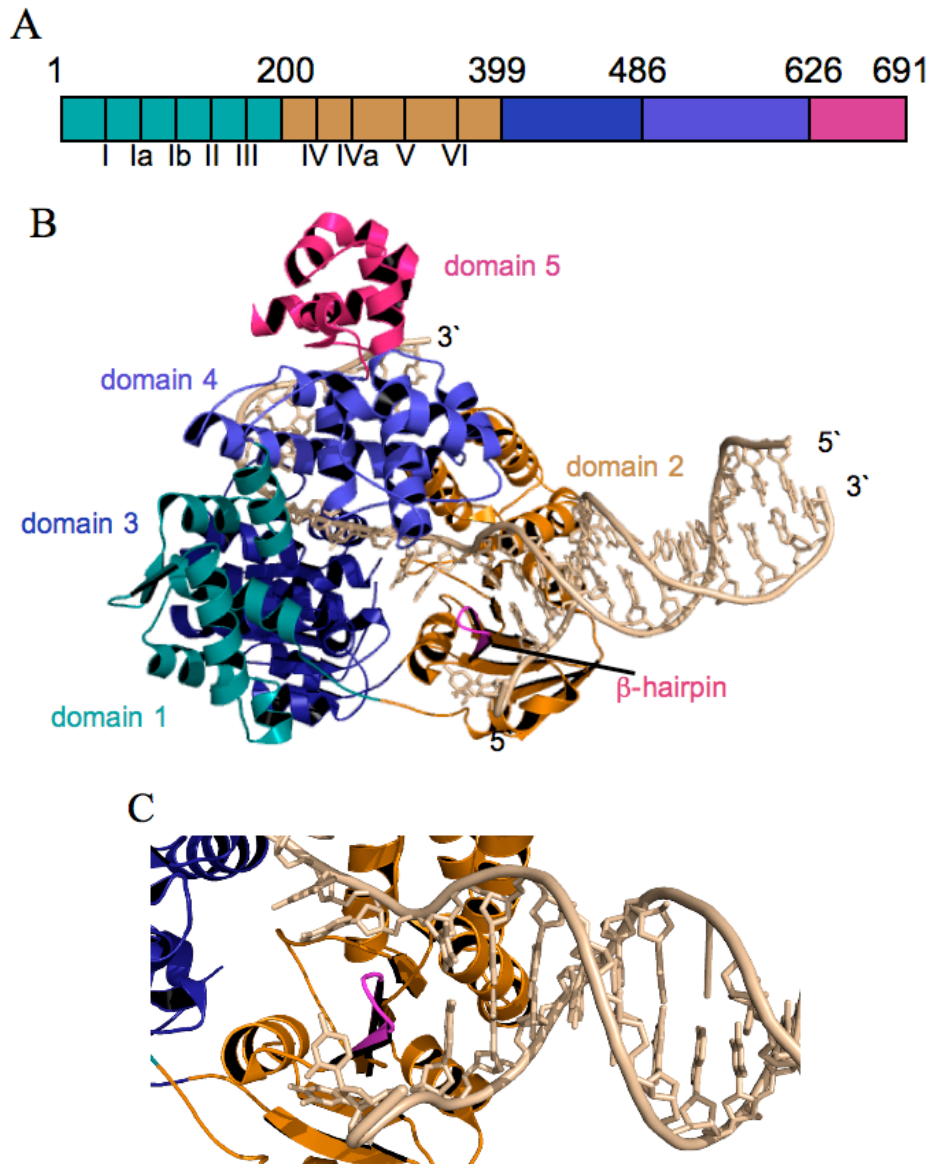
Figure 6.3: Structure of Afu Hel308

(A) Domain organisation of Afu Hel308. Numbering refers to the protein domain boundaries and roman numerals correspond to the helicase motifs and their approximate positions (Buttner et al., 2007). Domains coloured as follows: domain 1, jade; domain 2, gold; domain 3, dark blue; domain 4, light blue; domain 5, pink. (B) Ribbon structure of Hel308 in complex with DNA (15 base pair duplex with a 10 nucleotide single stranded region), shown as beige sticks (PDB code: *2p6r*). β-hairpin is shown in magenta. Domains coloured as above. (C) Close-up of domain 2 containing the β-hairpin loop, ideally positioned to disrupt base pairs of duplex DNA.

Based on this structural information, Hopfner and coworkers proposed a model for DNA translocation and unwinding by Hel308 (figure 6.4), in which ATP binding induces a conformational shift in domain 2 (Buttner et al., 2007). This movement is predicted to have two notable results: firstly, motif IV on domain 2 pushes

the DNA closer towards motif Ia on domain 1 and, secondly, domain 2 would also push on domain 4 by directly interacting with the ratchet helix. Since bases of the 3' tail stack with residues of domain 1, domain 2 and within the ratchet helix of domain 4, these movements are predicted to alter the position of the DNA such that the interaction between DNA and motif Ia weakens to allow the 3' tail to slide across the top of domain 1. Upon ATP hydrolysis, the structural domains are able to return to their original positions, enabling domain 2 to rebind at the ss-/ ds-DNA junction with the β-hairpin ideally positioned to melt another base pair of the duplex (Buttner et al., 2007).
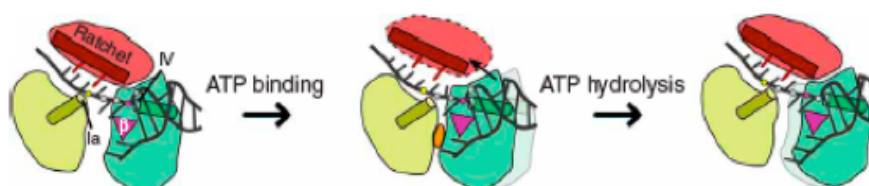
Figure 6.4: Model proposed for processive unwinding by archaeal Hel308

Binding of ATP (orange oval) to the cleft between domain 1 (yellow) and domain 2 (green) may trigger a conformational change such that domain 2 pushes on domain 4 (red) (indicated by arrow), via the ratchet helix (red helix), and domain 1 causing the 3' tail of DNA to slip across motif Ia (yellow helix). ATP hydrolysis then resets the position of domain 2 further upstream enabling the β-hairpin (pink triangle) to separate another base pair. Domains 3 and 5 have been omitted for simplicity. Taken from (Buttner et al., 2007).

This model predicts that Hel308 undergoes conformational changes during the hydrolysis and binding of ATP and DNA. However, the apo (Afu Hel308), DNA-bound (Afu Hel308) and nucleotide-bound structures (Pfu Hjm) of Hel308 superimpose well, revealing a very rigid structure that undergoes minimal conformational shifts upon ATP, ADP and DNA binding (figure 6.5). Hel308 may couple ATP binding and hydrolysis to DNA translocation and unwinding via subtle conformational changes that are not detectable using the current structural data. Indeed, the process of mechanochemical coupling by the Rep helicase has been difficult to elucidate from structural data (Korolev et al., 1997). Alternatively, the static views obtained of Hel308 by X-ray crystallography may simply capture the beginning and end points of the reaction cycle with the intervening period, potentially involving large-scale conformational changes, escaping analysis.
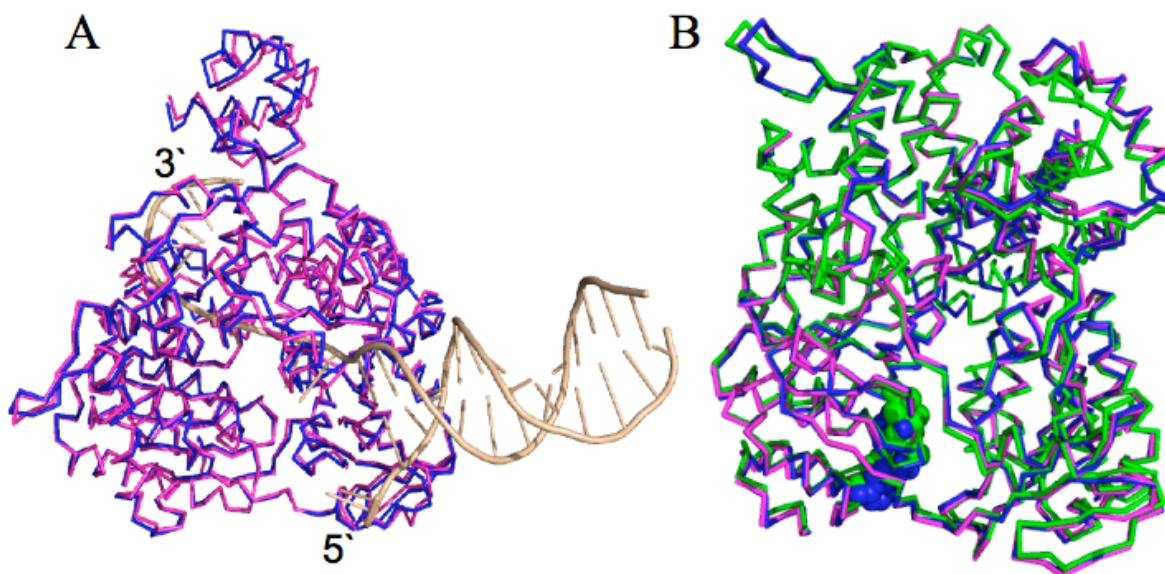
Figure 6.5: Superimposition of Hel308 in the presence and absence of nucleic acid and nucleotides

To investigate the structural flexibility of Hel308 during its reaction cycle, (A) the apo (magenta; PDB code: *2p6u*) and DNA-bound (blue; PDB code: *2p6r*) structures of Afu Hel308 were superimposed. The bound DNA molecule is coloured wheat. (B) The apo (magenta; PDB code: *2zj2*), ADP-bound (blue; PDB code: *2zj5*) and AMP-PNP (green; PDB code: *2zja*) structure of Pfu Hjm were also superimposed. The nucleotide cofactors are shown as space-filled spheres.

The structural work of Hel308 has highlighted key features that shall be addressed in this chapter. Of particular interest is the possibility of detecting conformational changes that occur during DNA binding and ATP binding and hydrolysis. In the former case, we are interested to investigate whether domains 2 and 4 move in relation to each other during DNA binding, to facilitate the entry of larger DNA molecules, e.g. stalled replication forks. Any movement that occurs between these two domains may support the proposed mechanism of action for Hel308 (figure 6.4), as proposed by (Buttner et al., 2007). In the latter case, we will investigate conformational changes between domains 1 and 2 during the reaction cycle; for instance, one aim is to determine whether an 'open' and 'closed' conformation of Hel308 exists during DNA binding and release, similar to that observed for *B. stearothermophilus* PcrA and *E. coli* Rep helicase. To render this possible, site-directed spin labelling and EPR spectroscopy have been employed since it offers extremely sensitive spatial resolution.

6.3   STRUCTURAL FLEXIBILITY OF HEL308

Defining inter-atomic and inter-molecular distances in biomolecules has been typically performed using X-ray crystallography (Tsvetkov et al., 2008). However, the formation of protein crystals required for this technique is both technically demanding and time limiting. EPR spectroscopy methods provide another avenue with which to explore protein structural dynamics. These methods do not require crystals and can be performed in either the liquid phase or in frozen solution, which are believed to be more reliable than crystals in terms of their ability to support biological activity (Tsvetkov et al., 2008). EPR spectroscopy was, therefore, chosen to investigate the structural flexibility of Hel308 during ATP hydrolysis and DNA binding. EPR spectroscopy study was performed in collaboration with Dr Olav Schiemann.

6.3.1   PELDOR EPR SPECTROSCOPY

Although Hel308 possesses two native cysteine residues, they are linked via a disulphide bond and unavailable for spin labelling, as determined by mass spectrometry (table 6.2). To investigate conformational changes that occur in Hel308 during ATP hydrolysis and DNA binding, paired residues were mutated to cysteine residues (K182C/ K409C; S302C/ S595C; L86C/ K284) and labelled with the nitroxide paramagnetic (1-oxyl-2,2,5,5-tetramethylpyrroline-3-methyl)-methanethiosulfonate spin label (MTSL), as outlined in Materials and Methods. The location of the MTSL and the motivation behind its attachment at specific sites in Hel308 is shown in table 6.1 and figure 6.6.

Table 6.1: MTSL attachment sites in Hel308. Inter-residue distance refers to the distance between the Cα atoms of each residue, as determined from the crystal structure of Hel308 (Buttner et al., 2007).

Site of MTSL	Inter-domain movement monitored	Inter-residue distance (Å)
C182/ C409	Domains 1 and 2 during ATP cycle	23
C302/ C595	Domains 2 and 4 during DNA binding	18
C86/ C284	Domains 1 and 2 during ATP cycle	27
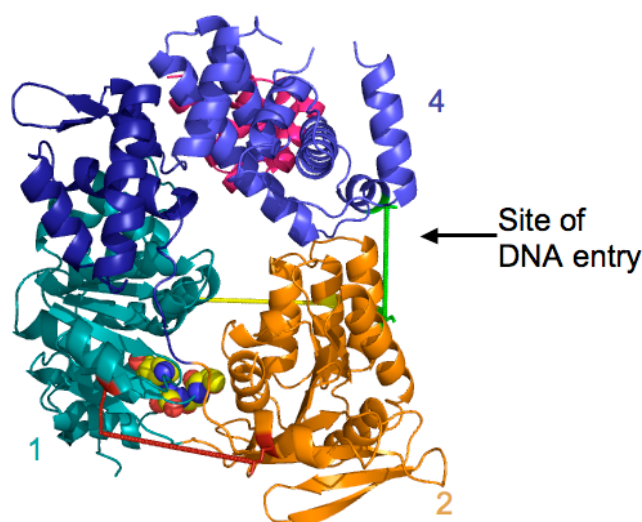
Figure 6.6: MTSL attachment sites in Sso PBL2025 Hel308

The location of the paired MTSL attachment sites in Hel308 (domains coloured as figure 6.3). Lines (C182/ C409, red; C302/ C595, green; C86/ C284, yellow) highlight inter-residue distance. DNA enters the molecule between domains 2 and 4, as indicated. The Walker A box is shown in space-fill and represents the site of ATP binding.

Attachment of MTSL to residues C182 and C409 will reveal distance changes between domains 1 and 2 during ATP binding and hydrolysis. This can be investigated by incubating the protein with a non-hydrolysable analog of ATP, AMPPNP, to trap the protein in the ATP-bound substrate configuration, or incubating the protein with ADP to trap the protein in the product conformation. The model proposed by Hopfner and coworkers predicts that ATP binding will bring domains 1 and 2 close together (Buttner et al., 2007), which may result in an opposing motion at the opposite end of domains 1 and 2. Spin labelling of residues C86 and C284, located at the top end of domains 1 and 2, will help to answer this question. Finally, MTSL-labelling of residues C302 and C595 will provide insight into any movement that occurs between domains 2 and 4 during DNA binding. The structure of Afu Hel308 implies that the cleft between these two domains must enlarge to accommodate bulky DNA molecules (Buttner et al., 2007). Furthermore, the relative positions of domains 1, 2 and 4 are expected to change during ATP hydrolysis and DNA binding. The position of MTSL in the three Hel308 cysteine mutants were selected to determine the accuracy of the model (figure 6.4) proposed by Hopfner and coworkers (Buttner et al., 2007).

Mass spectrometry was used to confirm successful attachment of MTSL to Hel308 (table 6.2). Addition of one spin label ($C_{10}H_{18}NO_3S_2$), and concomitant loss of sulfinic acid (CH_3SO_2), increases the protein mass by 185 Da (figure 2.3).

Table 6.2: Mass spectrometry of Hel308 before and after MTSL addition. Hel308 was incubated at 4 ºC for 2 hr in the presence (+ MTSL) or absence (- MTSL) of MTSL. ESI-TOF mass spectrometry was used to confirm attachment of MTSL to the protein. Addition of one spin label to Hel308 results in a theoretical mass increase of 185 Da.

Protein	Mass of Hel308		Mass change (Da)	Number of MTSL spin labels added
	- MTSL	+ MTSL		
WT	84251.2	84261.8	+ 10.6	0
C182/ C409	84375.7	84747	+ 371.3	2
C302/ S595	84469.1	84825.7	+ 356.6	2
C86/ C284	84374.6	84759.9	+ 385.3	2

Helicase assays were performed, as outlined in Materials and Methods, to ensure that MTSL-labelled Hel308 cysteine mutants retained the ability to unwind DNA, as observed for the WT protein. Figure 6.7 reveals that all spin-labelled Hel308 proteins were active helicases *in vitro*.
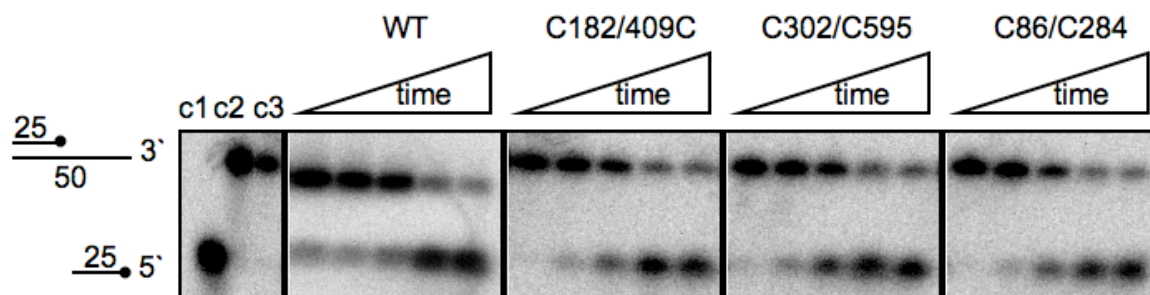
Figure 6.7: Helicase activity of MTSL-labelled Hel308 proteins

MTSL-labelled Hel308 cysteine mutants (500 nM) were incubated at 60 ºC with 10 nM 5' [^{32}P]-end labelled 3' overhang DNA (black circle; radiolabel; number of nucleotides shown). At specific time points (10, 30, 60, 120, 180 s), reactions were stopped and separated on native 12% (v/v) polyacrylamide: TBE gels. c1, size marker; c2, end-point reaction in absence of protein; c3, end-point reaction in absence of ATP/ MgCl$_2$. The figure shows representative gels obtained from experiments performed at least three times using different preparations of MTSL-labelled protein.

6.3.2   HEL308 IS STRUCTURALLY RIGID DURING REACTION CYCLE

PELDOR spectroscopy was performed on apo-Hel308 (Hel308 in the absence of nucleotides or nucleic acid), and resulting spectra revealed inter-spin distances of 30.8 Å, 23.2 Å, and 28.6 Å for Hel308 with MTSL attached to residues C182/ C409, C302/ C595, and C86/ C284 respectively (figure 6.8). These distances are slightly higher than the corresponding Cα-Cα distances (table 6.1) since they refer to the distance between MTSL spin labels, which are flexible in solution.
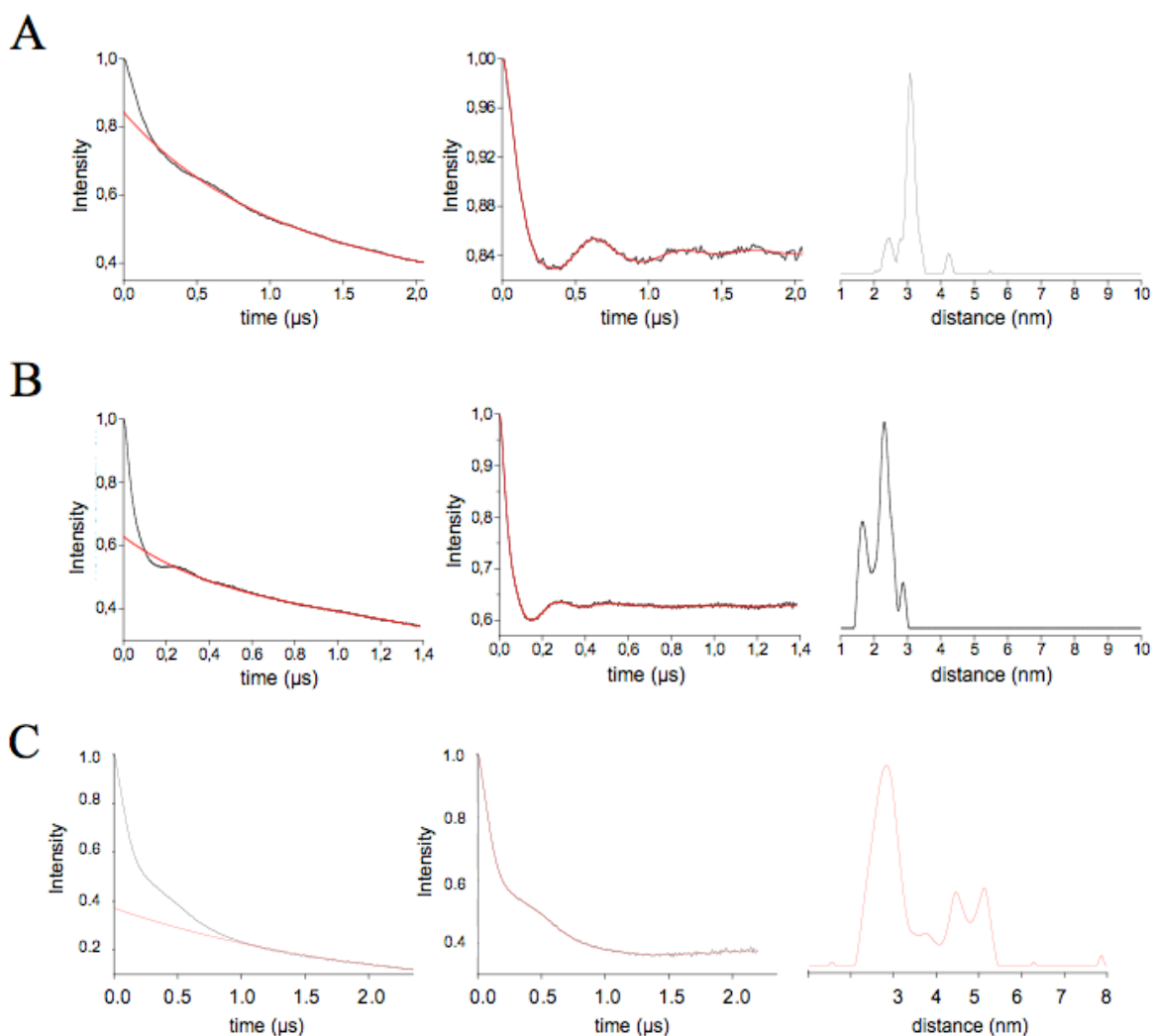
Figure 6.8: PELDOR spectroscopy of MTSL-labelled Hel308

EPR spectra of Hel308 with MTSL spin labels located on residues C182/ C409 (A), C302/ C595 (B), or C86/ C284 (C). In all cases, the original time trace is shown on the left, including intra- (black line) and inter-molecular (red line) spin spectra. The middle time trace shows the corrected intra-molecular spectra (intra-molecular spectrum divided by the inter-molecular spectrum). The right trace shows the intra-molecular spin distance, calculated via Tikhonov regularisation of the corrected intra-molecular time trace.

To investigate potential conformational changes during the Hel308 reaction cycle, the MTSL-labelled proteins (150 μM) were incubated with 20X excess ADP/ MgCl$_2$ or AMP-PNP/ MgCl$_2$ (non-hydrolysable analog of ATP) or 5X excess 34-mer DNA (Appendix 1, table A1.8), as outlined in Materials and Methods. Hel308 labelled on C182/ C409 (figure 6.9 A) and C302/ C595 (figure 6.9 B) appeared conformationally unchanged upon addition of these components, with no measurable shift in the intra-molecular spin distance (figure 6.12 and Appendix 6). This is

consistent with structural data of Hel308 in complex with DNA (Buttner et al., 2007) or nucleotide cofactors (Oyama et al., 2009).
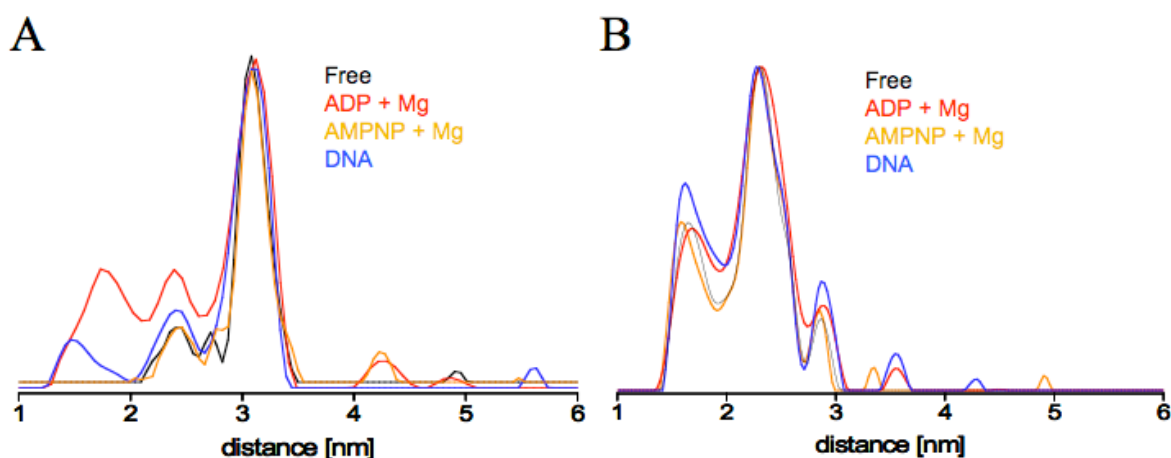
Figure 6.9: Intra-molecular spin distances during Hel308 reaction cycle

Distance changes between the (A) motor domains (measured using MTSL-labelled C182/ C409) and (B) domains 2 and 4 (measured using MTSL-labelled C302/ C595) in the presence of ADP/ MgCl$_2$, AMPPNP/ MgCl$_2$ or DNA. For each protein, the data represents triplicate measurements using a single protein preparation.

In contrast, the distance between MTSL spin labels on residues C86 and C284 of Hel308 increased from 28.6 Å (apo-Hel308) to 47 Å upon addition of ADP/ MgCl$_2$ (figure 6.10 A and Appendix 6). Importantly, the accompanying PELDOR spectrum lacked signal modulation (figure 6.10 A; left), a pre-requisite for reliable extraction of distance data. The experiment was thus repeated using 150 μM ADP/ MgCl$_2$ (representing a 1:1 molar concentration ratio with protein), resulting in EPR signal modulation and a distance measurement of 27.4 Å (figure 6.10 B); this equates to a small but reproducible shift between domains 1 and 2 of – 1.2 Å upon addition of ADP/ MgCl$_2$ (figure 6.12).
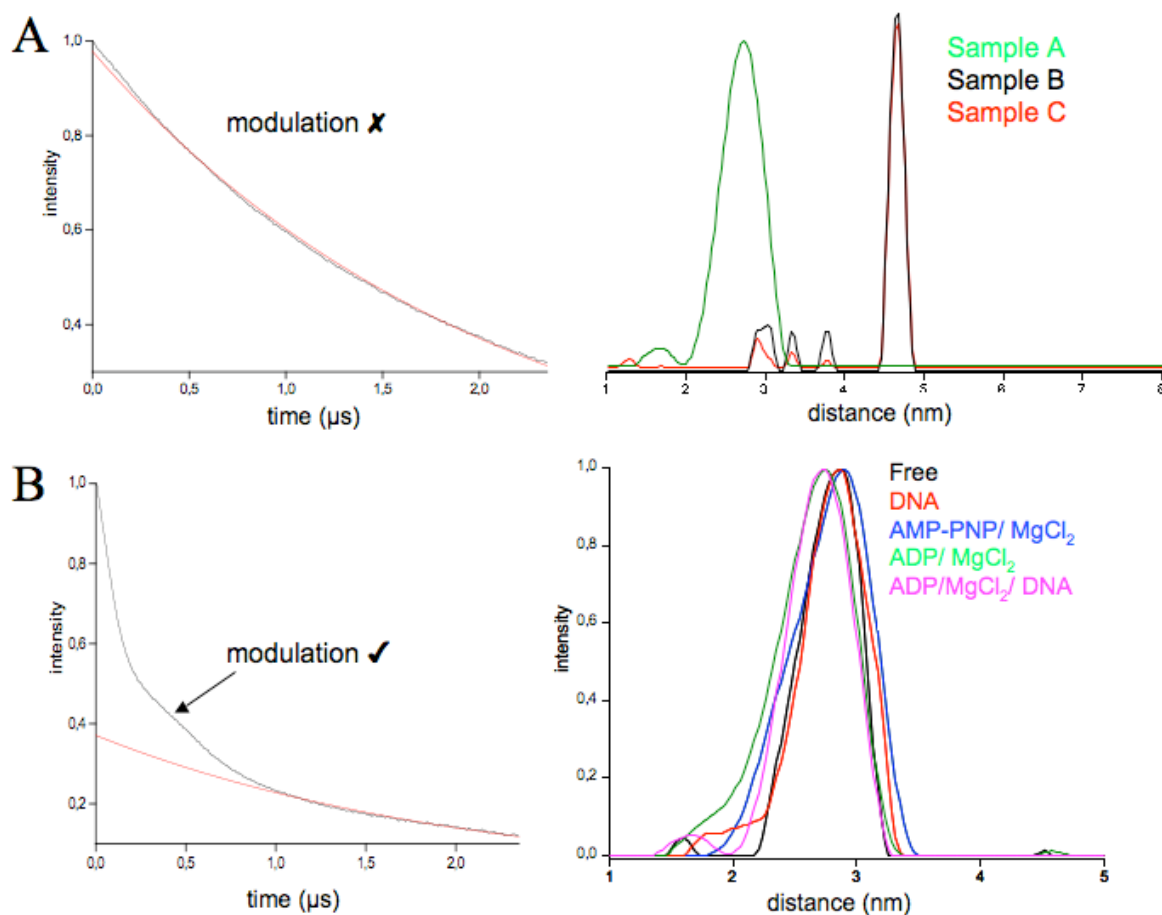
Figure 6.10: Distance between domains 1 and 2 during Hel308 reaction cycle

Hel308 with MTSL spin labels attached at residues C86 and C284. For both (A) and (B), the original time trace is shown on the left and the distance between the MTSL spin labels is shown on the right. (A) Hel308 (150 µM) in the absence (sample A) and presence of 20X excess ADP/ MgCl$_2$ (sample B). Control (sample C), ADP was added to a control sample containing protein and MgCl$_2$, previously revealing distances equal to protein only. (B) Hel308 (150 µM) in the presence of equimolar concentrations of MgCl$_2$ only, ADP/ MgCl$_2$, AMP-PNP/ MgCl$_2$ or 5X excess DNA. Each curve represents triplicate measurements using a single protein preparation.

The large apparent 18.4 Å distance change between MTSL labels located on residues C86 and C284 in the presence of 20X excess ADP/ MgCl$_2$ required further investigation. The absence of EPR signal modulation suggested that the spin labels had become very far apart, perhaps as a result of protein denaturation upon addition of high concentrations of nucleotide. To assess the impact of ADP/ MgCl$_2$ on the secondary structure of Hel308, far UV CD spectroscopy was performed on the MTSL-labelled protein before and after the addition of either equimolar or 20X molar concentrations of ADP/ MgCl$_2$. The spectra revealed that the overall structure of Hel308 remained

unchanged from the apo form, regardless of protein: ADP/ MgCl$_2$ molar concentration ratio used in the reaction (figure 6.11).
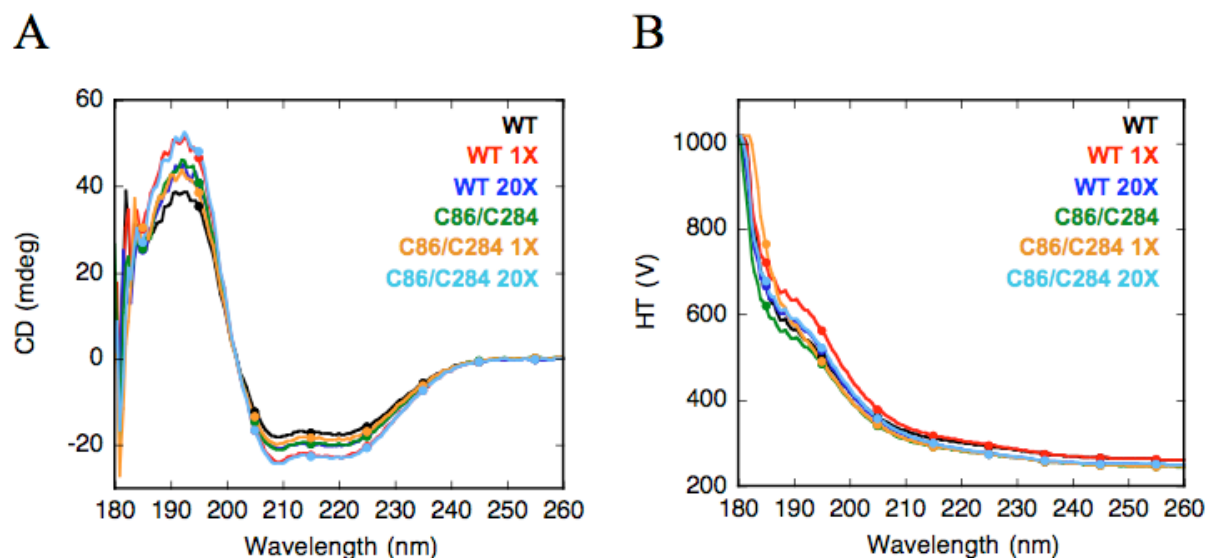
Figure 6.11: CD spectroscopy of MTSL-labelled Hel308 with ADP/ MgCl$_2$

(A) Far-UV CD measurements of WT and C86/ C284 Hel308 (0.5 mg/ ml) were recorded between 260 nm and 180 nm in the presence and absence of 0.5 mg/ml (1X) or 10 mg/ ml (20X) ADP/ MgCl$_2$. (B) High-tension voltage (HT) plot. Each curve in the graph represents the mean of six measurements using the same protein preparation.

Based on structural (Buttner et al., 2007; Oyama et al., 2009) and now, PELDOR spectroscopy data, it is unlikely that large conformational changes occur between domains 1, 2 and 4 during the Hel308 reaction cycle. The small 1.2 Å shift between domains 1 and 2 upon ADP binding may allude to the closure of the cleft between the motor domains in a manner akin to that observed with PcrA (Velankar et al., 1999). While this technique is very sensitive to distance changes, we cannot exclude the possibility that the positions of the MTSL spin label within the protein were not optimal for detecting small conformational changes, as suggested in the model (figure 6.4) proposed by Hopfner and coworkers (Buttner et al., 2007). Furthermore, we cannot exclude the possibility that EPR spectroscopy of AMPPNP- and ADP-bound Hel308 states reveals a static representation of the beginning and end of the Hel308 reaction cycle, failing to report on rapid conformational changes that may occur during the intervening period. For this reason, future EPR spectroscopy studies with Hel308 aim to use a quench-flow methodology in which nucleotide cofactors and nucleic acid molecules can be added to Hel308 in real-time during data acquisition.

C182/ C409	Distance (Å)
apo	30.8
+ MgCl$_2$	30.8
+ ADP/ MgCl$_2$	31.0
+ AMP-PNP/ MgCl$_2$	31.0
+ DNA	31.0

C302/ C595	Distance (Å)
apo	23.2
+ MgCl$_2$	23.0
+ ADP/ MgCl$_2$	23.0
+ AMP-PNP/ MgCl$_2$	23.0
+ DNA	23.0

C86/ C284	Distance (Å)
apo	28.6
+ MgCl$_2$	28.6
+ ADP/ MgCl$_2$[†]	27.4
+ AMP-PNP/ MgCl$_2$	28.6
+ DNA	28.6
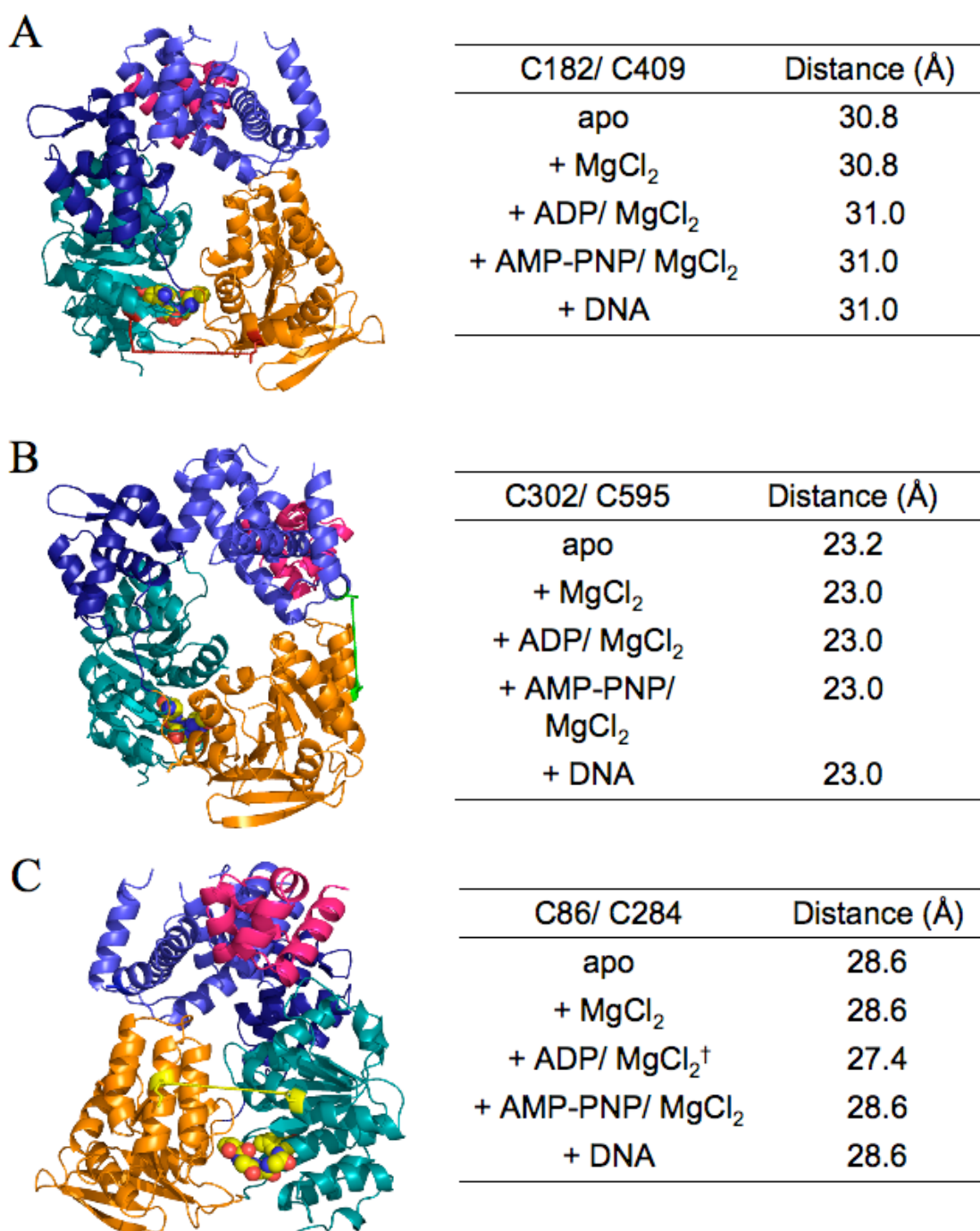
Figure 6.12: Distance changes between domains during Hel308 reaction cycle

Protein (150 μM) was incubated alone (apo) or with MgCl$_2$ (3 mM), ADP/ MgCl$_2$ (3 mM, except † where only 150 μM was added), AMP-PNP/ MgCl$_2$ (3 mM), or DNA (750 μM). PELDOR EPR data was collected and inter-spin distances are noted. (A) C182/ C409 Hel308; (B) C302/ S595 Hel308; (C) C86/ C284 Hel308, with the distance measured highlighted on the crystal structure with red, green and yellow lines, respectively. Error values reflecting distance distributions have not been noted because the distances are determined from a computer simulation and not directly from the experiment and, therefore, are not subject to noise.

196

6.4 SINGLE-MOLECULE FLUORESCENCE RESONANCE ENERGY TRANSFER STUDY OF HEL308

Single-molecule techniques were employed to measure the kinetic parameters of DNA binding, unwinding and translocation by individual Hel308 molecules with millisecond time resolution. Single-molecule measurements also facilitate the detection of rare conformational intermediates and the distribution of kinetic rates within a heterogeneous population of molecules during a reaction cycle (Ha, 2001). The high spatial and temporal sensitivity of fluorescence resonance energy transfer (FRET) spectroscopy made this particular technique attractive to study Hel308 at the single-molecule level. FRET exists as a powerful spectroscopic method for measuring distances between 10 and 80 Å, and is therefore relevant to the measurement of biomolecules (Ha, 2001; Lakowicz, 2006; Selvin, 2008). In recent years, several publications have reported the unique advantages of single-molecule FRET (smFRET) spectroscopy to probe the biochemical properties of DNA helicases including the hepatitis C virus NS3 helicase (Myong et al., 2007), the BLM helicase (Yodh et al., 2009) and the SF1 Rep helicase (Rasnik et al., 2004). This reinforces the high value of using smFRET spectroscopy to understand Hel308 biochemistry in greater detail.

6.4.1 LABELLING DNA AND HEL308 WITH FLUORESCENT DYES

To observe single molecules, DNA and Hel308 were labelled with cy3 and cy5 fluorescent dyes, respectively, as outlined in Materials and Methods. DNA oligonucleotides were synthesised with a biotin moiety at the 3' end to permit surface immobilisation and an amino group at the 5' end of the same oligonucleotide (figure 6.13 A) or a complementary oligonucleotide (figure 6.13 B) to enable conjugation to an amine-reactive cy3 dye. The resulting cy3-ssDNA (50T) and cy3- 3' overhang DNA substrates were purified by acrylamide gel electrophoresis and labelling efficiencies of 86% and 80%, respectively, were calculated from the absorbance spectrum of the final sample.
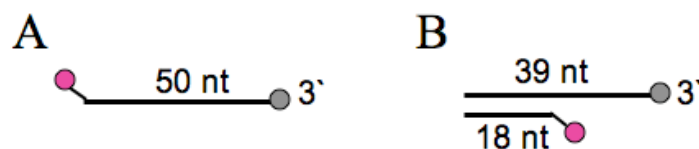
Figure 6.13: DNA substrates used for FRET experiments

Cy3-ssDNA (A) and cy3- 3'overhang (B) DNA substrates were generated. Grey circle, biotin; pink circle, cy3 dye. Lengths of oligonucleotides are indicated. Sequences of oligonucleotides are shown in Appendix 1 (table A1.6), entitled 50T (50 nt strand), sm2 (18 nt strand) and sm3 (39 nt strand).

Labelling of Hel308 with a thiol-reactive cy5-dye required site-directed mutagenesis of the protein to substitute residue S595 for a cysteine residue. This residue projects from domain 4 towards domain 2 (figure 6.14 A) such that it is positioned close to where DNA is expected to enter the protein (Buttner et al., 2007). The concentration and labelling efficiency of the protein (96%) was determined from the absorbance spectrum. A helicase assay, performed as described in Materials and Methods, confirmed that the cy5-labelled protein unwound DNA with an efficiency comparable to the WT (unlabelled) protein (figure 6.14 B).
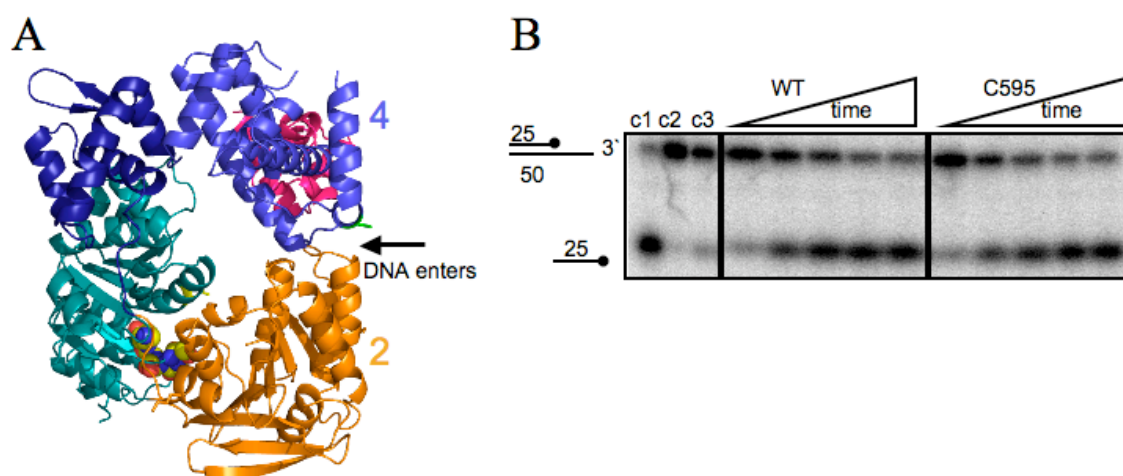
Figure 6.14: Cy5-labelled Hel308 helicase activity

(A) Cy5 was attached to Hel308 via the thiol group of residue C595 (represented as green stick). This residue is positioned between domains 2 and 4 (labelled), where DNA enters the protein (Buttner et al., 2007). (B) The cy5-labelled Hel308 protein (500 nM) unwound a 5' [^{32}P]-end labelled 3' overhang DNA substrate at 60 °C (time points: 10 s, 30 s, 60 s, 120 s, 180 s) at a rate comparable to the WT protein (unlabelled; 500 nM). c1, size marker; c2, end-point reaction in absence of protein; c3, end-point reaction in absence of ATP/ MgCl$_2$. Gel images are representative of triplicate experiments using separate cy5-labelled protein preparations.

6.4.2 BULK FRET ANALYSIS OF HEL308

Bulk FRET spectroscopy was performed to investigate the FRET signal obtained between cy5-Hel308 and each cy3-DNA substrate (figure 6.13). FRET efficiencies cannot be confidently determined from ensemble experiments because they do not account for incomplete labelling of reaction components, the presence of inactive molecules, and the possibility of more than one protein molecule binding to a single DNA molecule (Rasnik et al., 2004). Therefore, bulk FRET experiments were performed to simply investigate the ability to detect FRET between cy3-DNA and cy5-Hel308 molecules.

6.4.2.1 BULK FRET ANALYSIS OF DNA BINDING

FRET between cy5-Hel308 and cy3-ssDNA was determined in the absence of ATP/ MgCl$_2$. Cy5-Hel308 was titrated into buffer containing 40 nM cy3-ssDNA and the fluorescence intensity was determined by excitation of the cy3 dye at 547 nm. Data was collected between 557 nm and 700 nm. Increasing concentrations of cy5-Hel308 resulted in a decrease in donor (cy3-DNA) emission and concomitant increase in the acceptor (cy5-protein) emission signal, i.e. the ratio between acceptor and donor emission (with respect to a control sample lacking the cy5 acceptor) increased as a result of FRET (figure 6.15 A-B). A competitive titration of unlabelled Hel308 into a solution containing cy5-Hel308 (30 nM) bound to cy3-ssDNA (30 nM) was also performed. However, cy5-Hel308 did not appear to be displaced from the cy3-DNA using concentrations of unlabelled protein up to 2.5 µM (figure 6.15 C-D). It is possible that at high concentrations of cy5-Hel308, interactions between cy3 and cy5 affect the binding affinity of Hel308 for DNA such that higher concentrations of unlabelled protein are required to displace the pre-bound labelled protein. The direct titration experiment performed, however, suggested that binding of cy3-ssDNA by cy5-Hel308 produces a measurable FRET signal.
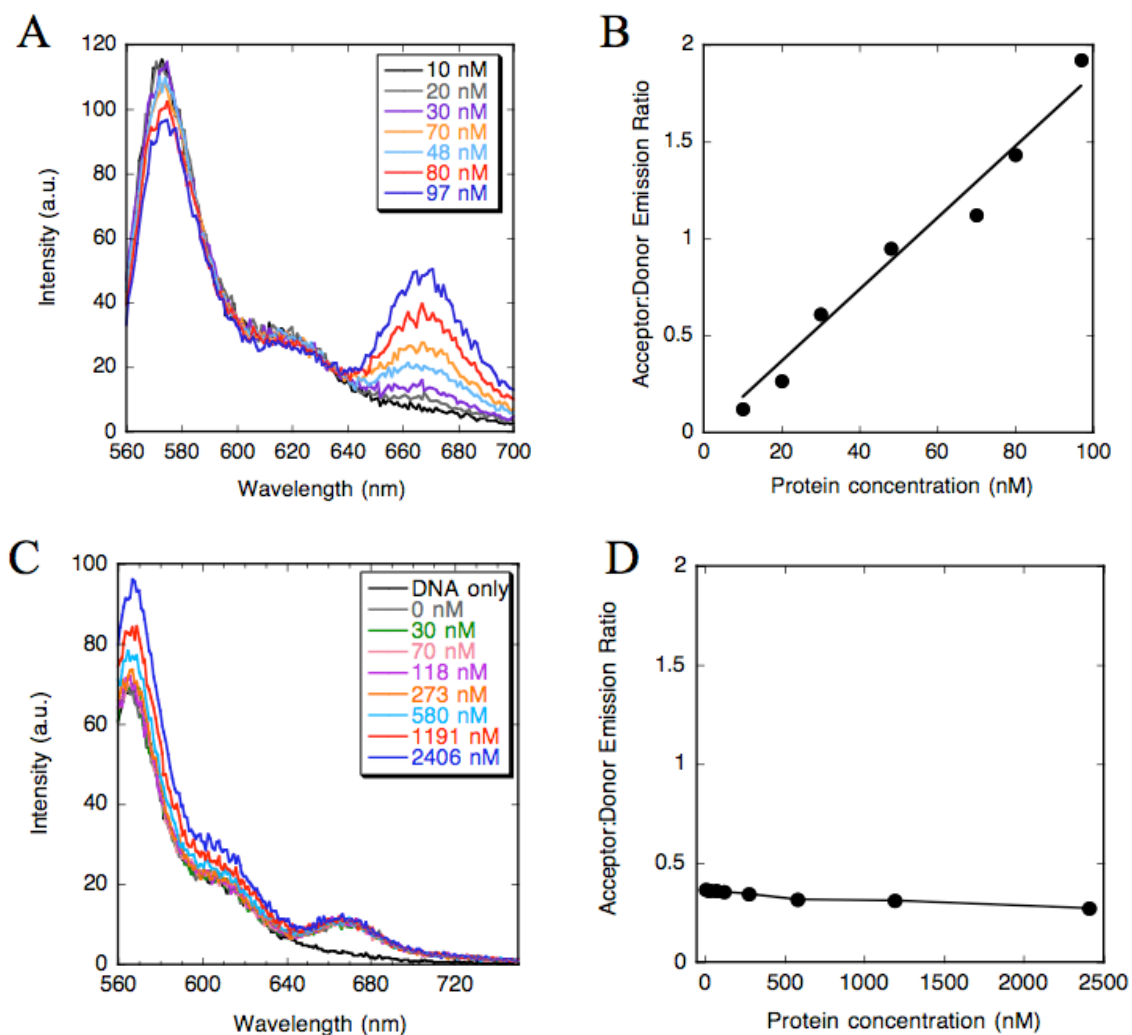
Figure 6.15: Bulk FRET analysis of DNA binding by Hel308

(A) Direct titration of cy5-Hel308 into cy3-ssDNA. The sample was excited at 547 nm and fluorescence emission was recorded between 557-700 nm. (B) Ratio of acceptor to donor emission (with respect to a control sample lacking cy5 molecules) with increasing cy5-protein concentration. This plot was constructed by measuring the ratio of the area under the donor (560.9 – 567.35 nm) and acceptor (650 – 680 nm) peaks in comparison to a control sample lacking cy5 molecules. (C) Competitive titration of unlabelled Hel308 into buffer containing cy3-ssDNA/ cy5-Hel308. Sample was excited at 547 nm and emission collected between 557-700 nm. (D) The ratio of acceptor to donor emission (with respect to a control sample lacking cy5 molecules) with increasing unlabelled protein concentration. This plot was constructed as described for (B). These experiments were performed once using a single protein preparation.

FRET between cy5-Hel308 and cy3-DNA requires protein concentrations approaching the dissociation constant of Hel308 for ssDNA (Richards et al., 2008b). This reflects the distance-dependent nature of FRET; for FRET to occur, cy5-Hel308 must bind ssDNA within 20-25 nucleotides of the cy3 label. Low protein concentrations

will limit the number of protein molecules bound within this region, and thus limit the detection of DNA binding by Hel308 via FRET.

6.4.2.2   BULK FRET ANALYSIS OF DNA TRANSLOCATION

Bulk FRET spectroscopy was used to determine the efficacy of using cy5-Hel308 and cy3-ssDNA to measure the DNA translocation activity of the protein. Reactions were performed in a final reaction volume of 150 µl containing 40 nM cy3-DNA and 250 nM cy5-protein. In the absence of ATP/ MgCl$_2$, these conditions permit DNA binding by Hel308 but not DNA translocation. An increase in FRET upon addition of ATP/ MgCl$_2$ may reflect 3'-5' translocation of cy5-protein towards the 5'-cy3 labelled end of the DNA molecule, whilst a decrease in FRET would suggest that the protein had translocated in a 5'-3' direction or had dissociated from the DNA. Alternatively, an increase in FRET upon addition of ATP/ MgCl$_2$ may simply reflect stronger DNA binding by the protein.

Multi-wavelength scanning was performed to collect the emission at 570 nm (donor) and 670 nm (acceptor) simultaneously for 10 min following excitation at 547 nm. Before the addition of ATP/ MgCl$_2$, cy3 donor molecules emitted more energy than the cy5 acceptor molecules (figure 6.16). Injection of ATP/ MgCl$_2$ led to an increase in acceptor emission reflecting increased FRET efficiency. To verify Hel308 translocation along ssDNA, the sample was excited at 547 nm and the emission between 557 and 700 nm was recorded before and after the addition of ATP/ MgCl$_2$ (figure 6.16 B). This data was normalised to correct for the overall reduction in fluorescent intensity observed following addition of ATP/ MgCl$_2$ (reflecting sample dilution upon ATP/ MgCl$_2$ addition) and indicated that ATP/ MgCl$_2$ had triggered the transfer of excitation energy from donor to acceptor fluorophores. These experiments support the use of the cy3-ssDNA (figure 6.13 A) and cy5-Hel308 to measure translocation via FRET. Since FRET increased upon addition of ATP/ MgCl$_2$, it suggests that cy5-Hel308 moved along DNA, in a 3'-5' direction, bringing the cy5 dye into close proximity to the cy3 dye at the 5' end of DNA.
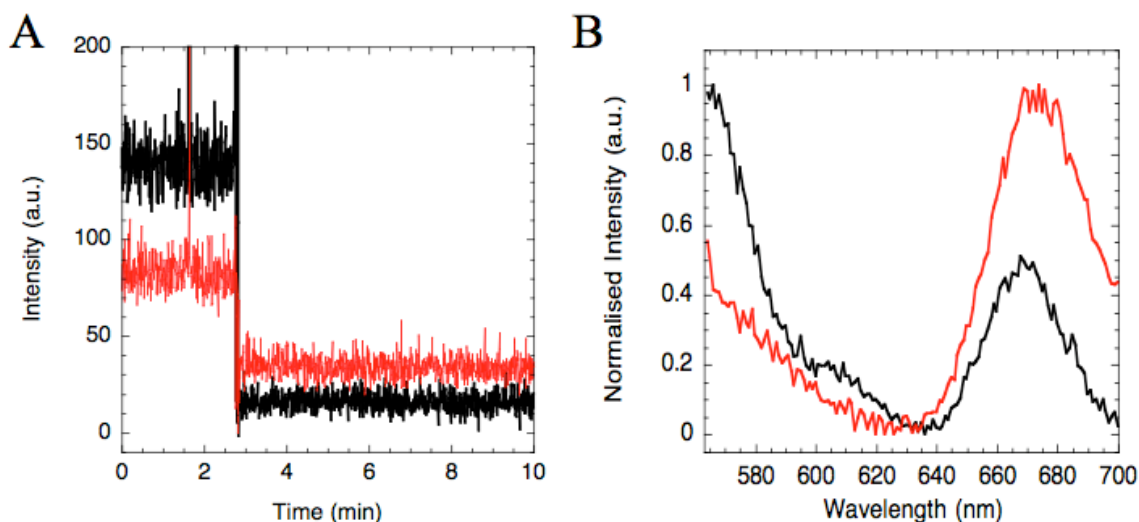
Figure 6.16: Bulk FRET analysis of Hel308 translocation along DNA

(A) Emission spectrum of donor (at 570 nm; black trace) and acceptor (at 670 nm; red trace) molecules for 10 minutes upon excitation of the donor at 547 nm. ATP/ MgCl$_2$ was injected at 3 min. (B) The sample was excited at 547 nm and the fluorescence intensity of energy emitted between 557 and 700 nm was recorded before (black trace) and after (red) addition of ATP/ MgCl$_2$. To correct for dilution effects, the data was normalised between 0 and 1. This experiment was performed once using a single protein preparation.

6.4.2.3 BULK FRET ANALYSIS OF DNA UNWINDING

Bulk methods were employed to ensure that unwinding of the cy3-labelled 3'-overhang DNA substrate (figure 6.13 B) by cy5-Hel308 was accompanied by a measurable FRET signal. The assay was performed in a final reaction volume of 150 µl comprising 40 nM DNA and 250 nM protein. Multi-wavelength scanning was employed to measure the fluorescence intensity of the energy emitted by both cy3- (570 nm) and cy5-dyes (670 nm) simultaneously for 16 min following excitation at 547 nm. ATP/ MgCl$_2$ (1 mM) was added to the sample after ~ 1.5, 4.5 and 12.5 min, each addition resulted in an increase in donor emission and concomitant decrease in acceptor emission, indicative of a loss of FRET (figure 6.17). This result was clarified by recording the emission spectrum between 557 and 700 nm, following excitation at 547 nm, before and after the addition ATP/ MgCl$_2$.
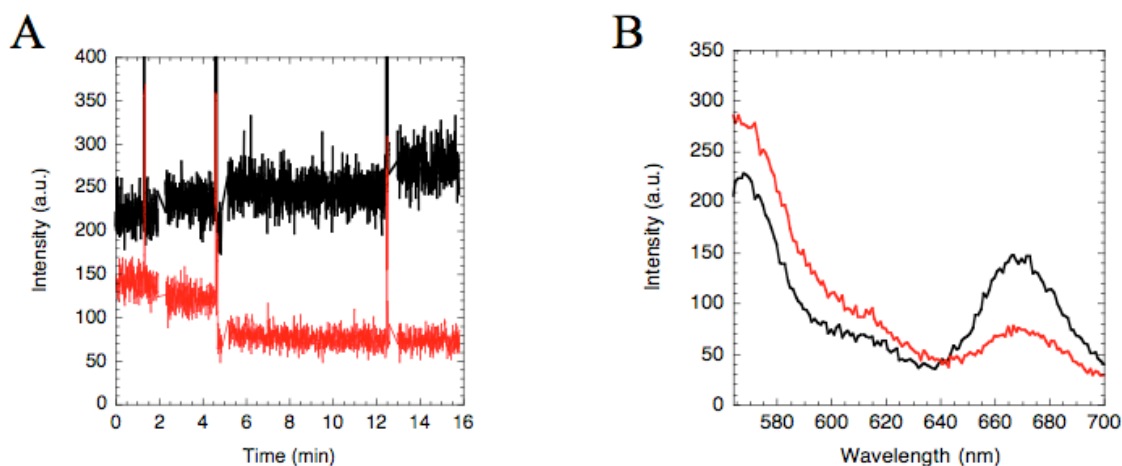
Figure 6.17: Bulk FRET analysis of DNA unwinding by Hel308

(A) Emission spectrum of donor (black) and acceptor (red) molecules following donor excitation at 547 nm. ATP/ MgCl$_2$ was injected at ~ 1.5, 4.5 and 12.5 min. (B) Fluorescent emission was collected between 557 and 700 nm before (black trace) and after (red trace) the addition of ATP/ MgCl$_2$. This experiment was performed once using a single protein preparation.

The high FRET values observed prior to the addition of ATP/ MgCl$_2$ denote the close proximity of the cy3- and cy5- dye of the DNA and protein, respectively, reflecting the ability of Hel308 to bind, but not unwind, DNA. ATP-dependent unwinding by Hel308 led to diminished FRET, reflecting the displacement of the cy3-labelled DNA strand and consequential separation of the cy3 and cy5 dyes. Despite repeated injection of ATP/ MgCl$_2$, acceptor emission was not completely lost (figure 6.17 A). This suggests that incomplete DNA unwinding occurred, possibly reflecting low helicase activity at the assay temperature of 20 ºC. Given that the host organism, *S. solfataricus*, thrives at temperatures above 70 ºC, it is unsurprising that optimal helicase activity was not observed at 20 ºC.

6.4.3 SMFRET ANALYSIS OF HEL308 ACTIVITY

Bulk FRET experiments revealed that measurable FRET signals were expected between cy5-Hel308 and both the cy3-ssDNA and cy3-labelled 3' overhang DNA substrate (figure 6.13). Therefore, smFRET spectroscopy was performed to provide greater insight into the heterogeneity of Hel308 biochemistry. The focus of this work was to confirm the directionality of Hel308-catalysed DNA translocation and unwinding and to investigate the rates (and the distribution of rates) of these processes.

6.4.3.1   IMMOBILISATION OF INDIVIDUAL BIOMOLECULES

To study the kinetic parameters of Hel308 for extended periods of time, surface immobilisation of individual molecules at a suitable density was necessary. One particular issue was how to immobilise positively-charged, DNA-binding protein molecules without non-specific adsorption to negatively-charged quartz microscope slides. Previous smFRET studies of DNA helicases have described the use of polyethylene glycol (PEG)-coated quartz slides to minimise this problem (Ha, 2001; Rasnik et al., 2004; Myong et al., 2007; Yodh et al., 2009). Accordingly, slides were prepared as outlined in figure 6.18 and in Materials and Methods.
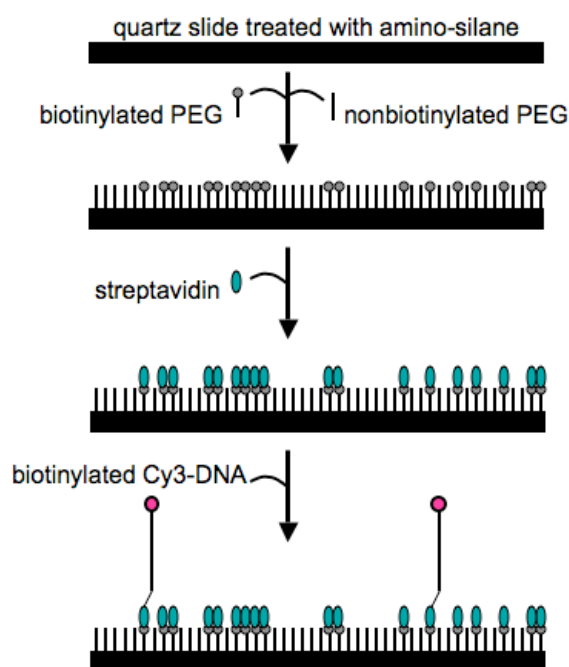
Figure 6.18: Preparation of microscope slides for smFRET experiments

Quartz microscope slides were coated with aminosilane to produce a positively charged surface to which negatively-charged PEG molecules could bind. A mixture of biotinylated and non-biotinylated PEG was added (ratio 1:100), providing available binding sites for streptavidin to immobilise a low-density surface of biotinylated cy3-DNA.

The efficacy of the slide preparation method to generate slides free from fluorescent impurities and immune from non-specific adsorption of Hel308 was tested (figure 6.19). Each PEG-coated slide was visualised prior to the addition of fluorescently labelled molecules, revealing few fluorescent spots per field of view (50 μm x 100 μm). In the absence of streptavidin, biotinylated cy3-DNA did not bind to the

slide; an adequate coating of DNA was achieved by addition of 200 pM cy3-DNA to a streptavidin-coated PEG slide. In the absence of DNA, few cy5-Hel308 (10 nM) bound non-specifically to the negatively charged quartz slide. However, in the presence of 200 pM DNA, adequate Hel308 immobilisation was achieved. These experiments support the use of PEG-coated quartz slides to visualise the specific interaction between fluorescently labelled DNA and protein molecules.
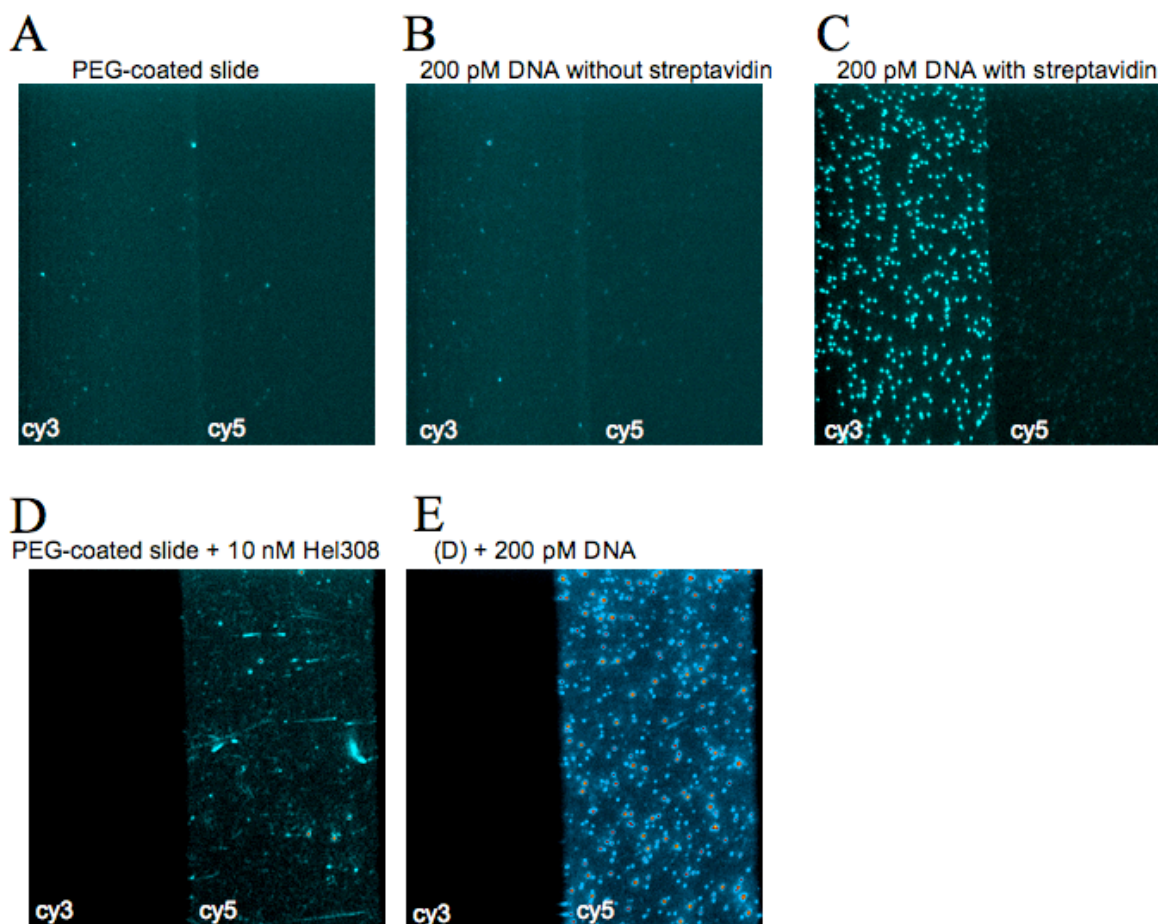
Figure 6.19: Immobilisation of DNA and Hel308 on PEG-coated quartz slides

A PEG-coated quartz microscope slide was imaged, by excitation at 532 nm (A-C), in the absence of fluorescently labelled DNA or protein (A) to determine the level of fluorescent impurities. (B) 200 pM cy3-DNA was injected onto the slide, in the absence (B) and presence (C) of streptavidin. The quartz slide was imaged by excitation at 638 nm (D-E): 10 nM cy5-Hel308 and 1 mM ATP/ MgCl$_2$ were injected onto the slide in the absence (D) and presence (E) of 200 pM cy3-DNA. This experiment was repeated prior to every single-molecule experiment requiring a new biotinylated quartz slide.

6.4.3.2   SMFRET ANALYSIS OF DNA BINDING

To determine the optimal protein concentration at which to observe protein: DNA binding by smFRET, a titration of cy5-Hel308 onto DNA-bound PEG-coated slides was performed (figure 6.20). Movies of 2000 frames (1 frame per 50 ms) were recorded and analysed to determine the efficiency of energy transfer between donor (cy3) and acceptor (cy5) molecules.
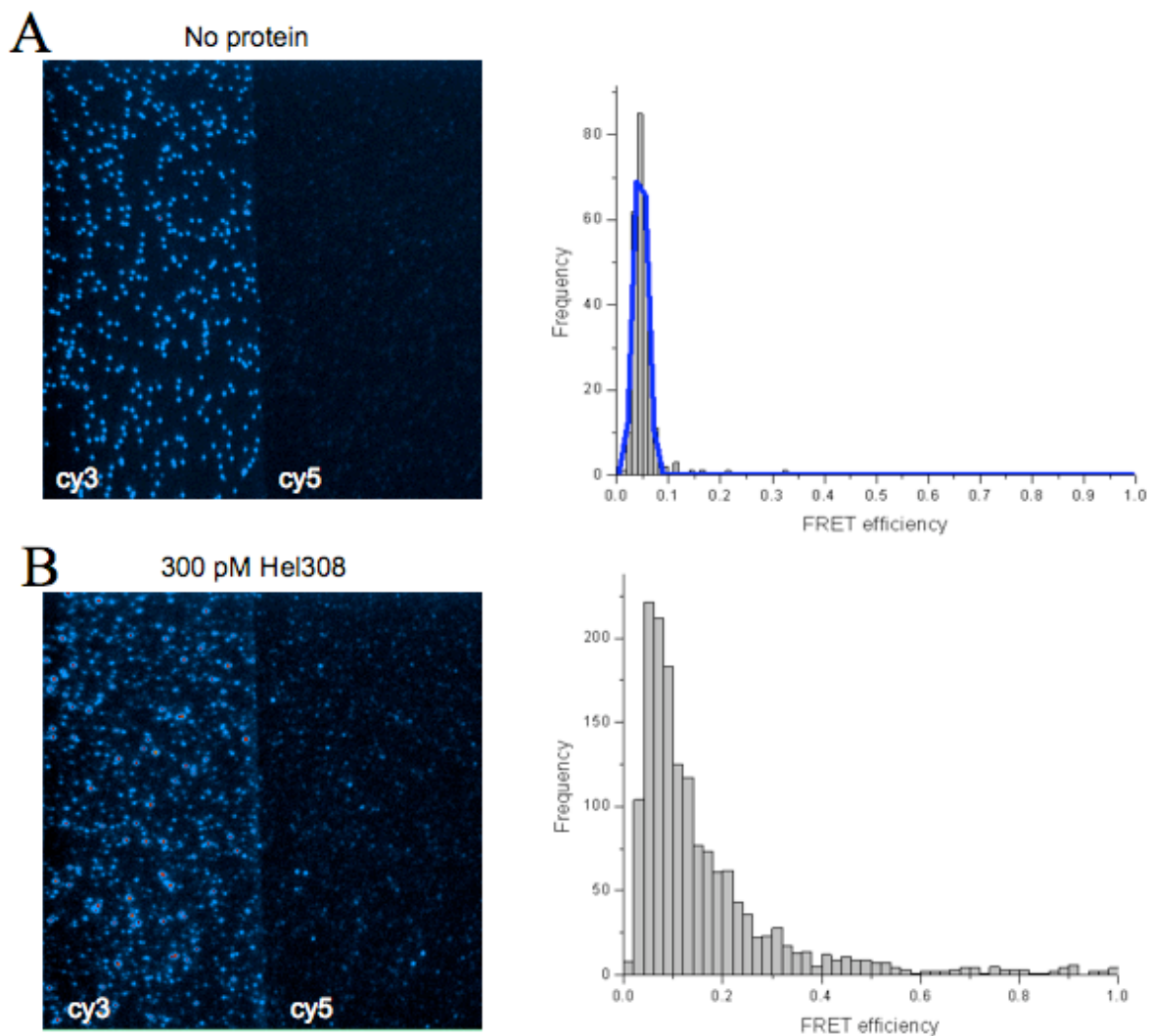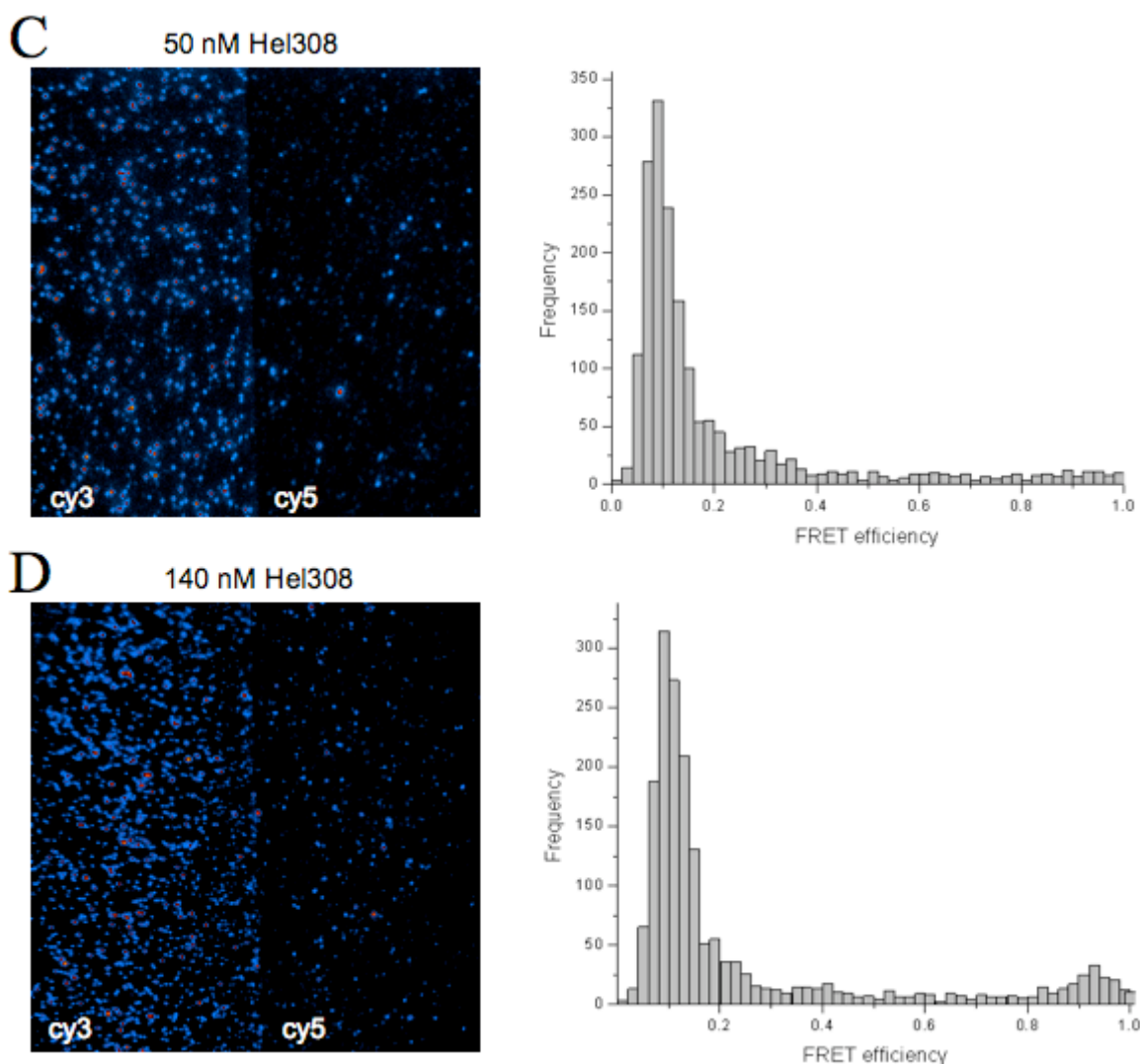
Figure 6.20: Hel308 concentration-dependence of FRET during DNA binding

Cy5-Hel308 was titrated onto a PEG-coated quartz slide with immobilised cy3-DNA (200 pM) and slides were imaged by excitation of the cy3-fluorophore at 532 nm. Left, image produced following excitation of cy3 molecules. Since cy5 molecules are not directly excited, molecules are only visualised in the cy5 channel (right hand panel) as a result of energy transfer from cy3 molecules. Right, histogram revealing the FRET efficiency between each cy3-cy5 pair (histogram produced using MATLAB software). A, no protein; B, 300 pM Hel308; C, 50 nM Hel308; D, 140 nM Hel308. This result is representative of triplicate experiments performed using three different protein preparations on three different biotinylated quartz slides.

As the concentration of Hel308 approached the dissociation constant (140 nM), FRET could be observed between cy3 and cy5 molecules readily (figure 6.20). However, at protein concentrations above 50 nM, multiple protein molecules were bound to DNA. Since this complicates analysis, subsequent smFRET DNA binding experiments were performed using 200 pM DNA and 10 nM protein, as outlined in Materials and Methods.
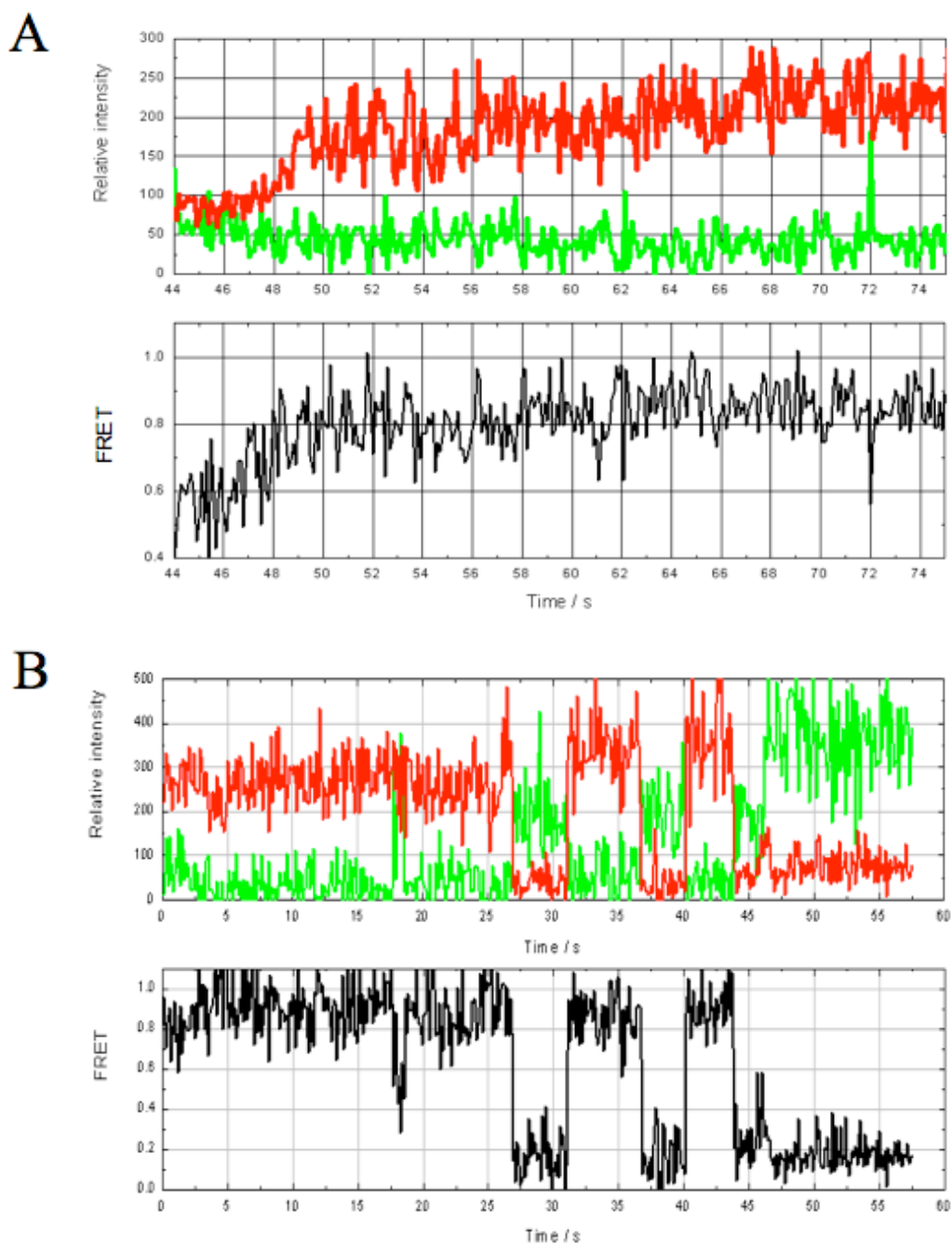
Figure 6.21: smFRET analysis of DNA binding by Hel308

Traces showing FRET between cy3-ssDNA (200 pM) and cy5-Hel308 (10 nM). For both (A) and (B): upper trace shows the relative emission intensity of cy3 (green) and cy5 (red) molecules over time; lower trace, FRET efficiency between cy3 and cy5 molecules over time. In (A), only the cy5 signal is observed indicating extended protein: DNA association; In (B), donor and acceptor emission intensities alternate over time. These traces show the behaviour of two molecules (from a field of view containing ~200 molecules) and reveal the heterogeneity of Hel308 binding activity.

Figure 6.21 displays representative traces obtained from single-molecule DNA binding experiments. For each molecule, one trace is generated to reveal the relative emission intensity of the donor and acceptor fluorophores over the course of the

experiment (figure 6.21 A/ B, upper trace). A second trace is produced to reveal the FRET efficiency between the donor and acceptor pair during the experiment (figure 6.21 A/ B, lower trace). Two predominant types of FRET traces were observed during DNA binding experiments: approximately half of the traces showed acceptor-only signals (figure 6.21 A), and the remainder showed donor and acceptor signals alternating over time (figure 6.21 B). In the former case, Hel308 may have bound to a cy3-DNA molecule for an extended period of time without dissociation or the bound DNA molecule may possess an inactive cy3 molecule. Control experiments (figure 6.19) suggest that an acceptor-only signal was unlikely to arise from non-specific adsorption of the cy5-protein molecule to the quartz slide within close proximity (within 10 nm) to an immobilised cy3-DNA molecule. It can be concluded that the majority of protein-binding events observed during these smFRET experiments were DNA-dependent. Figure 6.21 B shows repetitive binding and dissociation of Hel308 from DNA. Since these experiments were performed in the absence of ATP/ $MgCl_2$, DNA translocation could not account for the observed FRET signal. In the trace shown in figure 6.21 B, a single cy5-Hel308 molecule was bound to the DNA for ~ 28 s, when it dissociated briefly (resulting in a decrease in FRET) before re-binding the DNA molecule ~ 5 s later (indicated by an increase in FRET/ increase in acceptor fluorescence intensity). This was repeated once more before the protein molecule completely dissociated from the DNA. Alternatively, this trace may result from the dissociation of one Hel308 molecule from the DNA and subsequent binding of a different Hel308 molecule.

6.4.3.3   SMFRET ANALYSIS OF DNA TRANSLOCATION

Cy5-Hel308 was titrated onto quartz slides containing immobilised cy3-DNA, in the presence of ATP/ $MgCl_2$. After the addition of protein, slides were imaged by excitation of the cy3 dye. By contrast with DNA binding results in figure 6.20, the addition of ATP/ $MgCl_2$ significantly improved the efficiency of energy transfer between cy3 and cy5 molecules (figure 6.22), either reflecting translocation of the protein in a 3'-5' direction or an increased DNA binding affinity.
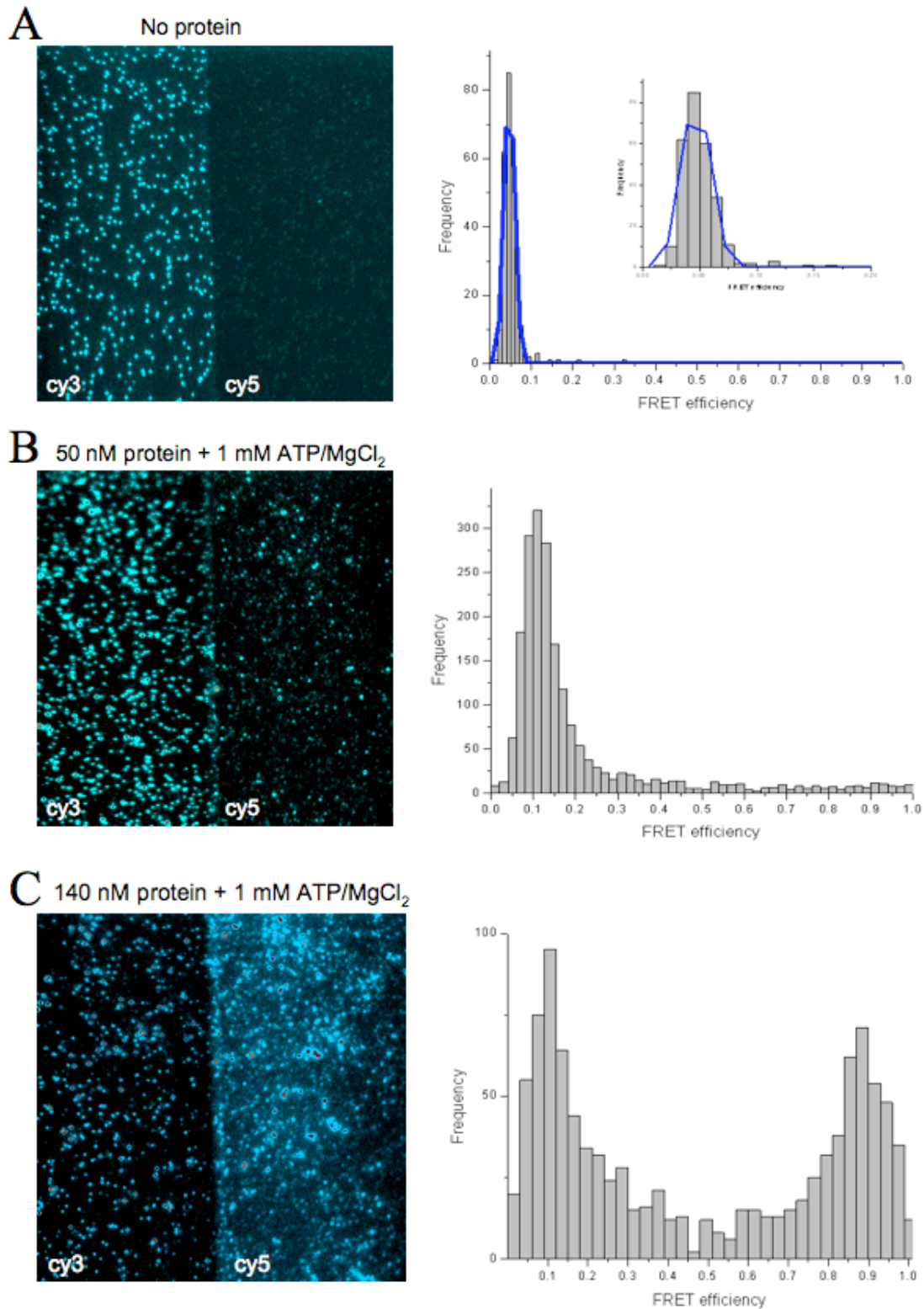
Figure 6.22: Hel308 concentration-dependence of FRET during DNA translocation

Cy5-Hel308 and ATP/ MgCl₂ were added to PEG-coated quartz slides possessing cy3-DNA. Histograms (right) reveal the FRET efficiencies between each cy3-cy5 pair. This result is representative of triplicate experiments performed using three different protein preparations on three different biotinylated quartz slides.

Although these experiments show that DNA translocation by Hel308 took place, they fail to report reaction kinetics. To resolve the limitations of this experiment, translocation studies must be performed in real-time. In other words, imaging of the protein: DNA complexes must begin prior to the addition of ATP/ MgCl$_2$. Therefore, a flow-cell chamber was designed (figure 6.23) to permit ATP/ MgCl$_2$ injection during the imaging process. The flow-cell chamber was mounted on the microscope scanning stage and held tightly in place with metal clamps.
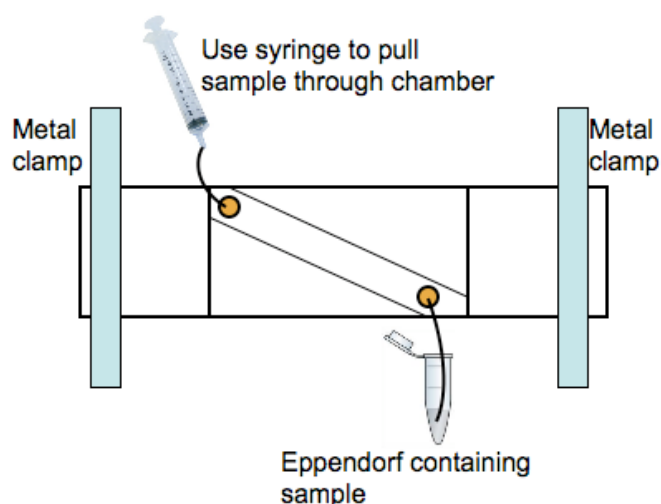
Figure 6.23: Flow-cell chamber for real-time smFRET studies of Hel308

A standard PEG-coated quartz microscope slide equipped with an input and output hole (yellow). The head of a hypodermic needle attached to a piece of plastic tubing (black curved line) was secured to each hole. One piece of tubing was attached to a 1 ml syringe and the other was inserted into an eppendorf tube containing the sample. To load the sample, the syringe was pulled forcing the sample from the eppendorf tube into the chamber. Metal clamps secured the slide to the microscope scanning stage.

To monitor DNA translocation of Hel308 in real-time, 200 pM cy3-DNA was immobilised on quartz slides, as above. Cy5-labelled Hel308 (10 nM) was bound to the DNA and complexes were imaged by excitation of the cy3 dye at 532 nm. During imaging, ATP/ MgCl$_2$ was loaded onto the slide (as described in figure 6.23) to stimulate translocation. Figure 6.24 documents the translocation of Hel308 along DNA with a series of images taken before, immediately after and 150 s after the injection of ATP/ MgCl$_2$. In the absence of ATP/ MgCl$_2$, low FRET was observed; since figure 6.21 shows that Hel308 can bind to DNA under these conditions, this data suggests that the protein was bound > 20 nt from the cy3 fluorophore (figure 6.24 A). Immediately following the addition of ATP/ MgCl$_2$, FRET values of ~ 0.5 were observed, consistent

with translocation of Hel308 along DNA (figure 6.24 B). After 120 s, the apparent efficiency of FRET approached 1 suggesting that nearly all molecules within the field of view had achieved a state of FRET resulting from Hel308 translocation (figure 6.24 C). To view the full-length movie of this experiment, please refer to the attached CD-ROM and Appendix 7.

The translocation experiments represented in figure 6.24 shows that FRET values increased upon addition of ATP/ MgCl$_2$. This suggests that the majority of Hel308 molecules in a heterogeneous population translocate along DNA with 3'-5' polarity, bringing the cy5 fluorophore into close proximity to the cy3 fluorophore at the 5' end of the DNA. However, analysis of the FRET vs time traces (figure 6.25) for each immobilised molecule imaged suggested that Hel308 molecules could translocate in a 5'-3' direction (figure 6.25 A), indicated by a gradual decrease in FRET over time, or in a 3'-5' direction (figure 6.25 B) indicated by a gradual increase in FRET over time. Furthermore, several traces revealed that individual Hel308 molecules were capable of bidirectional translocation during a single DNA binding event (figure 6.25). These results are consistent with gel-based assays reported in chapter 5.

Unfortunately, reaction kinetics for DNA translocation by Hel308 could not be reliably extracted from real-time smFRET experiments as a result of microscope defocusing upon ATP/ MgCl$_2$ addition. For this reason, further experiments must be performed to firmly establish the polarity of Hel308 translocation along DNA. Microscope defocusing also prohibited visualisation of DNA unwinding by Hel308 via smFRET since real-time data collection was also necessary for these experiments. Future experiments will, therefore, rely on the optimisation of the flow-cell chamber to permit real-time smFRET analysis of Hel308 function.
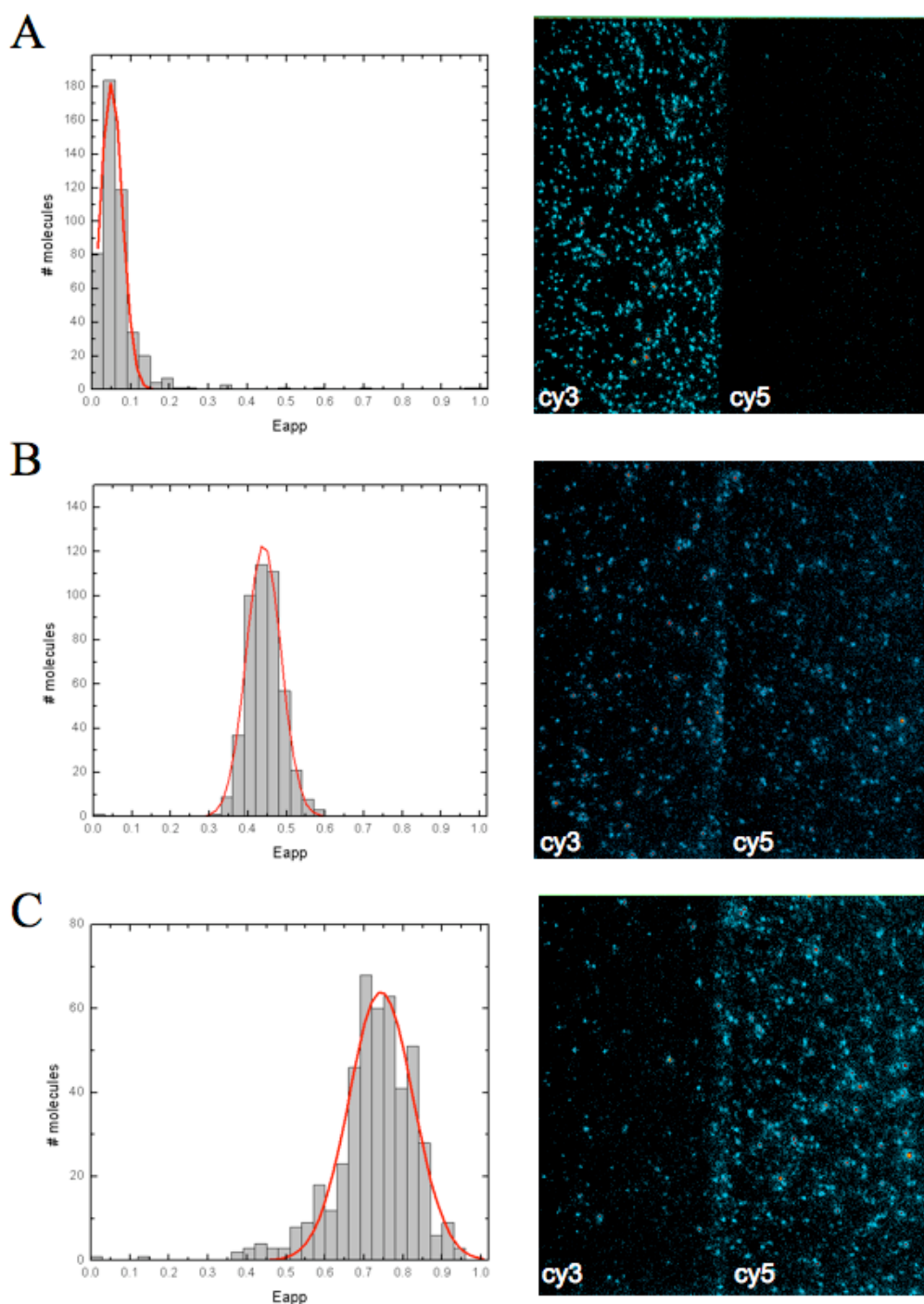
Figure 6.24: Real time smFRET analysis of DNA translocation by Hel308

smFRET images (right) and accompanying histograms (left) showing the apparent FRET (Eapp) of each cy3-cy5 pair during Hel308 translocation along DNA. Cy5-Hel308 (10 nM) was bound to cy3-ssDNA (200 pM) on PEG-coated quartz slides and a 2000-frame movie (frame rate: 1 frame per 50 ms) was recorded. Images are presented prior to the injection (A), 20 s after injection (B) and 150 s after injection (C) of ATP/ MgCl$_2$. This result is representative of triplicate experiments performed using three different protein preparations on three different biotinylated quartz slides.
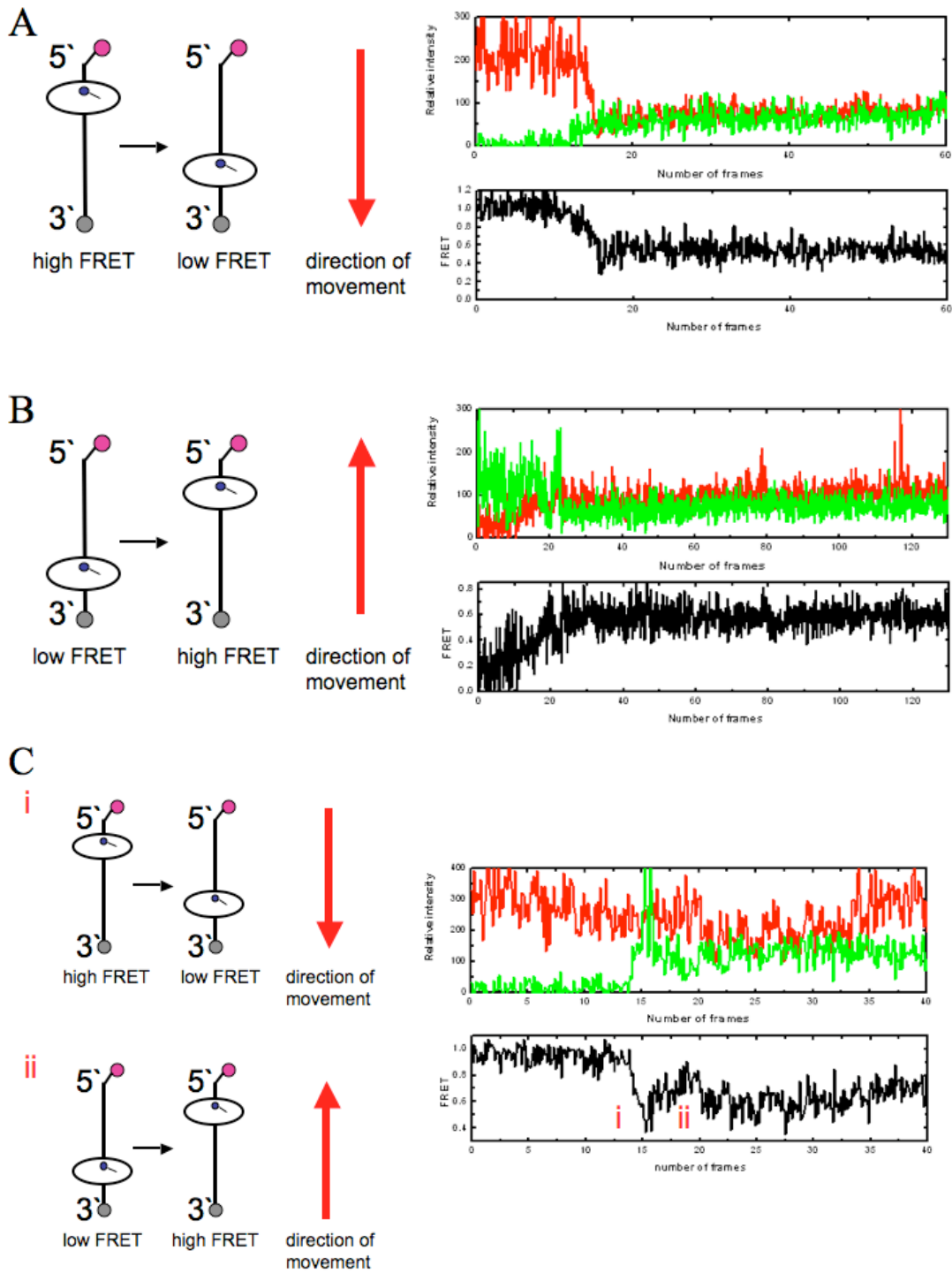
Figure 6.25: Bidirectional DNA translocation by Hel308

smFRET traces showing cy5-Hel308 translocation along cy3-ssDNA with (A) 5'-3' and (B) 3'-5' polarity. Bidirectional movement of Hel308 (i, 5'-3'; ii, 3'-5') was observed by individual molecules during a single DNA binding event, as represented in (C). Upper trace, emission intensity of cy3 (green) and cy5 (red) over time; lower trace, FRET efficiency between the cy3-cy5 pair visualised; pink circle, cy3; blue circle, cy5. The traces show the behaviour of three molecules (taken from a field of view containing ~200 molecules) and provides an insight into the heterogeneity of Hel308 activity.

6.5   DISCUSSION AND CONCLUDING REMARKS

This chapter described the use of PELDOR EPR spectroscopy and single-molecule FRET spectroscopy to investigate the mechanistic properties of Sso PBL2025 Hel308.

Structural analysis of apo-Hel308 and Hel308 in complex with DNA (Buttner et al., 2007; Richards et al., 2008b) or with nucleotide cofactors (Oyama et al., 2009) led to the proposal of a working model to describe how the protein coupled ATP binding and hydrolysis to DNA translocation and unwinding. One of the central features of the protein was that domains 1-4 formed a ring structure that could accommodate the displaced ssDNA following strand displacement. However, it was shown that the entry point of DNA, between domains 2 and 4, could not accommodate a large DNA structure, such as a stalled replication fork previously described as the preferentially substrate of Hel308 (Guy and Bolt, 2005; Richards et al., 2008b). It was hypothesised that the protein would have to undergo conformational changes to increase the gap size between domains 2 and 4 such that a branched molecule could enter (Buttner et al., 2007). Furthermore, to permit the transduction of ATP hydrolysis to DNA translocation and unwinding, linked movements of domains 1, 2 and 4 were expected to occur during the Hel308 reaction cycle (Buttner et al., 2007).

The data we obtained using PELDOR spectroscopy was not fully consistent with this model. A small but reproducible decrease in distance (1.2 Å) between domains 1 and 2 was observed upon ADP binding, indicating that the cleft between the motor domains may open/ close during ATP binding/ hydrolysis in a manner akin to that reported for PcrA (Velankar et al., 1999). However, no measurable change in the distance between domains 2 and 4 was detected upon DNA binding, suggesting that the passage through which DNA enters the protein does not alter significantly *in vivo.* These data are consistent with the structural analysis of Afu Hel308 (Buttner et al., 2007) and the Hel308 homologue from *P. furiosus*, Hjm (Oyama et al., 2009), that reveal superimposable structures in the presence and absence of DNA and nucleotide cofactors (figure 6.5). It is important to note that the positions of the MTSL spin labels may not have been optimal to observe conformational changes during the Hel308 reaction cycle by PELDOR EPR spectroscopy.

This chapter also presented a preliminary investigation of Hel308 biochemistry using single-molecule fluorescence resonance energy transfer techniques.

The use of polyethylene glycol (PEG)-coated quartz microscope slides proved sufficient to prohibit non-specific adsorption of fluorescently labelled protein molecules to the quartz slide, and permitted the observation of specific DNA-protein interactions. Despite observing association and dissociation events between DNA and protein, multiple protein binding per DNA molecule at high protein concentrations prohibited verification of the ssDNA dissociation constant of Hel308 by smFRET techniques. DNA translocation was observed by smFRET, supporting the bidirectional nature of movement observed by *in vitro* gel-based studies presented in Chapter 5. The data obtained by smFRET also revealed the distribution of such translocation events, implying that DNA translocation predominantly took place with 3'-5' polarity. Further work is required to fully understand the bidirectional nature of Hel308 translocation. Owing to technical difficulties, kinetic parameters of DNA translocation and unwinding by Hel308 could not be elucidated and remains the focus of future work.

CHAPTER 7

STRUCTURAL AND FUNCTIONAL CHARACTERISATION OF SSO2452

A RADA PARALOGUE FROM *S. SOLFATARICUS* WITH ANTI-RECOMBINASE ACTIVITY

7.1   INTRODUCTION

The RecA protein family, comprising RecA in bacteria, Rad51 in eukarya, and RadA in archaea, are among the most universally conserved DNA repair proteins known (Lin et al., 2006). These proteins are DNA recombinases that catalyse the strand exchange reaction central to homologous recombination (HR) and double-strand break repair (DSBR) (Cox, 2003) (figure 7.1). This reaction involves the initial formation of a nucleoprotein filament between the protein and single-stranded DNA (ssDNA) that subsequently invades a duplex DNA of cognate sequence to exchange genetic information. This strand exchange reaction leads to the formation of joint molecule recombination intermediates, such as heteroduplexes, D-loops and Holliday junctions (Kowalczykowski and Eggleston, 1994; Wyman et al., 2004). Disruption of RecA in bacteria, or Rad51 in yeast, is highly deleterious to the cell, whereas Rad51 in metazoa is an essential protein (Sonoda et al., 1998). This may reflect the fact that HR/ DSBR is the primary pathway for the rescue of stalled or collapsed replication forks (McGlynn and Lloyd, 2002), a phenomenon known as Recombination Dependent Replication (RDR).
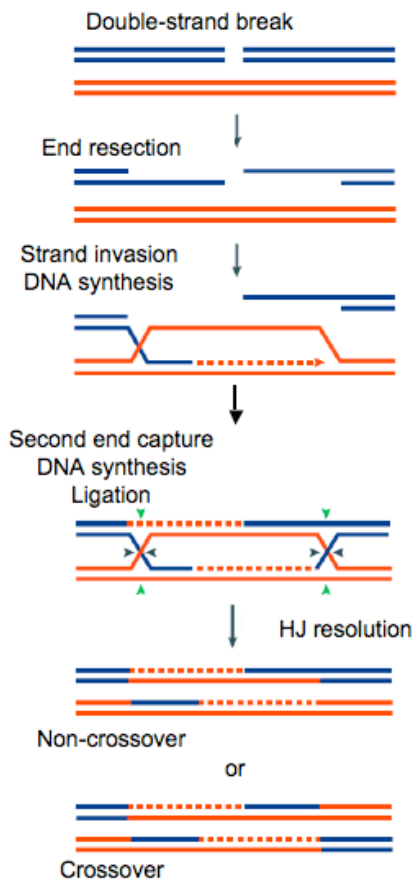
Figure 7.1: HR/ DSBR pathway

Double-strand breaks (DSB) can be repaired by various pathways of homologous recombination (HR), including double-strand break repair (DSBR). Repair is initiated following the resection of a DSB to generate 3' ssDNA tails that are subsequently bound by a RecA-like DNA recombinase to facilitate its invasion into a homologous duplex. The invading strand uses the homologous duplex as a template for DNA synthesis. This process is repeated with the second end of the DSB to form an intermediate with two Holliday junctions (HJs). Repair synthesis and ligation preceeds resolution of the HJs in either a non-crossover (black arrow heads at both HJs) or crossover mode (green arrow heads at one HJ and black arrow heads at the other HJ). Modified from (Sung and Klein, 2006).

In bacteria, a single DNA recombinase (RecA) is responsible for catalysing strand exchange. RecA is the most extensively studied DNA recombinase and research has revealed the molecular features necessary for the catalysis of strand exchange. Firstly, although recombinases have been observed in ring, filamentous, and monomeric configurations (Forget et al., 2006), strand exchange requires the cooperative assembly of RecA on ssDNA to form an elongated, helical nucleoprotein filament that can interact with dsDNA and undergo homology recognition (Sarai et al., 2006). Electron microscopy has proved a fundamental technique for direct observation of such nucleoprotein filament formations (Register et al., 1987; McIlwraith et al., 2001; Esashi et al., 2007), whilst single molecule techniques exploiting magnetic tweezers have provided greater insight into the dynamics and structural changes that occur during recombination (van der Heijden et al., 2008). Strand exchange by RecA is an ATP-dependent process: while homology recognition events, involving the formation of joint molecules, relies only on ATP binding, complete strand exchange requires ATP hydrolysis (Seitz and Kowalczykowski, 2000).

In contrast to bacteria, eukaryotes encode several Rad51 paralogues, in addition to Rad51 itself (Sung et al., 2003; Lin et al., 2006). There are seven RAD51-like genes in humans, including RAD51A, RAD51B, RAD51C, RAD51D, XRCC2, XRCC3 and the meiosis-specific DMC1 (Sung et al., 2003). Five of these paralogues exist in two complexes *in vivo*: the BCDX2 complex (RAD51B, RAD51C, RAD51D, and XRCC2) and the RAD51C-XRCC3 complex (Masson et al., 2001; Liu et al., 2002). These genes are all essential in mice, and the proteins cooperate with Rad51 in strand-exchange reactions *in vitro*, and are required for damage-specific Rad51 repair foci *in vivo* (Thacker, 2005; Liu et al., 2007). While the molecular roles of the Rad51 paralogues remains elusive, the Rad51C-XRCC3 complex has been implicated in the latter stages of the HR pathway (Liu et al., 2007). *S. cerevisiae* has only two Rad51 paralogues, Rad55/ Rad57, which forms a heterodimer and stimulates Rad51-mediated strand-exchange *in vitro* (Sung, 1997).

The archaea, although lacking a nucleus and bearing a superficial resemblance to bacteria, are most closely related to eukaryotes with respect to their informational processes, including DNA replication, recombination and repair. This is reflected in the fact that archaeal organisms possess at least one (and as many as four) RadA paralogue(s), but their functions remain unclear (Seitz et al., 1998; Waters et al., 2003). Several crenarchaeal genomes, in particular, encode multiple RadA paralogues, including three in *S. solfataricus* (Sso0777, Sso1861, and Sso2452) and four in *Pyrobaculum aerophylum*. Phylogenetic analysis reveals that these are monophyletic, with robust bootstrap values supporting the differentiation of this family of RadA paralogues from both euryarchaeal RadB and the archaeal RadA proteins (figure 7.2). The collective name "aRadC" (archaeal RadC) has recently been suggested for this group of RadA paralogues (Haldenby et al., 2009).
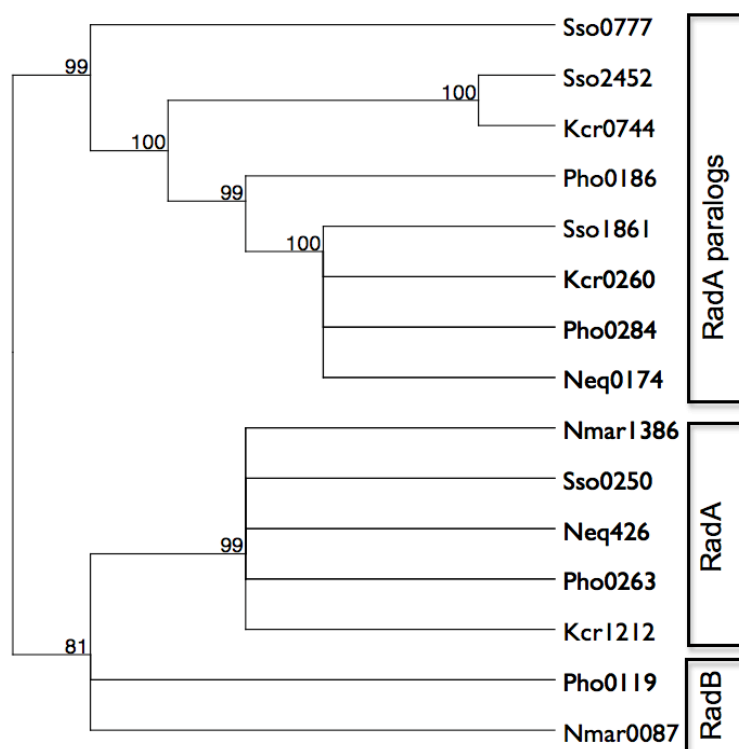
Figure 7.2: Phylogenetic analysis of archaeal RadA paralogues

This unrooted, bootstrapped phylogenetic tree shows the monophyletic family of uncharacterised archaeal RadA paralogues, with RadB from *Pyrococcus horikoshii* and RadA from *S. solfataricus* as representative outgroups. Each protein is represented by a three-letter code followed by the gene number from the respective genome sequences (Sso, *S. solfataricus*; Pho, *P. horikoshii*; Neq, *Nanoarchaeum equitans*; Nmar, *Nitrosophumilus martimus*; Kcr, *Korarchaeum cryptofilum*). This neighbour-joining tree was generated from a ClustalW alignment of the proteins using the programme MacVector, with pairwise distances between sequences uncorrected. The bootstrap values shown at each node represent the percentage of all trees (1000 total) agreeing with this topology. Reproduced from (McRobbie et al., 2009).

Structural studies of HR proteins from archaea, including Rad50, Mre11 and RadA, have supplied a great deal of information relevant to their eukaryal equivalents (Hopfner et al., 2001; Shin et al., 2004). Archaeal RadA is much more similar to Rad51 than to RecA, sharing the dsDNA binding N-terminal domain (NTD), and lacking the RecA-specific C-terminal domain (CTD) (Lin et al., 2006). A second Rad51 gene, RadB, has been described in some euryarchaea. RadB lacks the NTD present in RadA but has the core ATPase domain of the Rad51/ RecA family (Akiba et al., 2005). RadB appears to lack the strand-exchange activity of RadA, and turns over ATP very slowly (Komori et al., 2000a; Guy et al., 2006). Bioinformatic analysis supports the

differentiation of the aRadC family from the canonical archaeal recombinase RadA as, like RadB, they lack the NTD and are restricted to the core ATP binding domain.

Sso2452 is representative of the family of RadA paralogues in having only limited similarity (~ 30%) to the canonical archaeal recombinase RadA. In a previous screen for proteins from *S. solfataricus* that interact with ssDNA bound by the ssDNA binding protein (SSB), Sso2452 was identified as a prominent interacting protein (Cubeddu and White, 2005). Furthermore, *S. tokodaii* Sto0579, which exhibits 84% sequence identity to Sso2452, is reportedly upregulated in response to UV radiation and may play a role in overcoming the inhibitory effect of SSB on RadA-catalysed strand exchange (Sheng et al., 2008a). In this chapter, the initial biochemical characterisation and crystal structure of Sso2452 from *S. solfataricus* will be presented, which reveals key structural differences from the canonical RecA family recombinases that may explain its functional properties. The likely role of the archaeal RadA paralogues *in vivo* is discussed. The data presented in this chapter have been published in (McRobbie et al., 2009).

7.2 PURIFICATION OF RECOMBINANT SSO2452

The *sso2452* gene from *S. solfataricus* was amplified by PCR and cloned into the pDEST14 expression vector (performed by the Scottish Structural Proteomics Facility, University of St Andrews) for expression of N-terminal polyhistidine tagged protein. The gene was fully sequenced to confirm accurate cloning, free from base pair modifications. Recombinant protein was expressed in *E. coli* BL21 Rosetta cells and purified to homogeneity by heat-treatment followed by immobilised metal-affinity and size exclusion chromatography (figure 7.3), as described in Materials and Methods. The N-terminal poly-histidine tag was cleaved from the protein by overnight incubation with the tobacco etch virus protease at 22 ºC and further purified by immobilised metal-affinity chromatography to separate tagged from untagged protein. Typical yields of Sso2452 were 5-10 mg/ L of culture.
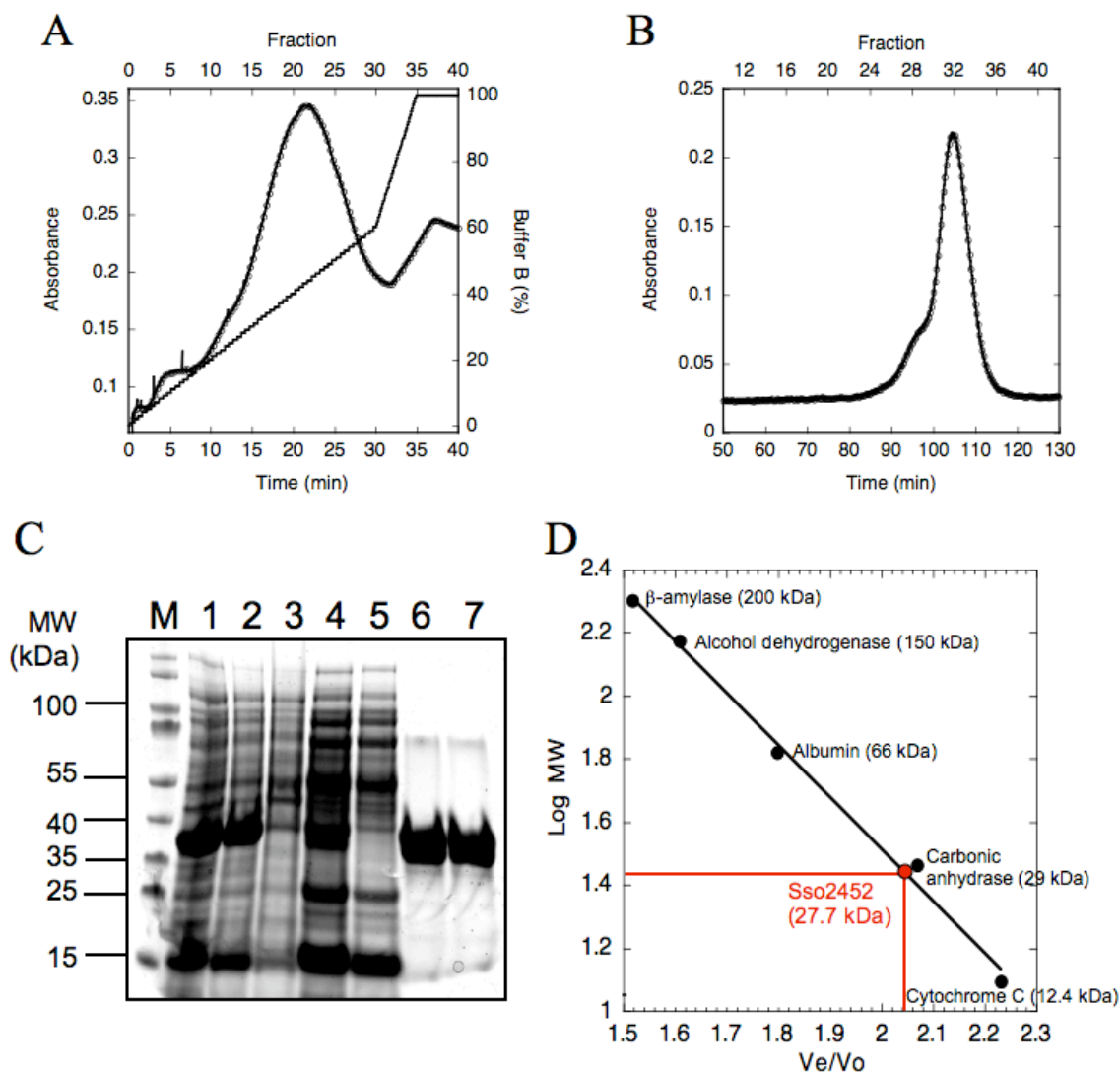
Figure 7.3: Purification of Sso2452

Elution chromatograms from (A) immobilised metal-affinity and (B) size exclusion chromatography. Absorbance recorded at 280 nm. (C) SDS-PAGE showing Sso2452 purification steps. From left to right: *E. coli* cell lysate before (1) and after (2) heat-treatment at 65 °C, insoluble material after heat-treatment (3), pooled fractions containing Sso2452 (4) and flow-through (5) after heparin column, Sso2452 after gel-filtration column before (6) and after (7) removal of polyhistidine tag. (D) Analytical gel filtration standard curve revealing the molecular mass of Sso2452 based on an elution volume of 63.7 ml.

Purified Sso2452 was analysed by SDS-PAGE to reveal an approximate molecular weight of 35 kDa (figure 7.3 C). The oligomeric state of the protein in solution was determined by analytical gel filtration. Proteins of known molecular weight were loaded onto a Superdex 200 10-300 column and a standard curve was generated (Appendix 5): the logarithm of the protein molecular weight was plotted against the ratio of its elution volume (Ve) divided by the void volume (Vo; elution

volume of Blue dextran). Sso2452 was loaded onto the column and eluted in a volume of 63.7 ml; using the equation of the standard curve, this corresponded to a molecular weight of 27.7 kDa, which was in good agreement with the theoretical mass of 30416.2 daltons and suggests that Sso2452 is a monomer in solution. To determine the mass of purified Sso2452, electrospray ionisation time-of-flight (ESI-TOF) whole protein mass spectrometry was performed. This technique revealed a mass of 30413.7 daltons (figure 7.4).
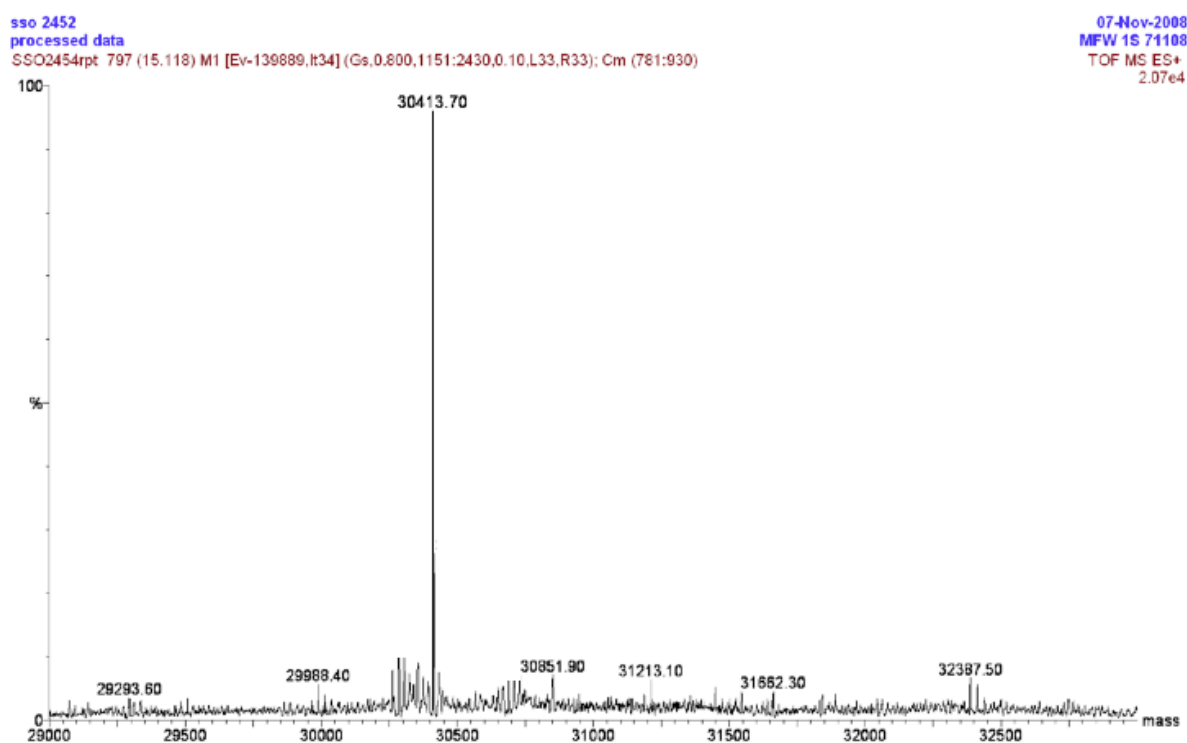
Figure 7.4: Whole mass determination of Sso2452 by ESI-TOF mass spectrometry

The molecular weight of Sso2452 was determined to confirm an intact full-length protein free from amino acid mutations. The major peak was calculated to a molecular mass of 30413.7 Da, which was in good agreement with the theoretical mass of 30416.2 Da.

7.3 EXPRESSION OF *S. SOLFATARICUS* RADA AND PARALOGUES

Although the *S. solfataricus* genome encodes three RadA paralogues (Sso0777, Sso1861, and Sso2452), there is little information available on their function or even whether they are all expressed *in vivo*. Using quantitative one-step reverse

transcriptase-PCR (RT-PCR), we quantified the levels of mRNA transcripts encoding RadA and its paralogues in exponentially-growing *S. solfataricus* cells.

7.3.1 RT-PCR EXPERIMENT AND DATA ANALYSIS

Forward and reverse primers were designed, using the web-based algorithm "Primer 3" (http://frodo.wi.mit.edu/), to enable amplification of ~200 bp of the gene of interest. To analyse the RT-PCR data, a standard curve was produced to provide a measure of reaction efficiency using each primer pair. A standard dilution series of *S. solfataricus* genomic DNA was amplified using gene-specific primers and resulting crossover (Ct) value. The Ct value corresponds to the number of PCR cycles required for the transcript concentration to exceed a pre-defined threshold assigned by the RT-PCR instrument (see Materials and Methods). Ct values were plotted against the logarithm of DNA concentration. The PCR efficiency (E) with each primer set was determined using the slope of the curve (figure 7.5), according to the equation $E = 10^{[-1/slope]}$. The efficiency of PCR using the reference gene, *sso0961*, was also tested. The expression level of this gene had been previously shown to remain constant over time and following treatment with various DNA damaging agents (Salerno et al., 2003; Gotz et al., 2007). A reference gene is required to allow normalisation of the amount or quality of RNA in the samples (Pfaffl, 2001; Radonic et al., 2004).

S. solfataricus cells were grown at 80 ºC until an OD_{600} of 0.2 was reached. RNA was extracted and 100 ng was used in each RT-PCR reaction. Negative controls were prepared without RNA to rule out the possibility of contamination. RT-PCR analysis of RadA and its paralogues from *S. solfataricus* generated amplification curves to reveal Ct values (figure 7.6 A) from which relative expression levels were deduced (table 7.1). Since SYBR® Green I fluoresces when it binds any dsDNA species, the reaction is very sensitive to DNA contamination. Melting curve analysis was employed to detect the presence of non-specific DNA contaminants (figure 7.6 B). Briefly, products of the RT-PCR reactions were incubated at 95 ºC to melt the DNA and dissociate any bound SYBR® Green I. Thus as melting proceeds, a decrease in fluorescent intensity is observed. Since the temperature at which a DNA species melts is sequence-dependent, a reaction devoid of contaminants should produce a single, symmetrical melting curve peak.
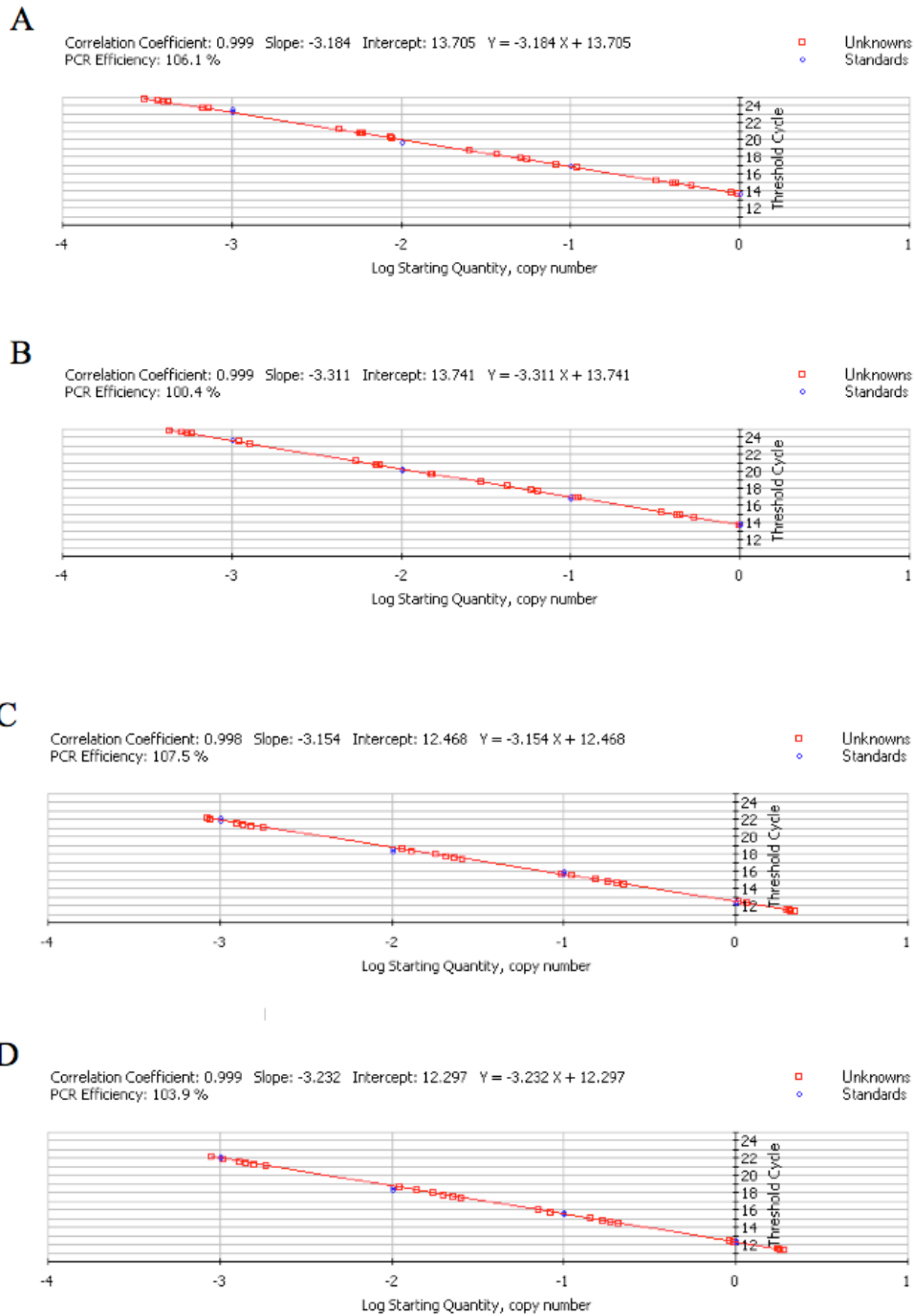
Figure 7.5: Amplification efficiency of RT-PCR primers

PCR reactions with the (A) *radA*, (B), *sso0777*, (C) *sso1861* and (D) *sso2452* primer pairs were prepared with a standard dilution series of genomic DNA. Ct values (Y-axis) of each reaction were plotted against the logarithm of the initial template concentration to reveal primer efficiencies. A straight line results if reaction efficiencies approximate 100%, and indicates that transcript levels double for each Ct unit increase.
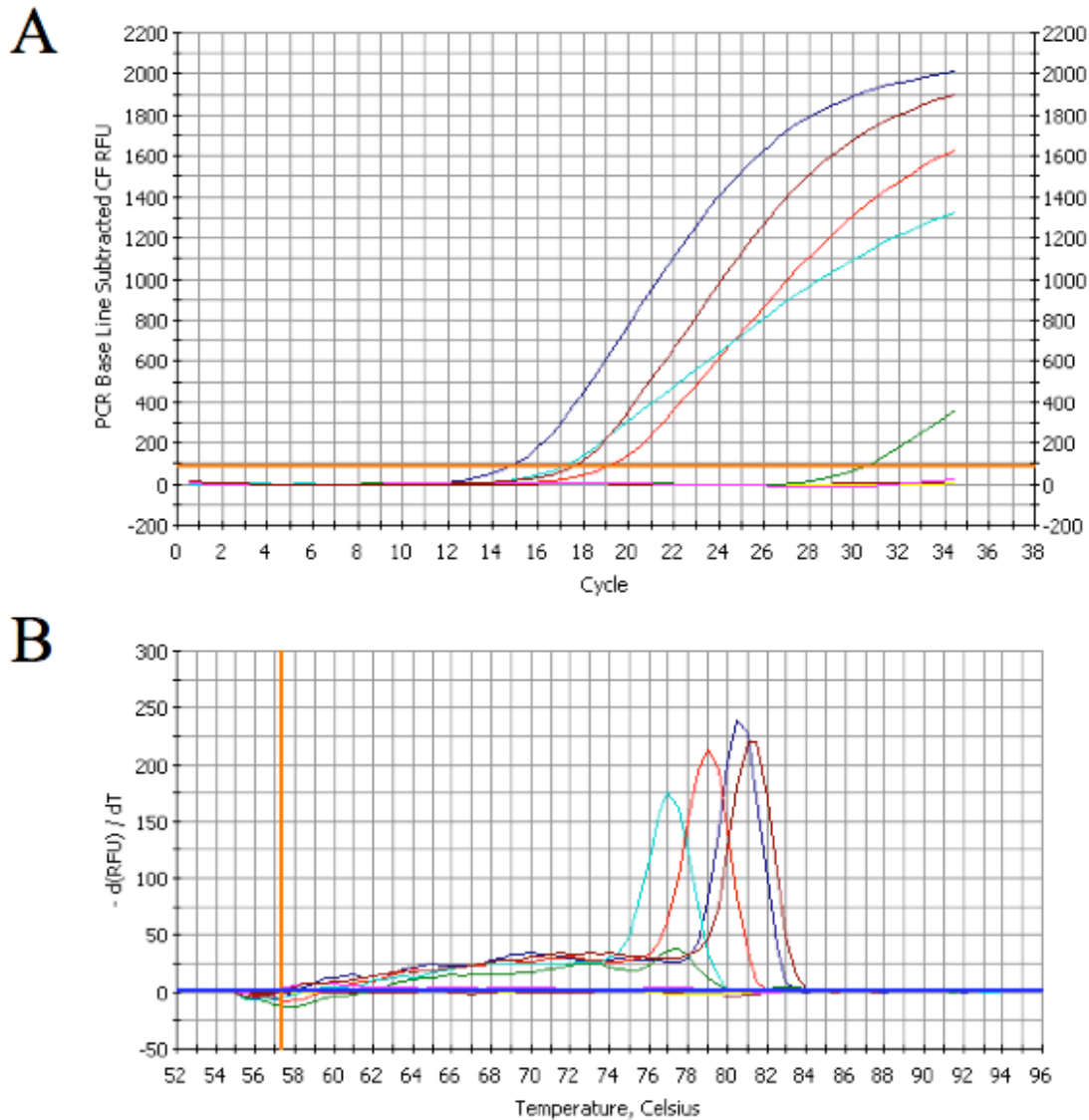
Figure 7.6: Relative mRNA levels of *radA* and paralogues from *S. solfataricus*

(A) The RT-PCR amplification curve reveals the number of PCR cycles required for *radA, sso0777, sso1861* and *sso2452* transcript concentration to cross a threshold (depicted by horizontal orange line). mRNA was extracted from growing *S. solfataricus* cultures and 100 ng was used in each reaction. Navy, *radA*; red*, sso0777*; cyan, *sso2452*; brown, *sso1861*. Negative controls lacking RNA are included: orange (*radA*), yellow (*sso0777),* green (*sso2452*) and pink (*sso1861*). (B) A RT-PCR melt curve was produced to identify any contaminating dsDNA species. The products of PCR were heated to 95 ºC and the temperatures at which they melted were recorded. Traces coloured as (A). The RT-PCR experiment was performed three times using different RNA preparations.

The data, summarised in Table 7.1, revealed that *radA* was the most abundant transcript in the *S. solfataricus* genome. The most highly transcribed paralogue was *sso1861*, with transcript levels only 5-fold lower than those for *radA*, suggesting that it may be present in appreciable quantities in the cell. *Sso2452* transcript

levels were also significant, at 11% of *radA* levels. This is consistent with previous observations that Sso2452 could be affinity purified from *S. solfataricus* cell extracts using a biotinylated oligonucleotide (Cubeddu and White, 2005). Transcripts encoding Sso0777 were present at the lowest levels (7% of *radA* levels). Although transcript abundance does not necessarily correspond directly to protein concentration in the cell, there is often a correlation. These data suggest that all three RadA paralogues are transcribed under normal growth conditions.

Table 7.1: Relative abundance of mRNA transcripts encoding RadA and paralogues in *S. solfataricus*.

Gene	Ct value (mean ± s.e.)	mRNA level relative to *radA* (%)
radA	13.7 ± 0.1	100
sso0777	17.5 ± 0.3	7
sso1861	16.3 ± 0.2	17
sso2452	16.8 ± 0.2	11
sso0961	22.3 ± 0.1	N/A

The relative levels of mRNA encoding RadA and paralogues in *S. solfataricus* were analysed by quantitative one-step RT-PCR. A total of 100 ng RNA was used in each reaction and measurements were performed in triplicate (using different RNA preparations) with standard errors (s.e.) shown. The Ct value is the cycle at which levels of PCR product cross a pre-defined threshold, with low values indicating a more abundant transcript.

7.4 RADA AND PARALOGUES ARE NOT UV-INDUCIBLE

Global microarray analysis and Western blotting have shown that *radA* expression is not induced appreciably by UV irradiation (Frols et al., 2007; Gotz et al., 2007). However, a two-fold increase in *S. tokadaii* RadA (Sto0579) protein levels in response to 200 J/ m^2 UV was recently reported (Sheng et al., 2008a). Both RadA and Sso0777 have been reported as up-regulated in response to DNA damage by actinomycin D in *S. solfataricus* (Abella et al., 2007). Transcript abundance in response to sub-lethal doses (200 J/ m^2) of UV radiation was measured for *radA* and its paralogues in *S. solfataricus* by quantitative RT-PCR. Relative gene expression of each transcript in the absence and presence of UV was calculated according to Equation 6 (Pfaffl, 2001). Briefly, relative gene expression of a sample gene is expressed as a ratio of the target gene expression in a sample versus control in comparison to a reference

gene. *Sso0961* was chosen as the reference gene since previous studies have shown that its expression remains constant over time and in the presence of UV radiation (Salerno et al., 2003; Gotz et al., 2007).

Equation 6:

$$\text{ratio} = \frac{(E_{target})^{\Delta Ct_{target}(control-sample)}}{(E_{ref})^{\Delta Ct_{ref}(control-sample)}}$$

E_{target}, RT-PCR efficiency of target gene transcript; E_{ref}, RT-PCR efficiency of a reference gene transcript; Ct, as described in text; ΔCt_{target}, Ct difference between control and sample (control – sample) of target gene transcript; ΔCt_{ref} = Ct difference between control and sample of reference gene transcript (Pfaffl, 2001).

Table 7.2 reveals that expression of all four proteins remained relatively stable up to 120 min after UV treatment, in agreement with microarray data for *S. solfataricus* (Gotz et al., 2007). It is possible that protein levels are influenced at a post-translational stage, explaining the modest induction in RadA and Sto0579 observed by Sheng and co-workers (Sheng et al., 2008a). Our findings suggest that UV radiation and actinomycin D do not induce similar patterns of transcription in *S. solfataricus* (Abella et al., 2007) and therefore may not elicit DNA repair via the same process.

Table 7.2: Effect of UV radiation on the relative expression of *radA, sso0777, sso1861,* and *sso2452* from *S. solfataricus*

Gene	Primer efficiency	Relative gene expression (UV/ control) after UV treatment			
		30 min	60 min	90 min	120 min
radA	2.06	1.00	1.04	1.01	1.00
sso0777	2.10	1.00	0.99	1.00	1.00
sso1861	2.04	1.00	0.94	0.96	0.94
sso2452	2.08	0.97	0.97	0.97	0.93
sso0961	2.00	1	1	1	1

A total of 100 ng RNA was used in each reaction and measurements were performed in triplicate (using different RNA preparations). Gene-specific RT-PCR primer efficiencies were calculated as described in the text, according to equation $E = 10^{[-1/slope]}$. The expression ratio refers to the relative levels of mRNA transcript present in cultures grown in the absence of UV treatment or following exposure to 200 J/ m^2 UV radiation, with respect to the reference gene (*sso0961*). A value of 1 represents no change in transcript level. Ratios were calculated using Equation 6 above, according to (Pfaffl, 2001).

7.5   SSO2452 BINDS SSDNA MORE TIGHTLY THAN RADA

The affinity of Sso2452 for ssDNA and dsDNA was determined by fluorescence anisotropy. Briefly, protein was directly titrated (between $0 - 10$ µM) into a solution containing 20 nM fluorescein-labelled (Fl-) single- or double-stranded DNA (45 nucleotides or base pairs, respectively) and changes in fluorescence anisotropy and intensity were recorded. Apparent equilibrium dissociation constants (K_Ds) were calculated from binding curves plotted using fluorescence anisotropy against protein concentration. Sso2452 bound ssDNA and dsDNA with apparent K_Ds of 120 nM ± 21 nM and 2.4 µM ± 0.22 µM, respectively (figure 7.7 A). To validate these binding constants, a competition assay was performed whereby unlabelled ssDNA (of the same sequence as the labelled strand) was titrated into a solution containing the Sso2452-Fl ssDNA complex (10 µM protein and 20 nM DNA). The binding curve, shown in figure 7.7 B, revealed a K_D for Sso2452-ssDNA binding of 82 nM ± 10 nM, which is in good agreement with the K_D obtained by direct titration. This suggests that the observed changes in fluorescence anisotropy reflected binding of Sso2452 to DNA and not fluorescein. This is consistent with the observed decrease of only 10% in total fluorescence intensity during both direct and competitive titrations (Appendix 3); previous research has shown that fluorescence intensity changes less than 30 % do not significantly influence anisotropy values (Reid et al., 2001).

RadA bound ssDNA and dsDNA with K_D values of 2.6 µM ± 0.3 µM and 5.3 µM ± 1 µM, respectively (figure 7.7 B). This suggests that Sso2452 binds much more tightly to DNA than RadA. These data, together with previously published work revealing that Sso2452 could be affinity purified from cell extracts of *S. solfataricus* using a biotinylated ssDNA oligonucleotide (Cubeddu and White, 2005), emphasises the potential for these RadA paralogues to form nucleoprotein filaments during HR/ DSBR.
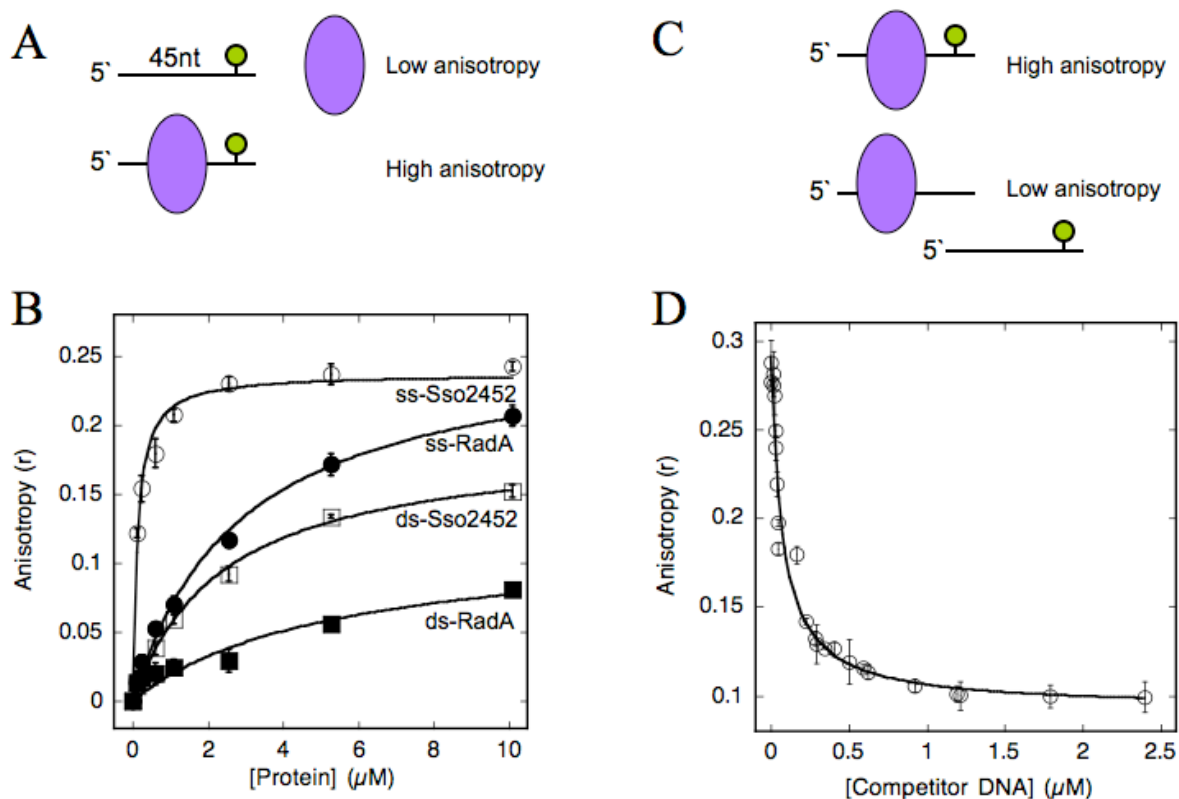
Figure 7.7: DNA binding affinity of Sso2452

(A) Protein (purple oval) was titrated into a solution containing 20 nM fluorescein-labelled DNA of 45 nucleotides or base pairs (green circle, fluorescein). (B) Anisotropy was measured and plotted against protein concentration. (C) Unlabelled, competitor DNA (same sequence as the labelled probe) was titrated into a solution containing Sso2452 (10 μM) bound to 20 nM fluorescein-labelled DNA. (D) Anisotropy was measured and plotted against the concentration of competitor DNA. Apparent dissociation constants were calculated from plots. The means of triplicate measurements were plotted with standard errors shown. For plots: Open shapes, Sso2452; closed shapes, RadA; circles, ssDNA; squares, dsDNA.

Analytical gel filtration revealed that Sso2452 was monomeric in solution (figure 7.3). However, upon incubation with a 34-nucleotide ssDNA (80 μM; Appendix 1, table A1.8), Sso2452 (149 μM) formed stable, large molecular weight (250-600 kDa) species (figure 7.8). These results were consistent with DNA-mediated assembly of the protein into a nucleoprotein complex as observed for other RecA family members including the ortholog from *S. tokodaii* (Sheng et al., 2008b).
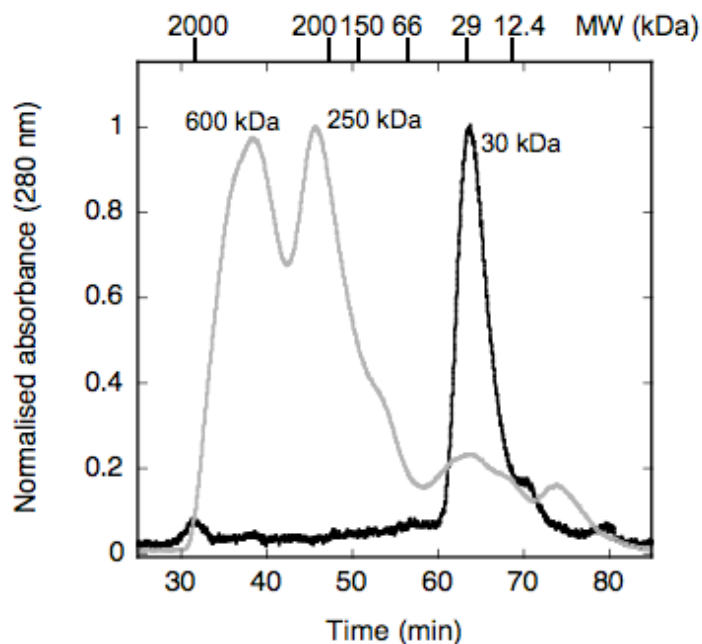
Figure 7.8: Analytical gel filtration of Sso2452: DNA complexes

A Superdex 200 10-300 gel filtration column was equilibrated with buffer (50 mM Tris, pH 7.2, 0.15 M NaCl, 1 mM EDTA, and 1 mM DTT) and calibrated with proteins of known molecular weight, whose elution positions are indicated at the top of the graph. Sso2452 (149 µM) was passed through the column alone (black trace) and in complex with a 34mer oligonucleotide (80 µM) (grey line). The maximum absorbance (280 nm) was normalised to 1. Original absorbance values at 280 nm: 600 kDa peak, 0.31; 250 kDa peak, 0.32; 30 kDa peak, 0.54. The presence of Sso2452 in eluted fractions was confirmed by SDS-PAGE. This experiment was performed once.

7.6   SSO2452 IS A DNA-DEPENDENT ATPASE

A characteristic feature of DNA recombinase proteins is their ability to catalyse DNA-dependent ATP hydrolysis. Therefore, the Sso2452 protein was tested for ATP hydrolysis activity in the presence and absence of single- and double-stranded DNA (10 nM). For temperature-dependent reactions, 1 µM Sso2452 was incubated at temperatures between 45 and 100 ºC for 60 min; for concentration-dependent reactions, Sso2452 (200 nM – 2 µM) was incubated at 60 ºC for 60 min; for time-dependent reactions, 1 µM protein was incubated with DNA at 60 ºC between 0 and 60 min.

A

B

C
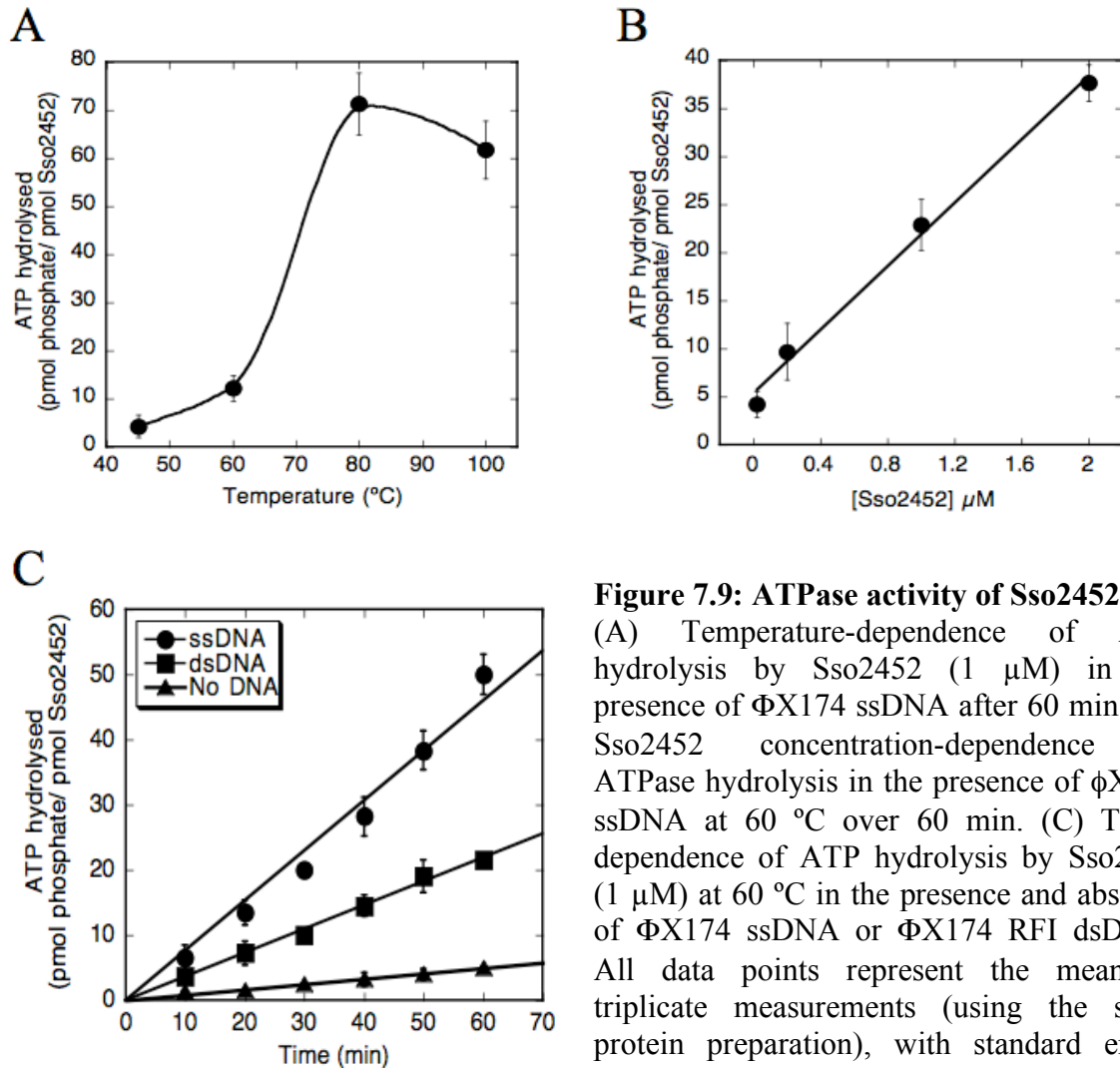
Figure 7.9: ATPase activity of Sso2452
(A) Temperature-dependence of ATP hydrolysis by Sso2452 (1 µM) in the presence of ΦX174 ssDNA after 60 min. (B) Sso2452 concentration-dependence of ATPase hydrolysis in the presence of φX174 ssDNA at 60 ºC over 60 min. (C) Time-dependence of ATP hydrolysis by Sso2452 (1 µM) at 60 ºC in the presence and absence of ΦX174 ssDNA or ΦX174 RFI dsDNA. All data points represent the mean of triplicate measurements (using the same protein preparation), with standard errors shown.

Sso2452 hydrolysed ATP to ADP in a protein-, temperature- and time-dependent manner (figure 7.9). The rate of Sso2452-mediated ATP hydrolysis was calculated to be approximately 0.8 ATP/ min in the presence of ssDNA at 60 ºC. This was higher than the rate of ATP hydrolysis observed for *S. solfataricus* RadA (0.1-0.2 ATP/ min) in the presence of ssDNA, monitored at 65 ºC (Seitz et al., 1998). ATP hydrolysis by Sso2452 was much less efficient in the presence of dsDNA and in the absence of DNA. Sso2452 ATPase activity was optimal at temperatures between 70 ºC and 80 ºC, consistent with the optimal growth temperature of the organism. ATP hydrolysis is not an inherent mechanistic requirement for the strand exchange process catalysed by RecA family members, but is thought to provide directionality and stability and allow more extensive strand exchange reactions to proceed (Jain et al., 1994). In contrast, RadB from *P. furiosus* does not catalyse multiple turnover of ATP and is not DNA stimulated (Guy et al., 2006).

233

7.7 SSO2452 PROMOTES HETERODUPLEX FORMATION

The defining event of HR/ DSBR is the exchange of homologous DNA strands to create a heteroduplex, or D-loop, species (figure 7.1). Bacterial RecA, eukaryotic Rad51, and archaeal RadA have all been reported to possess strand exchange activity (Cox and Lehman, 1982; Seitz et al., 1998; Seitz and Kowalczykowski, 2006). The ability of Sso2452 to catalyse heteroduplex formation was measured using a 5' [^{32}P]-end labelled 50-nt DNA strand and an unlabelled duplex of 25 base pairs (figure 7.10 A), modified from the protocol described by (Bugreev and Mazin, 2004). Assays were carried out at 60 ºC over a 10-min time course and products were analysed by native acrylamide gel electrophoresis. Sso2452 efficiently promoted the invasion of the single-stranded 50-mer into the duplex to form a heteroduplex product, composed of a 25-base pair region and a 25nt single-strand tail, at a rate comparable to RadA (figure 7.10 B). This activity was also observed when the two proteins were added consecutively, in either order, without any significant change in rate (figure 7.10 C). Neither reaction was completely ATP-dependent, an observation that is consistent with the behaviour of RecA family proteins when catalysing strand exchange reactions between relatively short DNA sequences (Menetski et al., 1990; Jain et al., 1994).

The stringency of the assay was tested using the *S. solfataricus* Alba1 protein, which binds both ss- and ds-DNA (Jelinska et al., 2005). The protein was found to efficiently stimulate ATP-independent strand exchange (figure 7.10 D). In contrast, the single-stranded DNA binding protein SSB from *S.* solfataricus did not support strand exchange (figure 7.10 D). These data suggest that proteins such as Sso2452, RadA and Alba1 that can bind to both ssDNA and dsDNA can promote limited strand exchange, potentially by passive equilibrium binding of all the species present.
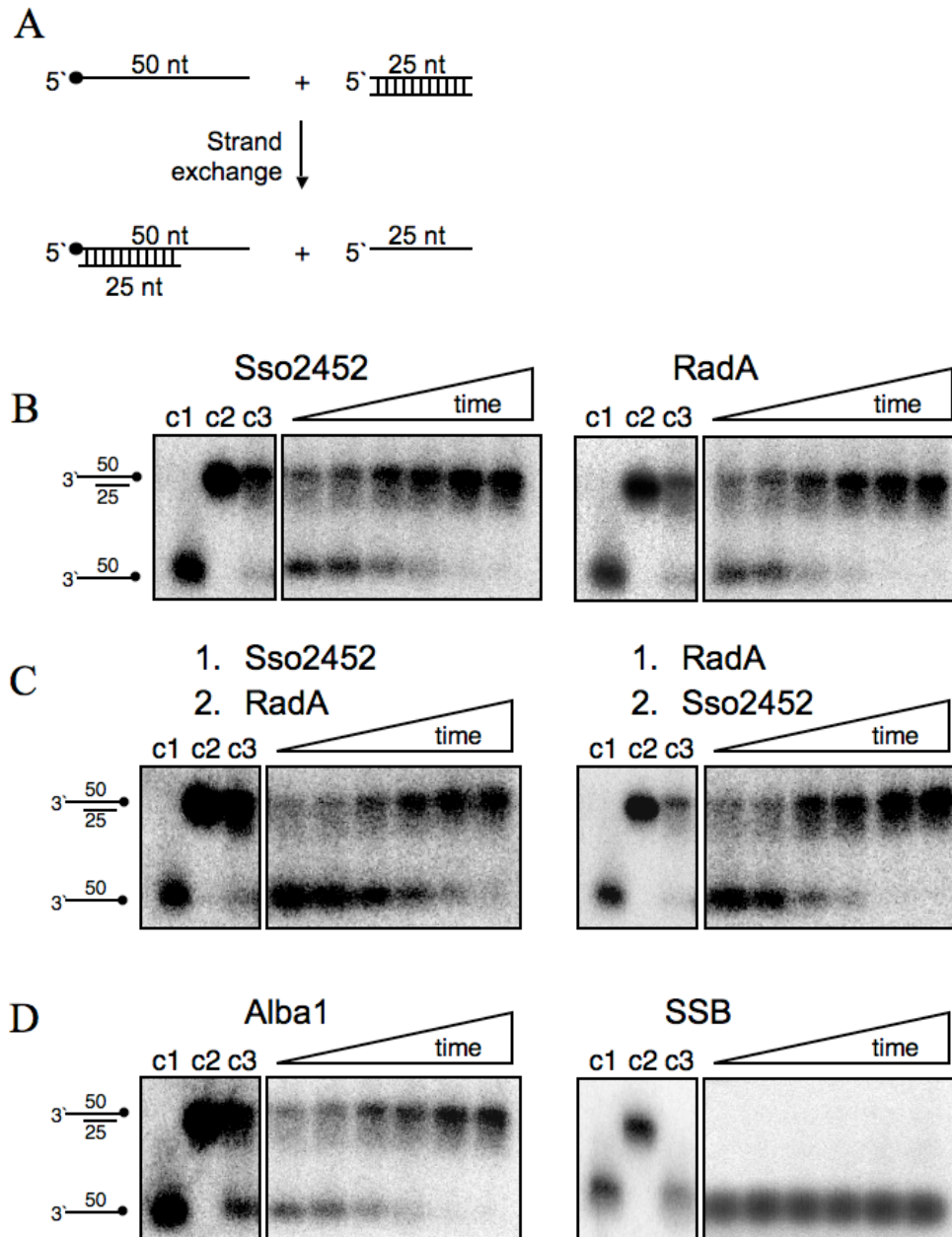
Figure 7.10: Strand-exchange activity of Sso2452 and RadA

(A) Schematic of strand-exchange reaction with length (nucleotides) of DNA indicated. Black circles, [^{32}P]-radiolabel. (B) Sso2452 (4 μM) or RadA (4 μM) was pre-incubated with [^{32}P]-end labelled 50mer for 3 min prior to addition of dsDNA (25mer) and incubation for up to 10 min at 60 °C. (C) Order of addition experiments. Protein (1) was pre-incubated with ssDNA for 3 min prior to the addition of protein (2). Following a further 3-min incubation, dsDNA was added to initiate the reaction. (D) Strand exchange reaction with Alba1 (10 μM) and SSB (10 μM). Reaction performed as described above. For all panels, time points shown are: 0.5, 1, 3, 5, 8, 10 min. c1, DNA species in the absence of protein after 10 min; c2, size marker for strand exchange product; c3, reactions lacking ATP/ MgCl$_2$ after 10 min. The results are representative of triplicate experiments (performed using different preparations of Sso2452 and RadA).

7.8 SSO2452 DOES NOT CATALYSE D-LOOP FORMATION

The strand-exchange assay above lacked sufficient stringency to test the ability of Sso2452 to function as a *bona fide* DNA recombinase. A more stringent test of recombinase activity is an assay that tests for D-loop formation, which monitors the ability of ssDNA to invade a duplex DNA plasmid (figure 7.11 A). Briefly, the protein was pre-incubated at 60 ºC with a 5' $[^{32}P]$-end labelled 80mer ssDNA and ATP/ $MgCl_2$, before reactions were initiated with double-stranded, supercoiled plasmid pUC19. Samples were taken at regular time intervals and deproteinised prior to agarose gel electrophoresis and phosphorimaging.

E. coli RecA and SsoRadA displayed robust ATP-dependent D-loop formation at 37 ºC and 60 ºC, respectively (figure 7.11 B). However, Sso2452 was unable to catalyse the formation of D-loops. Furthermore, Sso2452 inhibited the recombination activity of RadA, regardless of the order of addition, in a concentration-dependent manner (figure 7.11 C). This suggests that Sso2452 may displace RadA from RadA: ssDNA nucleoprotein filaments, either due simply to the higher ssDNA binding affinity of Sso2452 or due to an active nucleoprotein disassembly process. To mimic the *in vivo* environment, saturating concentrations of *S. solfataricus* SSB were incubated with ssDNA prior to the addition of RadA. SSB inhibited D-loop formation by RadA, consistent with studies in bacteria and eukarya (Morrical and Cox, 1990; Sung, 1997; Sigurdsson et al., 2001). To test whether Sso2452 could displace SSB to promote RadA nucleation along DNA, SSB-coated DNA was pre-incubated with RadA followed by Sso2452 (molar ratio of 1 SSB: 5 RadA: 5 Sso2452). However, no D-loops were formed (figure 7.1 D), suggesting that Sso2452 cannot function as a 'mediator protein' by displacing SSB and, consequently, promoting RadA-mediated strand exchange. These data are in contrast to those of Sheng and co-workers, who suggested such a mediator function for Sto0579, the orthologous protein from *S. tokodaii* (Sheng et al., 2008b).
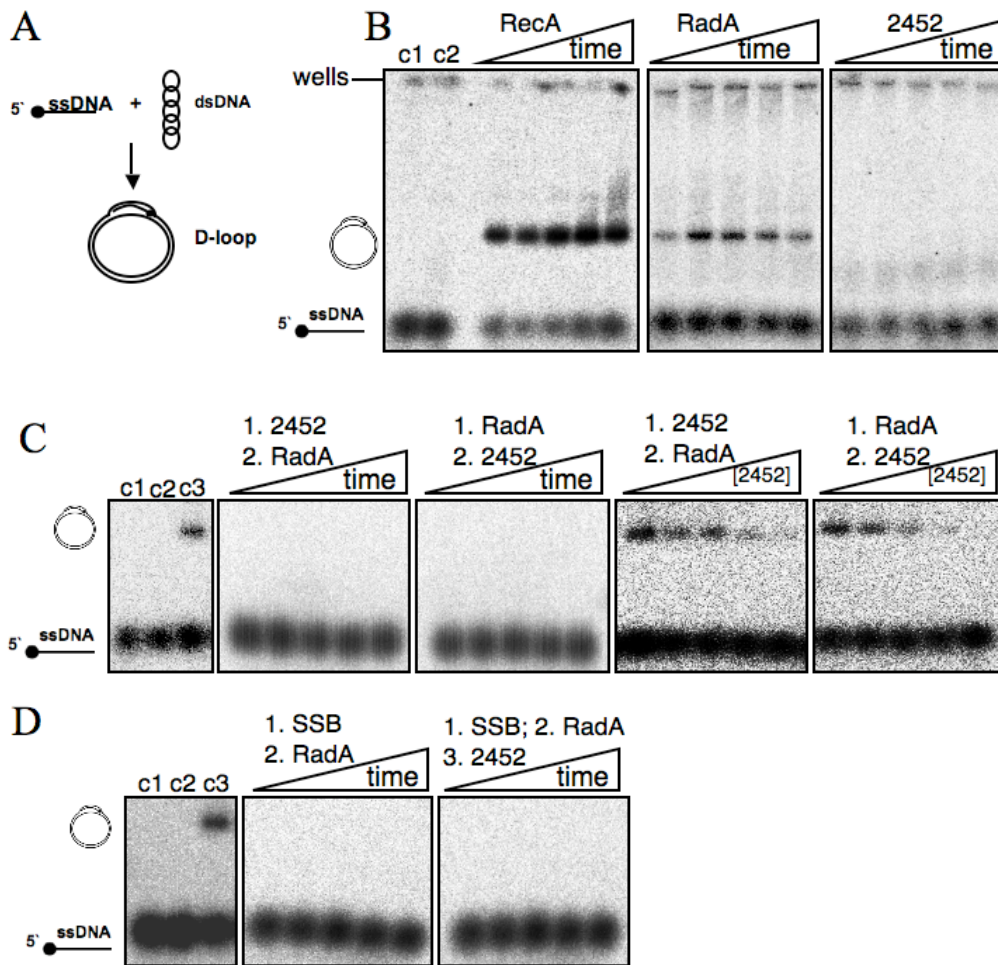
Figure 7.11: D-loop formation by RadA and Sso2452

D-loop reactions were performed over a 30-min time course (1, 5, 10, 20, 30 min) at 60 °C (unless otherwise stated). 5' [^{32}P]-end labelled 80mer was incubated with protein for 5 min prior to addition of dsDNA to initiate the reaction. Where addition proteins were added, the order of addition is indicated. Each protein addition was followed by a 5-min incubation. In all cases, reactions were initiated by dsDNA. (A) Schematic of D-loop reaction. (B) D-loop formation by RecA at 37 °C (time course: 1, 2, 4, 8, 16 min), RadA and Sso2452. (C) Left hand: Order of addition experiments where Sso2452 was pre-incubated with ssDNA prior to the addition of RadA (or vice versa). Right hand: inhibition of RadA as a function of Sso2452 concentration (titration: 0.5, 1, 2, 4, 6 µM). Reactions were initiated by dsDNA and incubated for 30 min. (D) Order of addition experiments with SSB, RadA and Sso2452. Controls in panels B-D: c1, without protein at 30 min; c2, without ATP/ MgCl$_2$ at 30 min; c3, RecA-catalysed D-loop formation after 1 min at 37 °C. The experiment was repeated at least three times, using different protein preparations.

7.9 IDENTIFICATION OF SSO2452 INTERACTING PARTNERS

Proteins often interact with other biomolecules in the cell to perform their *in vivo* function. To shed light on the biological role of Sso2452, a 'guilt by association'

approach was taken whereby potential Sso2452-interacting partners of known function were identified through their ability to co-elute with polyhistidine-tagged Sso2452 bound to a nickel-chelating affinity column (as outlined in Materials and Methods). Given that the eukaryal Rad51 paralogues can act as mediator proteins to displace the single-stranded DNA binding protein, RPA, to promote Rad51-catalysed strand exchange (Sung, 1997), the ability of Sso2452 to interact with archaeal RPA (known as SSB) and RadA was investigated. In contrast to a subsequent study involving the Sso2452 orthologue from *S. tokodaii* (Sheng et al., 2008b), Sso2452 from *S. solfataricus* did not interact with SSB or RadA (figure 7.12). Since Sso2452, RadA and SSB bind DNA, this study was performed in the presence of ethidium bromide, which prevents the indirect association of proteins by DNA bridging.
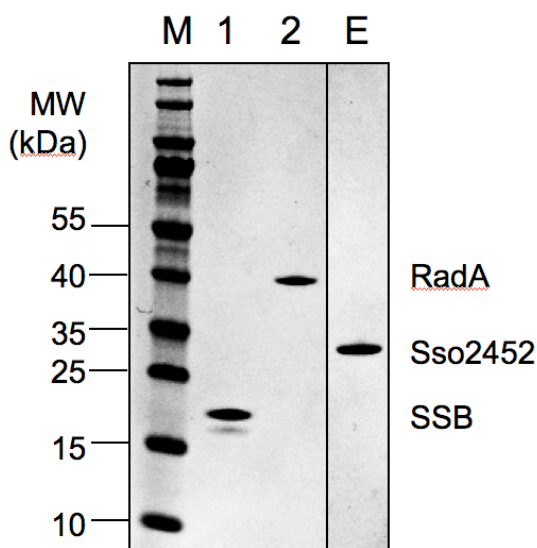
Figure 7.12: Sso2452-interacting partners Polyhistidine-tagged Sso2452 (100 µM) was bound to a 1 ml HiTrap Chelating HP column. Untagged SSB was loaded onto the column, which was then washed thoroughly. The column flow-through was collected and analysed by SDS-PAGE (1). Untagged RadA was loaded onto the column, which was then washed thoroughly. The column flow-through was collected and analysed by SDS-PAGE (2). Bound material was eluted from the column with 500 mM imidazole and analysed by SDS-PAGE (E). The experiment was repeated three times using different preparations of RadA and Sso2452.

7.10 X-RAY CRYSTALLOGRAPHIC STUDY OF SSO2452

Purified Sso2452 was submitted to the Scottish Structural Proteomics Facility (University of St Andrews) for crystallisation trials. Successful crystallisation of Sso2452 was achieved using the hanging-drop vapour-diffusion method under the following crystallisation conditions: 9% (w/v) PEG 8000, 0.1 M Mes (pH 5.5), and 0.14 M zinc-acetate. Crystals were cryo-protected by passing a solution of 10% (w/v) PEG 8000, 0.1 M Mes (pH 5.5), 0.15 M zinc-acetate and 30% (w/v) PEG 400 before being frozen by rapid immersion in liquid nitrogen. Crystal screening was performed at BM14UK (Grenoble; ESRF) resulting in a 2.0 Å data set revealing a single molecule within the asymmetric unit. For full crystallography and refinement data, refer to

(McRobbie et al., 2009). The structure was solved by the Scottish Structural Proteomics facility (University of St Andrews) and the analysis presented in this chapter was achieved by communication with Prof. J. H. Naismith and Dr L. Carter (University of St Andrews).

7.10.1 STRUCTURE OF SSO2452

The structure of Sso2452 was solved by molecular replacement using the structure available for RecA superfamily ATPase Pho0284 from *P. horikoskii* (PDB code: *2dr3*, unpublished). The final model contains 220 proteins residues, 98 water molecules, a pyrophosphate moiety and four zinc atoms (figure 7.13). Consistent with previous observations of proteins possessing RecA-like domains, Sso2452 consists of a large twisted central β-sheet, sandwiched between α-helices on both sides. The central β-sheet consists of six parallel strands (β3 to β8) followed by three anti-parallel strands (β9 to β11). There are small regions of disorder between residues 167-177 (ssDNA binding loop 2, "L2") and residues 93 to 96. Electron density was not observed for the 27 residues at the C-terminus of the protein. A single non-prolyl cis peptide bond linking serine 131 was observed, which is consistent with other RecA superfamily members and appears to be a hallmark of the protein class (Lusetti and Cox, 2002). The Sso2452 structure (which contains a monomer in the asymmetric unit) does not form any obviously biological relevant quaternary structure by crystal packing.
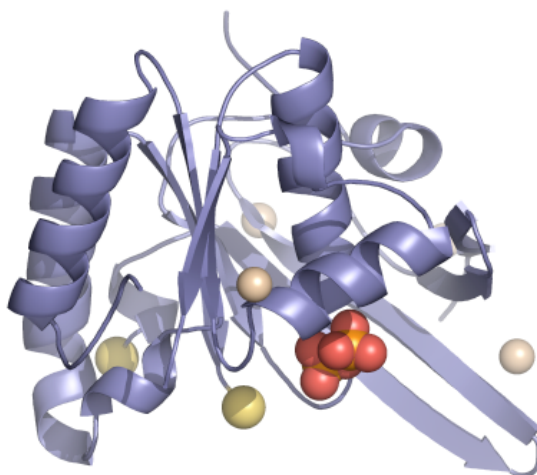
Figure 7.13: Crystal structure of Sso2452

Ribbon structure of the Sso2452 monomer (slate blue). Four Zn^{2+} ions, which are believed to be an artifact of crystallization, are shown as peach spheres. The termini of the disordered L2 DNA binding loop are shown as yellow spheres. Pyrophosphate was modeled into the structure and is shown in red/ orange space-filled sphers.

The four zinc atoms observed in the crystal structure of Sso2452 are coordinated by amino acids H212, H215, E235, H199, E78 and N64, as shown in figure 7.14. This does not suggest that Sso2452 is a metalloprotein; rather the zinc atoms are likely artefacts of the crystallisation conditions, which included zinc-acetate.
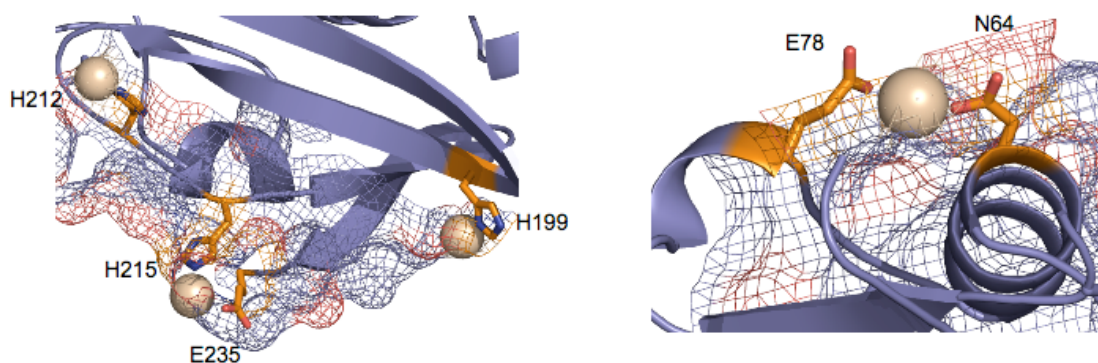
Figure 7.14: Sso2452 coordination of zinc atoms

The location, electron density and coordinating residues of the four zinc atoms bound to Sso2452 (blue slate ribbon) during protein crystallisation are shown. Zinc atoms, grey spheres.

Despite the absence of pyrophosphate (or compound with a similar atomic arrangement) in the crystallisation or cryoprotection solution, a large area of elongated density was observed with a tetragonal arrangement at each end. This density coincided with the approximate location of the Walker motifs of Sso2452 (figure 7.15 A) and in a position equivalent to the site of ADP binding in the homologous RecA-like protein Pho0284 from *P. horikoskii* (figure 7.15 B). Consequently, this region of density was modelled as a pyrophosphate moiety. It is likely that endogenous *E. coli* pyrophosphate bound to the protein and was retained during purification and crystallisation procedures.
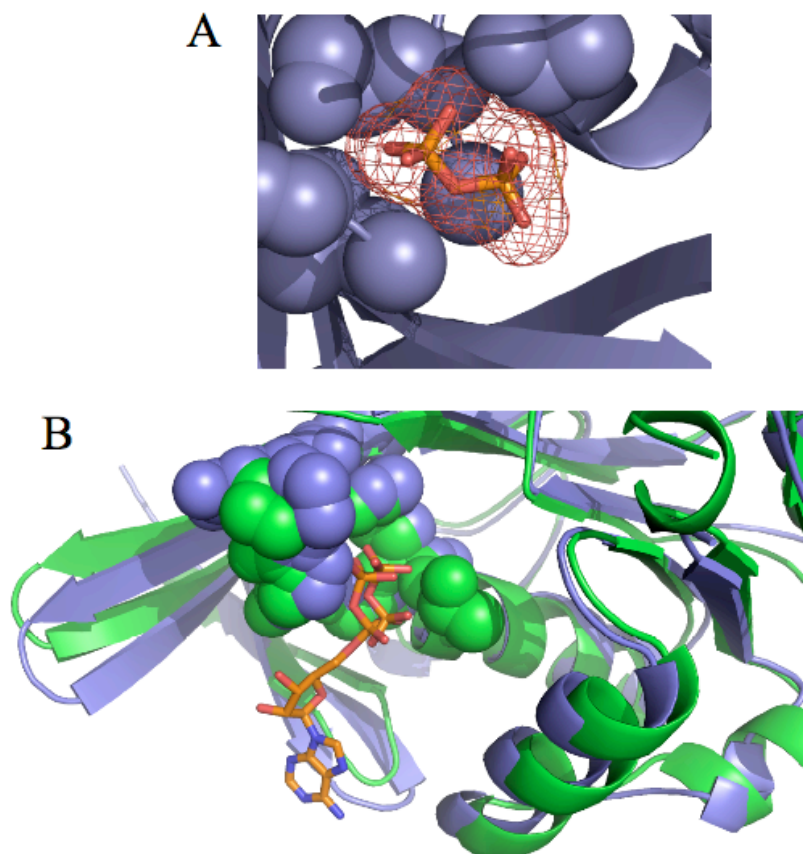
Figure 7.15: Expected site of NTP binding by Sso2452

(A) Unexpected electron density was observed in close proximity to the Walker A motif (blue slate spheres) of the Sso2452 (blue slate ribbon) crystal structure. Pyrophosphate (sticks) was modelled into this position since (B) superimposition of the Sso2452 (slate ribbon) and Pho0284 (green ribbon) structures revealed that ADP (sticks) occupied this site in the latter structure. Walker A motif shown as space-filled green (Pho0284) or slate (Sso2452) spheres.

7.10.2 STRUCTURAL COMPARISON OF SSO2452 TO OTHER RECA SUPERFAMILY PROTEINS

A structural similarity search conducted using secondary-structure matching (Krissinel and Henrick, 2004) confirms the similarity of Sso2452 to other members of the RadA/ Rad51/ RecA family. Sso2452 shares the highest structural similarity with the structure used for molecular replacement, the hexameric aRadC paralogue Pho0284 from *P. horikoshii* (PDB code: *2dr3*) with a root mean square deviation (RMSD) of 1.39 Å over 213 aligned Cα atoms. The hexameric quaternary structure of Pho0284 is shown in figure 7.16, in which one subunit is overlaid with the Sso2452 structure. Crystallographic studies found Sso2452 as a monomer in the asymmetric unit and found no evidence for higher order, multimeric structures from crystal packing. To catalyse

strand exchange, however, proteins from the RecA superfamily must multimerise on DNA to form extended, helical nucleoprotein filaments. Close scrutiny of the interfaces between the Pho0284 and Sso2452 structures revealed a 5 Å change in the location of the loop between β8 and β9 (figure 7.16 B). Residues within this loop, including Glu198, Glu201 and Leu203, and Asp200, are known to stabilise the Pho0284 interface by forming a salt bridge with Arg241 (positions indicated on structure in figure 7.16 B). The position of the zinc atom, partially coordinated by residues His199 and His215, in the Sso2452 crystal may have interfered with a similar protein:protein interface, disrupting any possible quaternary structure. Furthermore, the hexameric structure of Pho0284 reveals that ADP was bound at the active site (figure 7.16 A), which is located directly at the interface between individual Pho0284 monomers. In the equivalent position in the Sso2452 structure, pyrophosphate was bound (as described above). Since ATP/ ADP binding has been shown to play an important part in the RecA-family protein assembly (Logan et al., 2001), it is possible that the pyrophosphate has prevented Sso2452 from forming any higher-order structures.

Sso2452 is a good structural match to the core domain of RadA from *S. solfataricus* (PDB code: *2bke*) with 184 Cα atoms aligning with a RMSD of 1.84 Å (figure 7.17). Sso2452 and Pho0284, like RadB, lack the N-terminal helix-hairpin-helix domain that is implicated in dsDNA binding (Aihara et al., 1999; Kinebuchi et al., 2005), and the short β-strand polymerisation motif including a conserved phenylalanine that forms a "ball and socket" joint with adjacent RadA monomers (Shin et al., 2003). The lack of these features probably corresponds to a fundamentally different role for the aRadC paralogues, and an inability to function as recombinases on their own.
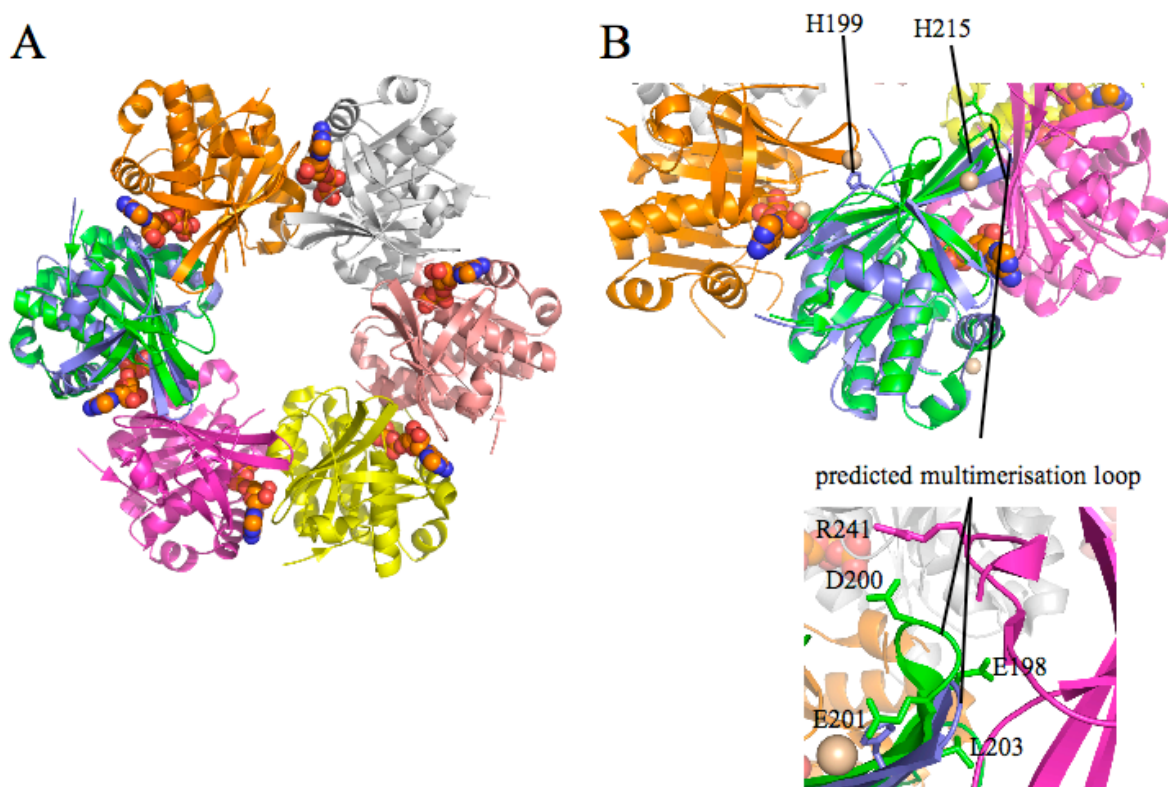
Figure 7.16: Structural superimposition of Sso2452 and Pho0284

(A) Ribbon diagram showing the structural superimposition of Sso2452 (blue slate) with a single monomer of the hexameric Pho0284 structure (PDB code: *2dr3*; each monomer highlighted by differential colouring; superimposing monomer shown in green). In Pho0284, ADP (shown as space-filled spheres) binds at each monomer interface. (B) Zoomed-in region showing superimposition of Pho284 (three monomers shown coloured orange, green and magenta) with Sso2452 (slate blue). The position of ADP (shown as spheres) in the Pho0284 structure may have stabilised the multimerisation of the protein. The position of zinc ions (grey spheres), coordinated by residues H199 and H215, may have prevented the assembly of higher-order structures of Sso2452. Finally, the position of the predicted multimerisation loop, containing residues essential for Pho0284 hexamerisation (D200, E201, E198, L203 and R241 shown as sticks) was shifted by 5 Å in the Sso2452 structure.

243
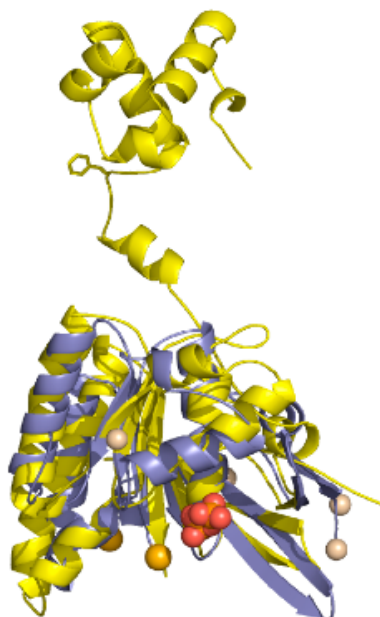
Figure 7.17: Structural superimposition of Sso2452 and RadA from *S. solfataricus*

Ribbon diagram showing the structural superimposition of *S. solfataricus* RadA (yellow; PDB code: *2bke*) and Sso2452 (slate blue). RadA has an extra N-terminal domain linked to the core domain via a short polymerisation motif that includes a conserved phenylalanine residue (F73; side chain indicated). Pyrophosphate was modeled into the structure and is shown as red/ orange space-filled spheres.

7.10.3   DNA BINDING RESIDUES AND SURFACES

Sso2452 and Pho0284 lack the N-terminal helix-hairpin-helix domain that is implicated in dsDNA binding, and the short β-strand polymerisation motif including a conserved phenylalanine that forms a "ball and socket" joint with adjacent RadA monomers (figure 7.17). The ability of members of the RecA family to bind ssDNA has been attributed to two loops, named L1 and L2 (figure 7.18). The recent co-crystal structure of RecA with ssDNA revealed that, upon DNA binding, the L1 and L2 loops become ordered forming a short helix and a β-hairpin structure, respectively (Chen et al., 2007b). The L1 motif comprises three arginine residues that are conserved across the RecA/ Rad51/ RadA family. Each of these residues has been implicated in ssDNA binding in RadA from *S. solfataricus* (Chen et al., 2007b). Interestingly, the aRadC paralogues lack two of these residues; functional differentiation observed between RadA and the aRadC paralogues may be related to the absence of these conserved arginine residues. While the L2 motif is not strongly conserved between RecA and Rad51/ RadA, two consecutive glycine residues at the C-terminal end of the L2 motif are conserved in RadA and Rad51 orthologs. These residues have been implicated in

binding a phosphate group of ssDNA. Importantly, these residues are not present in the archaeal aRadC paralogues, potentially explaining the functional divergence observed between RadA and Sso2452.


```
                        α11          α12

        SsoRadA  RAEYPGRENLAVRQ
   L1   Sso2452  VSALFLDKPAMARK
        Pho0284  VTTLYINKPAMARS


        SsoRadA  NQVMARPDMFYGDPTVAVGGKTLYHV
   L2   Sso2452  YATSQYAITTSQAFGFGVEHVADG
        Pho0284  IFVSQVSVGERGFGGPGVEHGVDG
```
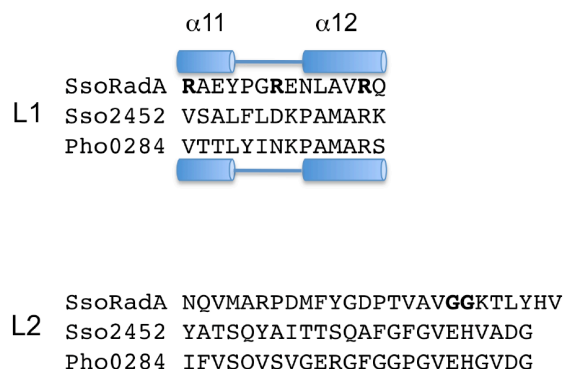
Figure 7.18: Sequence alignment of loops 1 and 2 implicated in RadA: ssDNA binding

The crystal structure of RadA (PDB code: *2bke*) from *S. solfataricus* revealed two loops (L1 and L2) implicated in ssDNA binding. In L1, three conserved arginine residues (highlighted in bold) are necessary for ssDNA interaction, while in L2, two glycine residues were shown to be important for RadA interaction with DNA. These residues are not conserved in archaeal aRadC paralogues, such as Sso2452 or Pho0284.

To investigate any other factors that may influence the strong binding affinity of Sso2452 for ssDNA (K_D of 120 nM) in comparison to *S. solfataricus* RadA (K_D of 2.4 µM), the electrostatic surface of Sso2452 was analysed (figure 7.19). However, the results failed to reveal any putative DNA binding sites and no striking distribution of charge on the surface of the protein was observed. The absence of the L2 DNA binding loop from the crystal structure may affect such analysis. The ability of proteins to bind ssDNA via a combination of ionic and hydrophobic interactions may also reflect the absence of a strong electrostatic "signal".
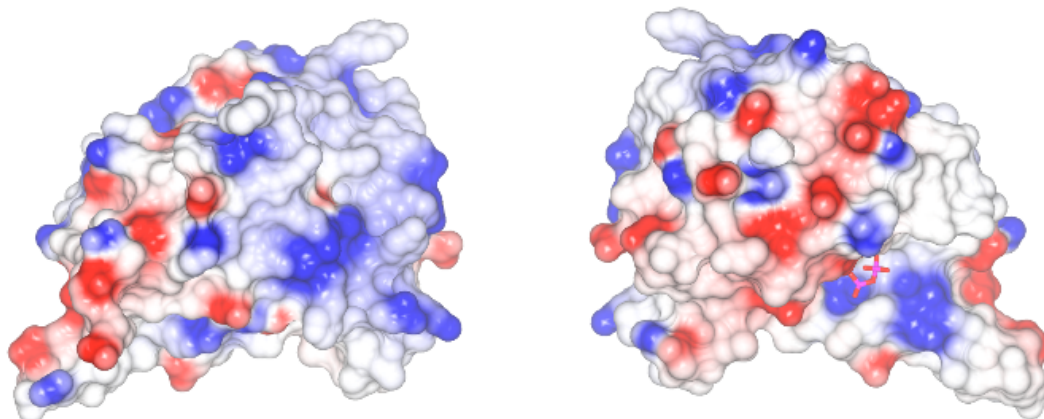
Figure 7.19: Electrostatic surface of Sso2452

To investigate potential DNA binding regions of Sso2452, the electrostatic surface of the front (left) and back (right) of the protein was analysed. Red, acidic residues; blue, basic residues; white, neutral residues.

The crystal structure of Sso2452 highlights the key differences between the aRadC paralogues and the canonical RadA that can be related to their differing functions. The ability of the aRadC paralogues to bind tightly to ssDNA deserves further investigation. Since the L1 and L2 loops of the aRadC family seem to have a less basic character than those of RadA, the interaction may involve the intercalation of hydrophobic residues between DNA bases as is seen in SSB proteins, as well as ionic interactions with the phosphodiester backbone.

7.11   DISCUSSION AND CONCLUDING REMARKS

In bacteria, a single DNA recombinase, RecA, is responsible for catalysing strand exchange during HR/ DSBR. In contrast, archaea and eukaryotes encode a primary DNA recombinase, RadA or RAD51 respectively, in addition to at least one paralogue. In humans, seven RAD51-like genes exist (RAD51A, RAD51B, RAD51C, RAD51D, XRCC2, XRCC3 and DMC1) and mutation of these proteins is associated with defective recombination and genomic instability (Masson et al., 2001). Thus, it has been proposed that the RAD51 paralogues regulate the recombinational activity of RAD51 (Masson et al., 2001). For instance, human RAD51B has been shown to promote the assembly of RAD51 nucleoprotein filaments during HR (Takata et al., 2000). Yeast, however, encode only two paralogues of Rad51: Rad55 and Rad57. These paralogues interact to form a heterodimer that has been reported to markedly enhance the recombination activity of Rad51 in the presence of the otherwise inhibitory single-

strand DNA binding protein, RPA (Sung, 1997). These findings from eukaryotic systems have stimulated research investigating the roles of the archaeal RadA paralogues. This chapter reports the biochemical and structural characterisation of one of the RadA paralogues, Sso2452, from *S. solfataricus*. Based on sequence analysis, Sso2452 appears to be a representative member of the large family of aRadC paralogues.

The *S. solfataricus* genome encodes three RadA paralogues (Sso0777, Sso1861, and Sso2452) with mRNA transcript levels present at approximately 10-20% of RadA, which is itself a highly abundant protein. While mRNA levels often correlate well with protein concentration, it is important to clarify that this is not always the case and significant post-translational regulation of protein levels could dramatically alter the final protein levels of RadA and paralogues in the cell. Komori and coworkers (2000a) reported that RadB mRNA was present in *P. furiosus* cells at a concentration 2-5 fold less than RadA mRNA; however, Western blot analysis revealed that final protein levels of RadB were 200-fold less than those of RadA (Komori et al., 2000a).

Like *radA*, UV radiation did not affect the expression of the three paralogues from *S. solfataricus*. Similarly, *radA* and *radB* from the hyperthermophilic archaeon *P. furiosus* were not upregulated in the presence of UV radiation, indicating that they were both constitutively expressed *in vivo* (Komori et al., 2000a). In contrast, the expression of eukaryotic RecA (Fernandez de Henestrosa et al., 1991) and Rad51 (Basile et al., 1992) was shown to be UV inducible. High constitutive expression may prove to be a general characteristic of proteins from hyperthermophilic organisms, which require highly efficient DNA repair to survive high temperature environments.

The structure of Sso2452 reveals a typical RecA-like fold with high similarity to the unpublished structure of Pho0284, an aRadC member from *P. horikoskii*. The overall structural configuration of Sso2452 is likely to be highly representative of the aRadC family in general. Hexamers, heptamers, and helices have all been observed in RadA/ RecA/ Rad51 structures, created either by non-crystallographic symmetry or as the result of crystal packing (Story et al., 1992; Yu and Egelman, 1997; Shin et al., 2003; Ariza et al., 2005). Additionally, helical structures have been observed by electron micrographs (Yu and Egelman, 1990; Yang et al., 2001; Shin et al., 2003). The helical arrangement of subunits produces a large positively charged outer-groove, which is believed to bind dsDNA (Shin et al., 2003). In some structures, the ATPase domain is at the subunit interface, where it facilitates ATPase

cooperativity (Shin et al., 2003). This would seem to be the case in the Pho0284 structure from *P. furiosus*. The hexameric quaternary structure of Pho0284 is consistent with the known propensity of RecA-family recombinases to adopt 6-, 7- and 8-membered rings as well as a variety of helical forms. Together with the observation of helical order in the RadB crystal structure (Akiba et al., 2005), this suggests that aRadC proteins can multimerise despite the lack of the polymerisation motif and N-terminal domain that are present in RadA/ Rad51. Despite this, evidence for multimerisation of Sso2452 was lacking: gel filtration analysis suggested that Sso2452 existed as a monomer in solution, a single Sso2452 molecule was observed in the asymmetric unit and analysis of the structure did not reveal any obviously biological relevant quaternary structure by crystal packing. The Sso2452 structure was analysed by PISA (Protein Interfaces Surfaces and Assemblies), which predicts likely quaternary structures that may arise based on the interfaces and surfaces of a macromolecule (Krissinel and Henrick, 2005). The program provides a "complexation significance score" (CSS) which denotes the significance of an interface/ surface in terms of its ability of promote the assembly of a higher-order structure (communication with Prof. J. Naismith). Sso2452 received a CSS score of zero, where scores range from zero to one as the likelihood of protein complexation increases. A possible explanation for the absence of higher-order assemblies of Sso2452 arises upon structural superimposition of Sso2452 and Pho0284: some minor changes in conformation were observed, with a movement of a loop containing a number of residues shown to be essential for multimerisation of the Pho0284 protein. It is important to note that the Pho0284 protein was crystallised in the presence of ADP, which has been shown to bind at the interface between monomers, and has been proposed to potentially stabilise the hexameric structure. Since Sso2452 was crystallised in the absence of nucleotide cofactors, it as not possible to ascertain the impact of such molecules on the multimerisation of the protein. In addition, the position of two zinc ions, coordinated by residues H199 and H215, may have prevented complexation of Sso2452.

The biochemical data presented in this chapter revealed that Sso2452 differs from RadA. While both catalyse ATP hydrolysis with similar efficiency, Sso2452 binds ssDNA with significantly higher affinity (22-fold greater) than RadA. It is likely that the tight binding of Sso2452 influences its biological activity, providing a possible explanation for the functional divergence of Sso2452 and RadA observed throughout this chapter.

Although supporting limited ATP-independent strand exchange *in vitro*, Sso2452 cannot catalyse D-loop formation, suggesting that it does not function as a recombinase *in vivo*. This is consistent with the known properties of *P. furiosus* RadB (Sung, 1997; Komori et al., 2000a) and the yeast Rad55/ Rad57 heterodimer (Sung, 1997). The ATP-independence of the reaction suggests that these results may simply reflect passive equilibrium binding of DNA by all proteins. While SSB binds ssDNA with a similar affinity as Sso2452 (Kernchen and Lipps, 2006), it failed to promote limited strand exchange. This suggests that while SSB may "trap" ssDNA by its tight association with the nucleic acid, Sso2452 has a different function in the cell.

Despite lacking *bona fide* recombinase activity, the Rad55/ Rad57 heterodimer has recently been shown to promote Rad51-catalysed homologous recombination in the presence of the single-strand DNA binding protein RPA (Sung, 1997). RPA, as we have shown for SSB, inhibits the recombinase activity of Rad51 (Sung, 1997). This has highlighted the possibility for the archaeal RadA paralogues to occupy a similar "mediator" role *in vivo* by influencing the activity of RadA. Our results reveal that the SSB-coated DNA inhibits the recombinase activity of RadA *in vitro*. However, Sso2452 failed to relieve this inhibition, suggesting that it does not function as a mediator of RadA, akin to that observed for the Rad55/ Rad57 heterodimer described above. It is possible, however, that Sso2452 fulfils a mediator function in collaboration with additional proteins, *in vivo*, such as DNA helicases. Our data also suggest that Sso2452 can prevent RadA-mediated D-loop formation when incubated with ssDNA either before or after the addition of RadA. Similar observations have been reported for RadB from *P. furiosus*, which appears to lack recombinase activity of its own and inhibits the ability of RadA to catalyse D-loop formation (Komori et al., 2000a). One possible explanation for this finding is the significantly higher binding affinity of Sso2452 for ssDNA in comparison to RadA. In this light, it is unsurprising that Sso2452 may sequester ssDNA, inhibiting RadA accessing the DNA to catalyse strand exchange. However, this argument is perhaps not consistent with the observation that Sso2452 inhibited D-loop formation by pre-formed RadA-ssDNA nucleoprotein filaments. This particular finding suggests that Sso2452 actively inhibits the recombinase activity of RadA. The yeast DNA helicase Srs2 (Krejci et al., 2003; Veaute et al., 2003) and human Bloom's and Werner's syndrome helicases (Bohr, 2008) have also been shown to possess anti-recombinase activity. The regulation of homologous

recombination in eukaryotes is important for genome stability and the aRadC paralogues may perform a similar function.

Based on the Sso2452 biochemistry reported in this chapter, certain parallels can be drawn with the euryarchaeal RadA paralogue, RadB (Komori et al., 2000a). Although RadB lacks ATPase activity, which is in contrast to Sso2452, the crystal structures of Sso2452 and RadB exhibit high levels of similarity and both proteins bind ssDNA tightly. As observed for Sso2452, RadB inhibits RadA-mediated strand exchange when incubated before and after the formation of RadA nucleoprotein filaments, suggesting a similar anti-recombinogenic function. Unlike Sso2452, however, RadB fails to catalyse ATP hydrolysis. *P. furiosus* RadB has been shown to interact directly with RadA and the Holliday junction resolving enzyme, Hjc (Komori et al., 2000a). Although the Sso2452 orthologue from *S. tokodaii* reportedly interacts with SSB and RadA (Sheng et al., 2008a; Sheng et al., 2008b), a preliminary investigation into potential Sso2452-interacting partners revealed no obvious candidates. It is important to note that such interaction studies are suspceptible to the generation of false positives when proteins have a DNA binding activity in common. Together, these data suggest that aRadC and RadB may have different functions *in vivo*, which is supported by the observation that many euryarchaea, including *P. furiosus*, have aRadC orthologs in addition to radB. The lack of evidence to support stable protein:protein interactions involving Sso2452 limits our understanding of the *in vivo* function of this protein. Further biochemical and genetic characterisation of the other archaeal RadA paralogues will be required to delineate their role in homologous recombination.

CHAPTER 8

CONCLUSIONS AND FUTURE WORK

8.1   MOLECULAR BASIS OF DISEASE-ASSOCIATED MUTATIONS OF XPD

The molecular basis of five human XPD mutations that cause *xeroderma pigmentosum* (XP) and trichothiodystropy (TTD) was investigated by introducing the equivalent mutation in the homologous protein from *Sulfolobus acidocaldarius*. Three of these mutations, A204Y (SacXPD) that confers XP in humans, and R373Q (SacXPD) and R531Q (SacXPD), both of which are associated with TTD, exhibited biochemical traits that deviated from those observed for the WT enzyme. Low yields of purified A204Y XPD were obtained, relative to the WT protein, and biochemical studies revealed that the mutation conferred thermal sensitivity. Structural assessment of the mutation revealed that the A204 residue projects into a tightly-packed helical bundle within the Arch domain of XPD, suggesting that its substitution for a bulky tyrosine residue would cause structural perturbations. The Arch domain of XPD has been implicated in mediating protein: protein interactions between XPD and other protein subunits of the TFIIH complex in humans. Thus, the A204Y mutation may not only disrupt the structural stability of XPD, but also the overall stability of the TFIIH complex in humans.

Although R373Q and R531Q XPD purified according to the WT protein, they bound DNA with reduced affinity and failed to function as active helicases. The R531Q mutant also hydrolysed ATP at a rate 170 % compared to WT XPD, indicative of an uncoupling between ATPase and helicase activity. By contrast, the R373Q mutant

displayed a reduced capacity to hydrolyse ATP. Upon inspection of the XPD structure, it was observed that both residues were located along the expected path taken by ssDNA through the protein. Mutation of positively charged residues that engage with the negatively charged DNA backbone for a neutral glutamine residue could explain the diminished capacity of these mutants for DNA binding. Taken together, the data suggests that the R373Q and R531Q mutations of XPD cause XP by affecting the biochemical functions of XPD rather than the stability of TFIIH. This is consistent with the proposal that XP-related mutations affect the NER, as opposed to the transcription, role of TFIIH (Bootsma and Hoeijmakers, 1991).

Chapter 3 revealed the potential of using an archaeal enzyme to delineate the molecular properties of a eukaryotic protein and, ultimately, aid our understanding of the development of human disease.

8.2   PURIFICATION AND CHARACTERISATION OF DING FROM *STAPHYLOCOCCUS AUREUS*

Chapter 4 described the purification and characterisation of DinG from *Staphylococcus aureus* (SarDinG), an XPD-related protein. Sequence analysis of SarDinG revealed six canonical helicase motifs and an N-terminal extension that was absent from XPD and other XPD-related proteins. Despite the presence of helicase motifs, an ability to bind both single- and double-stranded DNA and hydrolyse ATP, SarDinG failed to unwind a broad range of DNA substrates. The XPD protein from *S. acidocaldarius* (SacXPD) encodes an FeS cluster, which is essential for the helicase activity of the enzyme (Rudolf et al., 2006). Indeed, the helicase activity of several proteins related to XPD, including human FancJ, *S. cerevisiae* Rad3 and *E. coli* DinG, also depend on the presence of an FeS cluster (Rudolf et al., 2006; Pugh et al., 2008; Ren et al., 2009). Unlike XPD, SarDinG does not contain an FeS cluster domain, which may explain the absence of helicase activity.

It has become increasing clear that several putative helicases do not possess helicase activity despite encoding helicase motifs. Rather, these proteins often couple ATP hydrolysis to DNA translocation (Singleton et al., 2007). Therefore, the ability of DinG to translocate along ssDNA was investigated using a streptavidin-displacement assay (Byrd and Raney, 2004). The results revealed that DinG could not displace streptavidin from the biotinylated oligonucleotide. This does not necessarily imply that

DinG cannot translocate along DNA; SacXPD functions as a *bona fide* DNA helicase despite its inability to disrupt the streptavidin: biotin linkage in this assay (thesis of Dr J. Rudolf, University of St Andrews). This experiment does reveal, however, that DinG does not move along DNA with sufficient force to remove DNA-bound proteins, e.g. the single-stranded DNA binding protein (SSB), from DNA, as has been observed previously for DNA helicases such as archaeal Hel308 and the Dda helicase from bacteriophage T7 DinG (Byrd and Raney, 2006; Richards et al., 2008b).

Unlike XPD, SarDinG also encoded an N-terminal putative nuclease domain that exhibited homology to the epsilon subunit of polymerase III from *E. coli*, the archetype member of the DnaQ superfamily of nucleases. Biochemical analysis revealed that DinG catalysed degradation of DNA with 3'-5' polarity, as expected for a DnaQ superfamily protein, an activity that was inhibited by ATP. Site-directed mutagenesis was employed to generate a DinG mutant encoding a disrupted Walker A motif (K304A DinG). Disruption of the Walker A motif significantly reduces the ability of the protein to bind ATP and completely abrogates its ATPase activity. The K304A DinG protein possessed 3'-5' exonuclease activity that was inhibited by ATP and ADP. Given that this mutant cannot hydrolyse ATP, these findings suggest that DinG nuclease activity was inhibited by ATP binding and hydrolysis. The fact that ATP binding within the helicase domain can inhibit the nuclease domain suggests that the nuclease activity of DinG is under the control of the helicase domain. Furthermore, the ability of DinG to digest ssDNA suggests that nuclease activity does not depend on DNA binding by the helicase domain. The study also suggested that DinG required a single-stranded region to access the DNA, and failed to digest DNA from a blunt, fully base-paired end. However, it is clear that further study is required to fully understand the mechanism by which DinG digests DNA.

A preliminary model (figure 4.26) was proposed to explain the ATP-regulated exonuclease activity of DinG. In the absence of ATP, DinG binds DNA. Upon ATP binding and hydrolysis, a conformational change takes place within the helicase domain that is anticipated to alter the conformation of the nuclease domain or its position on the DNA such that nuclease activity is prohibited. Release of ADP activates DinG nuclease activity and one base pair of the DNA is digested in the 3'-5' direction. Rebinding of ATP within the helicase domain switches the nuclease domain off and the cycle continues.

8.3 MECHANISTIC AND FUNCTIONAL CHARACTERISATION OF ARCHAEAL HEL308

Chapter 5 and 6 provided a structural, biochemical and biophysical assessment of the DNA helicase, Hel308, from *S. solfataricus* (strain PBL2025). Previous work supported the ability of Hel308 to preferentially unwind stalled replication forks *in vitro* but failed to provide a consistent model for its role *in vivo*: the Hel308 homologue from *M. thermoautotrophicus* (Mth Hel308) (Guy and Bolt, 2005) and *P. furiosus* (Pfu Hjm) (Fujikane et al., 2006) was reported to possess 3'-5' helicase activity and thus unwind the lagging strand of replication forks, while the homologue from *S. tokodaii* (StoHjm) appeared to exhibit bidirectional helicase activity consistent with an ability to unwind both the leading and lagging strand of a replication fork (Li et al., 2008). It had been proposed that these findings reflected the evolutionary divergence of the Crenarchaeota and Euryarchaeota (Li et al., 2008). However, work reported in chapter 5 reveals that the crenarchaeal *S. solfataricus* PBL2025 Hel308 (Sso PBL2025 Hel308) protein unwinds DNA, in an ATP-dependent manner, with 3'-5' polarity, contradicting this hypothesis. While Sso PBL2025 Hel308 exhibited unidirectional helicase activity, it appeared to translocate along DNA in both the 3'-5' and 5'-3' direction using the streptavidin-displacement assay (Byrd and Raney, 2004). Preliminary studies of Hel308 using single-molecule fluorescence resonance energy transfer supported this observation with individual molecules of Hel308 migrating along DNA in either the 3'-5' or 5'-3' direction, and in some cases, a single protein molecule was shown to migrate bidirectionally during a single binding event. While several DExD/ H RNA helicases have been shown to translocate along DNA without strict polarity (Tanaka and Schwer, 2006; Yang and Jankowsky, 2006), this represents a rare observation for a SF2 DNA helicase and further study is required to understand the mechanism and implications of this finding.

Together, these data suggest that Hel308 can translocate along the lagging or leading strand template in a 3'-5' direction to unwind the lagging strand or parental duplex, respectively. Alternatively, Hel308 can translocate along the lagging or leading strand in a 5'-3' direction but, in this instance, is not accompanied by DNA unwinding. This suggests that Hel308 may bind DNA in two distinct orientations such that the β-hairpin loop of domain 2, implicated in strand displacement, is only correctly positioned to catalyse unwinding in one binding mode.

The elucidation of the Hel308 crystal structure led to the proposal that domains 3-5 of Hel308 were necessary for coupling ATP hydrolysis (performed by domains 1 and 2) to DNA translocation and unwinding (Buttner et al., 2007). The data presented in this thesis supports this hypothesis since Hel308 lacking domains 3-5 failed to translocate along and unwind DNA but exhibited an enhanced ATPase activity compared to the WT enzyme. Consistent with previous work (Richards et al., 2008b), the data reveals that domain 5 of Hel308 is an autoinhibitory domain that serves to limit the unwinding activity of the protein. Removal of domain 5 did not influence the directionality of DNA translocation or unwinding, suggesting that it does not play a role in determining the substrate preference of the enzyme. This contradicts the proposal that domain 5 is involved in binding branched nucleic acids, including replication forks, and positioning the protein to unwind the lagging strand (Buttner et al., 2007).

The ability of Hel308 to disrupt the biotin: streptavidin linkage during the streptavidin displacement assay (Byrd and Raney, 2004) suggested that the protein may be able to remove DNA-binding proteins during translocation. Indeed, during both DNA translocation and unwinding, Hel308 was shown to displace a variety of DNA-bound proteins, including the single-stranded DNA binding protein (SSB), the recombinase protein RadA, the RadA paralogue Sso2452, and the single- and double-stranded DNA binding protein Alba1. These findings are particular exciting considering the proposed role of Hel308 during the repair of stalled replication forks by homologous recombination. On the basis of these findings, we proposed a model for how Hel308 may function in the cell (figure 5.18). Upon replication fork stalling, Hel308 removes the lagging strand of the fork to produce a single-stranded region of DNA that is subsequently bound by SSB. For homologous recombination to proceed, RadA must bind this ssDNA tail to catalyse its invasion into a homologous duplex and form a D-loop. Our data reveal that SSB inhibits the recombination activity of RadA; however, Hel308 serves to relieve this inhibition, presumably through its ability to displace SSB from DNA. Thus, Hel308 may act to displace both DNA and DNA-bound proteins to facilitate RadA-catalysed recombination. This is reminiscent of the role played by the bacterial RecFOR complex, which has been shown to displace SSB to promote RecA-catalysed homologous recombination (Umezu et al., 1993; Morimatsu and Kowalczykowski, 2003).

However, our data also revealed that Hel308 can displace RadA from DNA; thus, the results may alternatively imply that Hel308 limits recombinational activity in a

manner analogous to the bacterial UvrD DNA helicase that disrupts RecA nucleoprotein filaments (Veaute et al., 2005). Our finding that Hel308 physically interacts with Hjc, the Holliday-junction resolving enzyme, may support such an anti-recombination role. Shen and coworkers reported that the physical association between the Hel308 (StoHjm) and Hjc (StoHjc) homologues from *S. tokodaii* served to inhibit Hjm helicase activity (Li et al., 2008), thus limiting recombinational repair in a manner akin to that described for the human RecQ DNA helicases, BLM and WRN, for which there are no known homologues in archaea (Bohr, 2008).

Based on the structure of Hel308 in complex with DNA, Hopfner and coworkers proposed a model to describe how the protein coupled ATP binding and hydrolysis to DNA translocation and unwinding (Buttner et al., 2007). The DNA molecule was shown to enter the protein between domains 2 and 4 before passing through a central channel formed between domains 1-4. The gap between domains 2 and 4 was deemed too narrow to accommodate a branched DNA molecule, such as a replication fork, leading to the proposal that these domains would have to move relative to each other to increase the gap width. In addition, it was anticipated that domains 1 and 2 (otherwise referred to as the core RecA-like motor domains) would move relative to each other during ATP binding and hydrolysis, in a manner akin to that observed for bacterial PcrA (Velankar et al., 1999). PELDOR EPR spectroscopy revealed that, in the presence of ADP, domains 1 and 2 moved closer together by 1.2 Å suggesting that the ATP-binding cleft closes upon ATP hydrolysis. Overall, however, the EPR data revealed that Hel308 was a structurally rigid molecule that did not rely on large conformational changes to couple ATP binding and hydrolysis to DNA translocation and unwinding, as reported for bacterial Rep DNA helicase (Korolev et al., 1997). This data is consistent with the observation that the structures of Hel308 in the presence and absence of DNA or NTP analogs are easily superimposed to reveal minimal structural perturbations. This may suggest that Hel308 translocates along and unwinds DNA by employing an inchworm-like mechanism.

Finally, chapter 6 presented a preliminary investigation of Hel308 activity using single-molecule fluorescence resonance energy transfer (smFRET). The data supports the use of polyethylene glycol (PEG)-coated quartz slides for visualisation of single molecules of Hel308 without non-specific adsorption to the negatively-charged surface. It provided a glimpse of Hel308 binding and dissociation from DNA and, of particular interest, the data provided support for the bidirectional translocation of

Hel308 along DNA, indicating that individual molecules can move along DNA in either a 3'-5' direction or 5'-3' direction before dissociation, or indeed can move in both directions during a single binding event. The data highlighted the importance of employing a methodology by which single molecules of Hel308 can be visualised in real-time, which would allow the determination of DNA translocation and unwinding rates. Although initial steps were taken to achieve this, resulting data suffered from the effects of microscope defocusing. This remains the focus of future work.

8.4 PURIFICATION AND CHARACTERISATION OF A RADA PARALOGUE FROM *SULFOLOBUS SOLFATARICUS*

Chapter 7 presented the purification and characterisation of a RadA paralogue, Sso2452, from *S. solfataricus,* that belongs to the recently defined aRadC family (Haldenby et al., 2009). RadA catalyses strand exchange during homologous recombination (HR)/ double strand break repair (DSBR) to facilitate the exchange of genetic material between homologous DNA sequences. Archaea encode a number of RadA paralogues that may be expected to function alongside RadA in a manner analogous to the RAD51 paralogues in eukaryotes (Masson et al., 2001).

S. solfataricus encodes three paralogues of RadA (Sso0777, Sso1861, and Sso2452) with mRNA transcript levels present at approximately 10-20% of RadA, which is itself a highly abundant protein. Like *radA*, the expression of these paralogues was not influenced by UV radiation. A similar finding was observed for the *radA* and *radB* genes from the hyperthermophilic archaeaon *P. furiosus* (Komori et al., 2000a). This suggests that RadA and its paralogues are constitutively expressed, possibly reflecting the high rate of DNA repair required by hyperthermophilic organisms.

The structure of Sso2452 revealed a typical RecA-like fold with high similarity to the unpublished structure of Pho0284, an aRadC member from *P. horikoskii*. It is anticipated that the Sso2452 structure will be representative of all members of the aRadC family. Although recombinases of the RecA family, including bacterial RecA, eukaryotic RAD51, and archaeal RadA, assemble into higher oligomeric structures (Story et al., 1992; Yu and Egelman, 1997; Shin et al., 2003; Ariza et al., 2005), structural data and gel filtration analysis suggested that Sso2452 existed as a monomer. This does not exclude the possibility that the protein oligomerises upon interaction with DNA or other cofactors. Indeed, gel filtration

analysis revealed a higher molecular weight species upon incubation of Sso2452 with DNA.

Biochemical analysis revealed significant differences between Sso2452 and RadA. Although both proteins catalysed DNA-dependent ATP hydrolysis at comparable rates, Sso2452 bound ssDNA with significantly higher affinity compared to RadA. As expected, RadA promoted the ATP-dependent invasion of ssDNA into a homologous double-stranded DNA plasmid to generate a D-loop. However, Sso2452 was not an active recombinase and inhibited the recombination activity of RadA in a concentration-dependent manner. This inhibition was observed regardless of which protein was pre-incubated with DNA suggesting that the inhibition did not simply reflect the lower dissociation constant of Sso2452, but rather that Sso2452 actively inhibited RadA activity. The results suggested that Sso2452 does not function as a mediator of RadA-catalysed recombination, as described for the yeast Rad51 paralogues (Rad55 and Rad57), but instead possesses antirecombinogenic activities akin to those described for the yeast DNA helicase Srs2 (Krejci et al., 2003; Veaute et al., 2003) and human BLM and WRN helicases (Bohr, 2008). The regulation of homologous recombination in eukaryotes is important for genome stability and the aRadC paralogues may perform a similar function.

Sso2452 appears to share several properties with the euryarchaeal RadA paralogue, RadB (Komori et al., 2000a): the crystal structures of Sso2452 and RadB are very similar, both proteins bind DNA with high affinity and exhibit anti-recombinogenic functions by inhibiting the strand exchange reaction catalysed by RadA. Furthermore, both Sso2452 and RadB have been shown to physically interact with the Holliday junction resolving enzyme, Hjc (Komori et al., 2000a). However, it is likely that aRadC proteins and RadB have different functions *in vivo*. This is supported by the observation that many euryarchaea, including *P. furiosus*, have aRadC orthologues in addition to RadB. Further biochemical and genetic characterisation of the other archaeal RadA paralogues will be required to delineate their role in homologous recombination.

REFERENCES

Abella, M., S. Rodriguez, S. Paytubi, S. Campoy, M. F. White and J. Barbe (2007). "The Sulfolobus solfataricus radA paralogue sso0777 is DNA damage inducible and positively regulated by the Sta1 protein." Nucleic Acids Res **35**(20): 6788-97.

Ahnert, P. and S. S. Patel (1997). "Asymmetric interactions of hexameric bacteriophage T7 DNA helicase with the 5'- and 3'-tails of the forked DNA substrate." J Biol Chem **272**(51): 32267-73.

Ahnert, P., K. M. Picha and S. S. Patel (2000). "A ring-opening mechanism for DNA binding in the central channel of the T7 helicase-primase protein." EMBO J **19**(13): 3418-27.

Aihara, H., Y. Ito, H. Kurumizaka, S. Yokoyama and T. Shibata (1999). "The N-terminal domain of the human Rad51 protein binds DNA: structure and a DNA binding surface as revealed by NMR." J Mol Biol **290**(2): 495-504.

Akiba, T., N. Ishii, N. Rashid, M. Morikawa, T. Imanaka and K. Harata (2005). "Structure of RadB recombinase from a hyperthermophilic archaeon, Thermococcus kodakaraensis KOD1: an implication for the formation of a near-7-fold helical assembly." Nucleic Acids Res **33**(10): 3412-23.

Allers, T. and M. Mevarech (2005). "Archaeal genetics - the third way." Nat Rev Genet **6**(1): 58-73.

Anderson, D. G., J. J. Churchill and S. C. Kowalczykowski (1997). "Chi-activated RecBCD enzyme possesses 5'-->3' nucleolytic activity, but RecBC enzyme does not: evidence suggesting that the alteration induced by Chi is not simply ejection of the RecD subunit." Genes Cells **2**(2): 117-28.

Andressoo, J. O., J. H. Hoeijmakers and J. R. Mitchell (2006). "Nucleotide excision repair disorders and the balance between cancer and aging." Cell Cycle **5**(24): 2886-8.

Arima, Y., C. Nishigori, T. Takeuchi, S. Oka, K. Morimoto, A. Utani and Y. Miyachi (2006). "4-Nitroquinoline 1-oxide forms 8-hydroxydeoxyguanosine in human fibroblasts through reactive oxygen species." Toxicol Sci **91**(2): 382-92.

Ariza, A., D. J. Richard, M. F. White and C. S. Bond (2005). "Conformational flexibility revealed by the crystal structure of a crenarchaeal RadA." Nucleic Acids Res **33**(5): 1465-73.

Assenmacher, N., K. Wenig, A. Lammens and K. P. Hopfner (2006). "Structural basis for transcription-coupled repair: the N terminus of Mfd resembles UvrB with degenerate ATPase motifs." J Mol Biol **355**(4): 675-83.

Atkinson, J., C. P. Guy, C. J. Cadman, G. F. Moolenaar, N. Goosen and P. McGlynn (2009). "Stimulation of UvrD helicase by UvrAB." J Biol Chem **284**(14): 9612-23.

Bachrati, C. Z. and I. D. Hickson (2008). "RecQ helicases: guardian angels of the DNA replication fork." Chromosoma **117**(3): 219-33.

Banham, J. E., C. M. Baker, S. Ceola, I. J. Day, G. H. Grant, E. J. J. Groenen, C. T. Rodgers, G. Jeschke and C. R. Timmel (2008). "Distance measurements in the borderline region of applicability of CW EPR and DEER: A model study on a homologous series of spin-labelled peptides." Journal of Magnetic Resonance **191**: 202-218.

Barns, S. M., C. F. Delwiche, J. D. Palmer and N. R. Pace (1996). "Perspectives on archaeal diversity, thermophily and monophyly from environmental rRNA sequences." Proc Natl Acad Sci U S A **93**(17): 9188-93.

Basile, G., M. Aker and R. K. Mortimer (1992). "Nucleotide sequence and transcriptional regulation of the yeast recombinational repair gene RAD51." Mol Cell Biol **12**(7): 3235-46.

Bell, S. D. and S. P. Jackson (1998). "Transcription and translation in Archaea: a mosaic of eukaryal and bacterial features." Trends Microbiol **6**(6): 222-8.

Beranek, D. T. (1990). "Distribution of methyl and ethyl adducts following alkylation with monofunctional alkylating agents." Mutat Res **231**(1): 11-30.

Berneburg, M., J. E. Lowe, T. Nardo, S. Araujo, M. I. Fousteri, M. H. Green, J. Krutmann, R. D. Wood, M. Stefanini and A. R. Lehmann (2000). "UV damage causes uncontrolled DNA breakage in cells from patients with combined features of XP-D and Cockayne syndrome." EMBO J **19**(5): 1157-66.

Bhaskara, V., A. Dupre, B. Lengsfeld, B. B. Hopkins, A. Chan, J. H. Lee, X. Zhang, J. Gautier, V. Zakian and T. T. Paull (2007). "Rad50 adenylate kinase activity regulates DNA tethering by Mre11/Rad50 complexes." Mol Cell **25**(5): 647-61.

Bianco, P. R. and S. C. Kowalczykowski (2000). "Translocation step size and mechanism of the RecBC DNA helicase." Nature **405**(6784): 368-72.

Bjornson, K. P., I. Wong and T. M. Lohman (1996). "ATP hydrolysis stimulates binding and release of single stranded DNA from alternating subunits of the dimeric E. coli Rep helicase: implications for ATP-driven helicase translocation." J Mol Biol **263**(3): 411-22.

Boal, A. K., E. Yavin, O. A. Lukianova, V. L. O'Shea, S. S. David and J. K. Barton (2005). "DNA-bound redox activity of DNA repair glycosylases containing [4Fe-4S] clusters." Biochemistry **44**(23): 8397-407.

Bohr, V. A. (2008). "Rising from the RecQ-age: the role of human RecQ helicases in genome maintenance." Trends Biochem Sci **33**(12): 609-20.

Bootsma, D. and J. H. Hoeijmakers (1991). "The genetic basis of xeroderma pigmentosum." Ann Genet **34**(3-4): 143-50.

Boubakri, H., A. Langlois de Septenville, E. Viguera and B. Michel (2009). "The helicases DinG, Rep and UvrD cooperate to promote replication across transcription units *in vivo*." EMBO J: advanced online publication.

Boucher, Y., C. J. Douady, R. T. Papke, D. A. Walsh, M. E. Boudreau, C. L. Nesbo, R. J. Case and W. F. Doolittle (2003). "Lateral gene transfer and the origins of prokaryotic groups." Annu Rev Genet **37**: 283-328.

Boyd, J. B., K. Sakaguchi and P. V. Harris (1990). "mus308 mutants of Drosophila exhibit hypersensitivity to DNA cross-linking agents and are defective in a deoxyribonuclease." Genetics **125**(4): 813-9.

Brochier, C., S. Gribaldo, Y. Zivanovic, F. Confalonieri and P. Forterre (2005). "Nanoarchaea: representatives of a novel archaeal phylum or a fast-evolving euryarchaeal lineage related to Thermococcales?" Genome Biol **6**(5): R42.

Broughton, B. C., A. F. Thompson, S. A. Harcourt, W. Vermeulen, J. H. Hoeijmakers, E. Botta, M. Stefanini, M. D. King, C. A. Weber, J. Cole and et al. (1995). "Molecular and cellular analysis of the DNA repair defect in a patient in xeroderma pigmentosum complementation group D who has the clinical features of xeroderma pigmentosum and Cockayne syndrome." Am J Hum Genet **56**(1): 167-74.

Bugreev, D. V. and A. V. Mazin (2004). "Ca2+ activates human homologous recombination protein Rad51 by modulating its ATPase activity." Proc Natl Acad Sci U S A **101**(27): 9988-93.

Bukowy, Z., J. A. Harrigan, D. A. Ramsden, B. Tudek, V. A. Bohr and T. Stevnsner (2008). "WRN Exonuclease activity is blocked by specific oxidatively induced

base lesions positioned in either DNA strand." Nucleic Acids Res **36**(15): 4975-87.

Busso, D., A. Keriel, B. Sandrock, A. Poterszman, O. Gileadi and J. M. Egly (2000). "Distinct regions of MAT1 regulate cdk7 kinase and TFIIH transcription activities." J Biol Chem **275**(30): 22815-23.

Buttner, K., S. Nehring and K. P. Hopfner (2007). "Structural basis for DNA duplex separation by a superfamily-2 helicase." Nat Struct Mol Biol **14**(7): 647-52.

Byrd, A. K. and K. D. Raney (2004). "Protein displacement by an assembly of helicase molecules aligned along single-stranded DNA." Nat Struct Mol Biol **11**(6): 531-8.

Byrd, A. K. and K. D. Raney (2006). "Displacement of a DNA binding protein by Dda helicase." Nucleic Acids Res **34**(10): 3020-9.

Cadet, J., E. Sage and T. Douki (2005). "Ultraviolet radiation-mediated damage to cellular DNA." Mutat Res **571**(1-2): 3-17.

Cantor, S., R. Drapkin, F. Zhang, Y. Lin, J. Han, S. Pamidi and D. M. Livingston (2004). "The BRCA1-associated protein BACH1 is a DNA helicase targeted by clinically relevant inactivating mutations." Proc Natl Acad Sci U S A **101**(8): 2357-62.

Carlson, M. and B. C. Laurent (1994). "The SNF/SWI family of global transcriptional activators." Curr Opin Cell Biol **6**(3): 396-402.

Chen, L., K. Brugger, M. Skovgaard, P. Redder, Q. She, E. Torarinsson, B. Greve, M. Awayez, A. Zibat, H. P. Klenk and R. A. Garrett (2005). "The genome of Sulfolobus acidocaldarius, a model organism of the Crenarchaeota." J Bacteriol **187**(14): 4992-9.

Chen, L. T., T. P. Ko, Y. C. Chang, K. A. Lin, C. S. Chang, A. H. Wang and T. F. Wang (2007a). "Crystal structure of the left-handed archaeal RadA helical filament: identification of a functional motif for controlling quaternary structures and enzymatic functions of RecA family proteins." Nucleic Acids Res **35**(6): 1787-801.

Chen, L. T., T. P. Ko, Y. W. Chang, K. A. Lin, A. H. Wang and T. F. Wang (2007b). "Structural and functional analyses of five conserved positively charged residues in the L1 and N-terminal DNA binding motifs of archaeal RADA protein." PLoS One **2**(9): e858.

Chepanoske, C. L., M. P. Golinelli, S. D. Williams and S. S. David (2000). "Positively charged residues within the iron-sulfur cluster loop of E. coli MutY participate in damage recognition and removal." Arch Biochem Biophys **380**(1): 11-9.

Coin, F., V. Oksenych and J. M. Egly (2007). "Distinct roles for the XPB/p52 and XPD/p44 subcomplexes of TFIIH in damaged DNA opening during nucleotide excision repair." Mol Cell **26**(2): 245-56.

Coin, F., L. Proietti De Santis, T. Nardo, O. Zlobinskaya, M. Stefanini and J. M. Egly (2006). "p8/TTD-A as a repair-specific TFIIH subunit." Mol Cell **21**(2): 215-26.

Coin, F., J.C. Marinoni, C. Rudolfo, S. Fribourg, A. M. Pedrini and J.M. Egly (1998). "Mutations in the XPD helicase gene result in XP and TTD phenotypes, preventing interaction between XPD and the p44 subunit of TFIIH. Nat Genet **20**(2): 184-8.

Coldren, C. D., H. W. Hellinga and J. P. Caradonna (1997). "The rational design and construction of a cuboidal iron-sulfur protein." Proc Natl Acad Sci U S A **94**(13): 6635-40.

Connolly, B., C. A. Parsons, F. E. Benson, H. J. Dunderdale, G. J. Sharples, R. G. Lloyd and S. C. West (1991). "Resolution of Holliday junctions in vitro requires

the Escherichia coli ruvC gene product." Proc Natl Acad Sci U S A **88**(14): 6063-7.

Connolly, B. and S. C. West (1990). "Genetic recombination in Escherichia coli: Holliday junctions made by RecA protein are resolved by fractionated cell-free extracts." Proc Natl Acad Sci U S A **87**(21): 8476-80.

Constantinesco, F., P. Forterre, E. V. Koonin, L. Aravind and C. Elie (2004). "A bipolar DNA helicase gene, herA, clusters with rad50, mre11 and nurA genes in thermophilic archaea." Nucleic Acids Res **32**(4): 1439-47.

Conway, A. B., T. W. Lynch, Y. Zhang, G. S. Fortin, C. W. Fung, L. S. Symington and P. A. Rice (2004). "Crystal structure of a Rad51 filament." Nat Struct Mol Biol **11**(8): 791-6.

Costa, R. M., V. Chigancas, S. Galhardo Rda, H. Carvalho and C. F. Menck (2003). "The eukaryotic nucleotide excision repair pathway." Biochimie **85**(11): 1083-99.

Cox, M. M. (2003). "The bacterial RecA protein as a motor protein." Annu Rev Microbiol **57**: 551-77.

Cox, M. M. and I. R. Lehman (1982). "recA protein-promoted DNA strand exchange. Stable complexes of recA protein and single-stranded DNA formed in the presence of ATP and single-stranded DNA binding protein." J Biol Chem **257**(14): 8523-32.

Cubeddu, L. and M. F. White (2005). "DNA damage detection by an archaeal single-stranded DNA-binding protein." J Mol Biol **353**(3): 507-16.

Dawid, A., V. Croquette, M. Grigoriev and F. Heslot (2004). "Single-molecule study of RuvAB-mediated Holliday-junction migration." Proc Natl Acad Sci U S A **101**(32): 11611-6.

De Felice, M., V. Aria, L. Esposito, M. De Falco, B. Pucci, M. Rossi and F. M. Pisani (2007). "A novel DNA helicase with strand-annealing activity from the crenarchaeon Sulfolobus solfataricus." Biochem J **408**(1): 87-95.

de Laat, W. L., N. G. Jaspers and J. H. Hoeijmakers (1999). "Molecular mechanism of nucleotide excision repair." Genes Dev **13**(7): 768-85.

Dillingham, M. S. and S. C. Kowalczykowski (2008). "RecBCD enzyme and the repair of double-stranded DNA breaks." Microbiol Mol Biol Rev **72**(4): 642-71, Table of Contents.

Dillingham, M. S., P. Soultanas and D. B. Wigley (1999). "Site-directed mutagenesis of motif III in PcrA helicase reveals a role in coupling ATP hydrolysis to strand separation." Nucleic Acids Res **27**(16): 3310-7.

Dorazi, R., J. L. Parker and M. F. White (2006). "PCNA activates the Holliday junction endonuclease Hjc." J Mol Biol **364**(3): 243-7.

Dubaele, S., L. Proietti De Santis, R. J. Bienstock, A. Keriel, M. Stefanini, B. Van Houten and J. M. Egly (2003). "Basal transcription defect discriminates between xeroderma pigmentosum and trichothiodystrophy in XPD patients." Mol Cell **11**(6): 1635-46.

Dudas, A. and M. Chovanec (2004). "DNA double-strand break repair by homologous recombination." Mutat Res **566**(2): 131-67.

Eady, R. R. (1988). "The vanadium-containing nitrogenase of Azotobacter." Biofactors **1**(2): 111-6.

Egelman, E. H., X. Yu, R. Wild, M. M. Hingorani and S. S. Patel (1995). "Bacteriophage T7 helicase/primase proteins form rings around single-stranded DNA that suggest a general structure for hexameric helicases." Proc Natl Acad Sci U S A **92**(9): 3869-73.

<center>REFERENCES</center>

Enemark, E. J. and L. Joshua-Tor (2006). "Mechanism of DNA translocation in a replicative hexameric helicase." <u>Nature</u> **442**(7100): 270-5.

Enemark, E. J. and L. Joshua-Tor (2008). "On helicases and other motor proteins." <u>Curr Opin Struct Biol</u> **18**(2): 243-57.

Esashi, F., V. E. Galkin, X. Yu, E. H. Egelman and S. C. West (2007). "Stabilization of RAD51 nucleoprotein filaments by the C-terminal region of BRCA2." <u>Nat Struct Mol Biol</u> **14**(6): 468-74.

Evans, E., J. G. Moggs, J. R. Hwang, J. M. Egly and R. D. Wood (1997). "Mechanism of open complex and dual incision formation by human nucleotide excision repair factors." <u>EMBO J</u> **16**(21): 6559-73.

Fan, L., A. S. Arvai, P. K. Cooper, S. Iwai, F. Hanaoka and J. A. Tainer (2006). "Conserved XPB core structure and motifs for DNA unwinding: implications for pathway selection of transcription or excision repair." <u>Mol Cell</u> **22**(1): 27-37.

Fan, L., J. O. Fuss, Q. J. Cheng, A. S. Arvai, M. Hammel, V. A. Roberts, P. K. Cooper and J. A. Tainer (2008). "XPD helicase structures and activities: insights into the cancer and aging phenotypes from XPD mutations." <u>Cell</u> **133**(5): 789-800.

Feaver, W. J., W. Huang, O. Gileadi, L. Myers, C. M. Gustafsson, R. D. Kornberg and E. C. Friedberg (2000). "Subunit interactions in yeast transcription/repair factor TFIIH. Requirement for Tfb3 subunit in nucleotide excision repair." <u>J Biol Chem</u> **275**(8): 5941-6.

Fernandez de Henestrosa, A. R., S. Calero and J. Barbe (1991). "Expression of the recA gene of Escherichia coli in several species of gram-negative bacteria." <u>Mol Gen Genet</u> **226**(3): 503-6.

Forget, A. L., M. M. Kudron, D. A. McGrew, M. A. Calmann, C. A. Schiffer and K. L. Knight (2006). "RecA dimers serve as a functional unit for assembly of active nucleoprotein filaments." <u>Biochemistry</u> **45**(45): 13537-42.

Forterre, P., S. Gribaldo and C. Brochier-Armanet (2009). "Happy together: genomic insights into the unique Nanoarchaeum/Ignicoccus association." <u>J Biol</u> **8**(1): 7.

Fousteri, M., W. Vermeulen, A. A. van Zeeland and L. H. Mullenders (2006). "Cockayne syndrome A and B proteins differentially regulate recruitment of chromatin remodeling and repair factors to stalled RNA polymerase II in vivo." <u>Mol Cell</u> **23**(4): 471-82.

Frols, S., P. M. Gordon, M. A. Panlilio, I. G. Duggin, S. D. Bell, C. W. Sensen and C. Schleper (2007). "Response of the hyperthermophilic archaeon Sulfolobus solfataricus to UV damage." <u>J Bacteriol</u> **189**(23): 8708-18.

Fujikane, R., H. Shinagawa and Y. Ishino (2006). "The archaeal Hjm helicase has recQ-like functions, and may be involved in repair of stalled replication fork." <u>Genes Cells</u> **11**(2): 99-110.

Garfinkel, D. J. and A. M. Bailis (2002). "Nucleotide Excision Repair, Genome Stability, and Human Disease: New Insight from Model Systems." <u>J Biomed Biotechnol</u> **2**(2): 55-60.

Gotz, D., S. Paytubi, S. Munro, M. Lundgren, R. Bernander and M. F. White (2007). "Responses of hyperthermophilic crenarchaea to UV irradiation." <u>Genome Biol</u> **8**(10): R220.

Groisman, R., J. Polanowska, I. Kuraoka, J. Sawada, M. Saijo, R. Drapkin, A. F. Kisselev, K. Tanaka and Y. Nakatani (2003). "The ubiquitin ligase activity in the DDB2 and CSA complexes is differentially regulated by the COP9 signalosome in response to DNA damage." <u>Cell</u> **113**(3): 357-67.

Gunnar, J. (2002). " Distance measurements in the nanometer range by pulse EPR." <u>ChemPhysChem</u> **3**(11): 927-32.

<center>264</center>

Guy, C. P. and E. L. Bolt (2005). "Archaeal Hel308 helicase targets replication forks in vivo and in vitro and unwinds lagging strands." Nucleic Acids Res **33**(11): 3678-90.

Guy, C. P., S. Haldenby, A. Brindley, D. A. Walsh, G. S. Briggs, M. J. Warren, T. Allers and E. L. Bolt (2006). "Interactions of RadB, a DNA repair protein in archaea, with DNA and ATP." J Mol Biol **358**(1): 46-56.

Guy, C. P., A. I. Majernik, J. P. Chong and E. L. Bolt (2004). "A novel nuclease-ATPase (Nar71) from archaea is part of a proposed thermophilic DNA repair system." Nucleic Acids Res **32**(21): 6176-86.

Ha, T. (2001). "Single-molecule fluorescence resonance energy transfer." Methods **25**(1): 78-86.

Hagensee, M. E. and R. E. Moses (1989). "Multiple pathways for repair of hydrogen peroxide-induced DNA damage in Escherichia coli." J Bacteriol **171**(2): 991-5.

Haldenby, S., M. F. White and T. Allers (2009). "RecA family proteins in archaea: RadA and its cousins." Biochem Soc Trans **37**(Pt 1): 102-7.

Hanawalt, P. C. and G. Spivak (2008). "Transcription-coupled DNA repair: two decades of progress and surprises." Nat Rev Mol Cell Biol **9**(12): 958-70.

Harris, P. V., O. M. Mazina, E. A. Leonhardt, R. B. Case, J. B. Boyd and K. C. Burtis (1996). "Molecular cloning of Drosophila mus308, a gene involved in DNA cross-link repair with homology to prokaryotic DNA polymerase I genes." Mol Cell Biol **16**(10): 5764-71.

Haseltine, C. A. and S. C. Kowalczykowski (2009). "An archaeal Rad54 protein remodels DNA and stimulates DNA strand exchange by RadA." Nucleic Acids Res **37**(8): 2757-70.

Heller, R. C. and K. J. Marians (2006a). "Replication fork reactivation downstream of a blocked nascent leading strand." Nature **439**(7076): 557-62.

Heller, R. C. and K. J. Marians (2006b). "Replisome assembly and the direct restart of stalled replication forks." Nat Rev Mol Cell Biol **7**(12): 932-43.

Heyer, W. D., X. Li, M. Rolfsmeier and X. P. Zhang (2006). "Rad54: the Swiss Army knife of homologous recombination?" Nucleic Acids Res **34**(15): 4115-25.

Hishida, T., Y. W. Han, T. Shibata, Y. Kubota, Y. Ishino, H. Iwasaki and H. Shinagawa (2004). "Role of the Escherichia coli RecQ DNA helicase in SOS signaling and genome stabilization at stalled replication forks." Genes Dev **18**(15): 1886-97.

Hoeijmakers, J. H. (2001). "Genome maintenance mechanisms for preventing cancer." Nature **411**(6835): 366-74.

Hopfner, K. P., A. Karcher, L. Craig, T. T. Woo, J. P. Carney and J. A. Tainer (2001). "Structural biochemistry and interaction architecture of the DNA double-strand break repair Mre11 nuclease and Rad50-ATPase." Cell **105**(4): 473-85.

Howard, J. (1996). "The movement of kinesin along microtubules." Annu Rev Physiol **58**: 703-29.

Huber, H., M. J. Hohn, R. Rachel, T. Fuchs, V. C. Wimmer and K. O. Stetter (2002). "A new phylum of Archaea represented by a nanosized hyperthermophilic symbiont." Nature **417**(6884): 63-7.

Hutton, R. D., J. A. Roberts, J. C. Penedo and M. F. White (2008). "PCNA stimulates catalysis by structure-specific nucleases using two distinct mechanisms: substrate targeting and catalytic step." Nucleic Acids Res **36**(21): 6720-7.

Ip, S. C., U. Rass, M. G. Blanco, H. R. Flynn, J. M. Skehel and S. C. West (2008). "Identification of Holliday junction resolvases from humans and yeast." Nature **456**(7220): 357-61.

Iyer, N., M. S. Reagan, K. J. Wu, B. Canagarajah and E. C. Friedberg (1996). "Interactions involving the human RNA polymerase II transcription/nucleotide excision repair complex TFIIH, the nucleotide excision repair protein XPG, and Cockayne syndrome group B (CSB) protein." Biochemistry **35**(7): 2157-67.

Jain, S. K., M. M. Cox and R. B. Inman (1994). "On the role of ATP hydrolysis in RecA protein-mediated DNA strand exchange. III. Unidirectional branch migration and extensive hybrid DNA formation." J Biol Chem **269**(32): 20653-61.

Jelinska, C., M. J. Conroy, C. J. Craven, A. M. Hounslow, P. A. Bullough, J. P. Waltho, G. L. Taylor and M. F. White (2005). "Obligate heterodimerization of the archaeal Alba2 protein with Alba1 provides a mechanism for control of DNA packaging." Structure **13**(7): 963-71.

Joo, C., S. A. McKinney, M. Nakamura, I. Rasnik, S. Myong and T. Ha (2006). "Real-time observation of RecA filament dynamics with single monomer resolution." Cell **126**(3): 515-27.

Keeney, S., G. J. Chang and S. Linn (1993). "Characterization of a human DNA damage binding protein implicated in xeroderma pigmentosum E." J Biol Chem **268**(28): 21293-300.

Kelman, Z. and M. F. White (2005). "Archaeal DNA replication and repair." Curr Opin Microbiol **8**(6): 669-76.

Kernchen, U. and G. Lipps (2006). "Thermodynamic analysis of the single-stranded DNA binding activity of the archaeal replication protein A (RPA) from Sulfolobus solfataricus." Biochemistry **45**(2): 594-603.

Kim, J. L., K. A. Morgenstern, J. P. Griffith, M. D. Dwyer, J. A. Thomson, M. A. Murcko, C. Lin and P. R. Caron (1998). "Hepatitis C virus NS3 RNA helicase domain with a bound oligonucleotide: the crystal structure provides insights into the mode of unwinding." Structure **6**(1): 89-100.

Kinebuchi, T., W. Kagawa, H. Kurumizaka and S. Yokoyama (2005). "Role of the N-terminal domain of the human DMC1 protein in octamer formation and DNA binding." J Biol Chem **280**(31): 28382-7.

Kirby, T. W., S. Harvey, E. F. DeRose, S. Chalov, A. K. Chikova, F. W. Perrino, R. M. Schaaper, R. E. London and L. C. Pedersen (2006). "Structure of the Escherichia coli DNA polymerase III epsilon-HOT proofreading complex." J Biol Chem **281**(50): 38466-71.

Kodadek, T. (1991). "Inhibition of protein-mediated homologous pairing by a DNA helicase." J Biol Chem **266**(15): 9712-8.

Kodadek, T. and B. M. Alberts (1987). "Stimulation of protein-directed strand exchange by a DNA helicase." Nature **326**(6110): 312-4.

Komori, K., T. Miyata, J. DiRuggiero, R. Holley-Shanks, I. Hayashi, I. K. Cann, K. Mayanagi, H. Shinagawa and Y. Ishino (2000a). "Both RadA and RadB are involved in homologous recombination in Pyrococcus furiosus." J Biol Chem **275**(43): 33782-90.

Komori, K., S. Sakae, R. Fujikane, K. Morikawa, H. Shinagawa and Y. Ishino (2000b). "Biochemical characterization of the hjc holliday junction resolvase of Pyrococcus furiosus." Nucleic Acids Res **28**(22): 4544-51.

Komori, K., S. Sakae, H. Shinagawa, K. Morikawa and Y. Ishino (1999). "A Holliday junction resolvase from Pyrococcus furiosus: functional similarity to Escherichia coli RuvC provides evidence for conserved mechanism of homologous recombination in Bacteria, Eukarya, and Archaea." Proc Natl Acad Sci U S A **96**(16): 8873-8.

Koonin, E. V., K. S. Makarova and L. Aravind (2001). "Horizontal gene transfer in prokaryotes: quantification and classification." Annu Rev Microbiol **55**: 709-42.

Korolev, S., J. Hsieh, G. H. Gauss, T. M. Lohman and G. Waksman (1997). "Major domain swiveling revealed by the crystal structures of complexes of E. coli Rep helicase bound to single-stranded DNA and ADP." Cell **90**(4): 635-47.

Kowalczykowski, S. C., D. A. Dixon, A. K. Eggleston, S. D. Lauder and W. M. Rehrauer (1994). "Biochemistry of homologous recombination in Escherichia coli." Microbiol Rev **58**(3): 401-65.

Kowalczykowski, S. C. and A. K. Eggleston (1994). "Homologous pairing and DNA strand-exchange proteins." Annu Rev Biochem **63**: 991-1043.

Krejci, L., S. Van Komen, Y. Li, J. Villemain, M. S. Reddy, H. Klein, T. Ellenberger and P. Sung (2003). "DNA helicase Srs2 disrupts the Rad51 presynaptic filament." Nature **423**(6937): 305-9.

Krissinel, E. and K. Henrick (2004). "Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions." Acta Crystallogr D Biol Crystallogr **60**(Pt 12 Pt 1): 2256-68.

Krissinel, E. and K. Henrick (2005). Detection of protein assemblies in crystals, Springer Berlin/Heidelberg.

Kvaratskhelia, M. and M. F. White (2000). "Two Holliday junction resolving enzymes in Sulfolobus solfataricus." J Mol Biol **297**(4): 923-32.

Lakowicz, J. R. (2006). Principles of Fluorescence Spectroscopy. Singapore, Springer Science and Business Media.

Legrand, M., C. L. Chan, P. A. Jauert and D. T. Kirkpatrick (2008). "Analysis of base excision and nucleotide excision repair in Candida albicans." Microbiology **154**(Pt 8): 2446-56.

Lehmann, A. R. (2001). "The xeroderma pigmentosum group D (XPD) gene: one gene, two functions, three diseases." Genes Dev **15**(1): 15-23.

Levran, O., C. Attwooll, R. T. Henry, K. L. Milton, K. Neveling, P. Rio, S. D. Batish, R. Kalb, E. Velleuer, S. Barral, J. Ott, J. Petrini, D. Schindler, H. Hanenberg and A. D. Auerbach (2005). "The BRCA1-interacting helicase BRIP1 is deficient in Fanconi anemia." Nat Genet **37**(9): 931-3.

Lewis, L. K., M. E. Jenkins and D. W. Mount (1992). "Isolation of DNA damage-inducible promoters in Escherichia coli: regulation of polB (dinA), dinG, and dinH by LexA repressor." J Bacteriol **174**(10): 3377-85.

Lewis, S.M. (1971). "International committee for standardization in hematology: proposed recommendations for measurement of serum iron in human blood." Am J Clin Pathol **56**(4): 543-5.

Li, J., M. J. Poi, D. Qin, T. L. Selby, I. J. Byeon and M. D. Tsai (2000). "Tumor suppressor INK4: quantitative structure-function analyses of p18INK4C as an inhibitor of cyclin-dependent kinase 4." Biochemistry **39**(4): 649-57.

Li, X. and W. D. Heyer (2008). "Homologous recombination in DNA repair and DNA damage tolerance." Cell Res **18**(1): 99-113.

Li, Z., S. Lu, G. Hou, X. Ma, D. Sheng, J. Ni and Y. Shen (2008). "Hjm/Hel308A DNA helicase from Sulfolobus tokodaii promotes replication fork regression and interacts with Hjc endonuclease in vitro." J Bacteriol **190**(8): 3006-17.

Lin, J. J., A. M. Phillips, J. E. Hearst and A. Sancar (1992). "Active site of (A)BC excinuclease. II. Binding, bending, and catalysis mutants of UvrB reveal a direct role in 3' and an indirect role in 5' incision." J Biol Chem **267**(25): 17693-700.

Lin, J. J. and A. Sancar (1992). "Active site of (A)BC excinuclease. I. Evidence for 5' incision by UvrC through a catalytic site involving Asp399, Asp438, Asp466, and His538 residues." J Biol Chem **267**(25): 17688-92.

Lin, Z., H. Kong, M. Nei and H. Ma (2006). "Origins and evolution of the recA/RAD51 gene family: evidence for ancient gene duplication and endosymbiotic gene transfer." Proc Natl Acad Sci U S A **103**(27): 10328-33.

Lindahl, T. and R. D. Wood (1999). "Quality control by DNA repair." Science **286**(5446): 1897-905.

Liu, H., J. Rudolf, K. A. Johnson, S. A. McMahon, M. Oke, L. Carter, A. M. McRobbie, S. E. Brown, J. H. Naismith and M. F. White (2008). "Structure of the DNA repair helicase XPD." Cell **133**(5): 801-12.

Liu, N., D. Schild, M. P. Thelen and L. H. Thompson (2002). "Involvement of Rad51C in two distinct protein complexes of Rad51 paralogs in human cells." Nucleic Acids Res **30**(4): 1009-15.

Liu, Y., M. Tarsounas, P. O'Regan and S. C. West (2007). "Role of RAD51C and XRCC3 in genetic recombination and DNA repair." J Biol Chem **282**(3): 1973-9.

Logan, K. M., A. L. Forget, J. P. Verderese and K. L. Knight (2001). "ATP-mediated changes in cross-subunit interactions in the RecA protein." Biochemistry **40**(38): 11382-9.

Lohman, T. M. and K. P. Bjornson (1996). "Mechanisms of helicase-catalyzed DNA unwinding." Annu Rev Biochem **65**: 169-214.

Lovett, S. T. and R. D. Kolodner (1989). "Identification and purification of a single-stranded-DNA-specific exonuclease encoded by the recJ gene of Escherichia coli." Proc Natl Acad Sci U S A **86**(8): 2627-31.

Lundblad, J. R., M. Laurance and R. H. Goodman (1996). "Fluorescence polarization analysis of protein-DNA and protein-protein interactions." Mol Endocrinol **10**(6): 607-12.

Lusetti, S. L. and M. M. Cox (2002). "The bacterial RecA protein and the recombinational DNA repair of stalled replication forks." Annu Rev Biochem **71**: 71-100.

Mackintosh, S. G. and K. D. Raney (2006). "DNA unwinding and protein displacement by superfamily 1 and superfamily 2 helicases." Nucleic Acids Res **34**(15): 4154-9.

Malta, E., G. F. Moolenaar and N. Goosen (2006). "Base flipping in nucleotide excision repair." J Biol Chem **281**(4): 2184-94.

Malta, E., C. P. Verhagen, G. F. Moolenaar, D. V. Filippov, G. A. van der Marel and N. Goosen (2008). "Functions of base flipping in E. coli nucleotide excision repair." DNA Repair (Amst) **7**(10): 1647-58.

Marians, K. J. (2004). "Mechanisms of replication fork restart in Escherichia coli." Philos Trans R Soc Lond B Biol Sci **359**(1441): 71-7.

Masson, J. Y., M. C. Tarsounas, A. Z. Stasiak, A. Stasiak, R. Shah, M. J. McIlwraith, F. E. Benson and S. C. West (2001). "Identification and purification of two distinct complexes containing the five RAD51 paralogs." Genes Dev **15**(24): 3296-307.

McGlynn, P. and R. G. Lloyd (2002). "Recombinational repair and restart of damaged replication forks." Nat Rev Mol Cell Biol **3**(11): 859-70.

McIlwraith, M. J., D. R. Hall, A. Z. Stasiak, A. Stasiak, D. B. Wigley and S. C. West (2001). "RadA protein from Archaeoglobus fulgidus forms rings, nucleoprotein filaments and catalyses homologous recombination." Nucleic Acids Res **29**(22): 4509-17.

McRobbie, A. M., L. G. Carter, M. Kerou, H. Liu, S. A. McMahon, K. A. Johnson, M. Oke, J. H. Naismith and M. F. White (2009). "Structural and functional characterisation of a conserved archaeal RadA paralog with antirecombinase activity." J Mol Biol **389**(4): 661-73.

Mellon, I. (2005). "Transcription-coupled repair: a complex affair." Mutat Res **577**(1-2): 155-61.

Menetski, J. P., D. G. Bear and S. C. Kowalczykowski (1990). "Stable DNA heteroduplex formation catalyzed by the Escherichia coli RecA protein in the absence of ATP hydrolysis." Proc Natl Acad Sci U S A **87**(1): 21-5.

Middleton, C. L., J. L. Parker, D. J. Richard, M. F. White and C. S. Bond (2003). "Crystallization and preliminary X-ray diffraction studies of Hje, a HolliDay junction resolving enzyme from Sulfolobus solfataricus." Acta Crystallogr D Biol Crystallogr **59**(Pt 1): 171-3.

Mimitou, E. P. and L. S. Symington (2009). "Nucleases and helicases take center stage in homologous recombination." Trends Biochem Sci **34**(5): 264-72.

Min, J. H. and N. P. Pavletich (2007). "Recognition of DNA damage by the Rad4 nucleotide excision repair protein." Nature **449**(7162): 570-5.

Mitchell, D. L. and R. S. Nairn (1989). "The biology of the (6-4) photoproduct." Photochem Photobiol **49**(6): 805-19.

Morel, P., J. A. Hejna, S. D. Ehrlich and E. Cassuto (1993). "Antipairing and strand transferase activities of E. coli helicase II (UvrD)." Nucleic Acids Res **21**(14): 3205-9.

Morimatsu, K. and S. C. Kowalczykowski (2003). "RecFOR proteins load RecA protein onto gapped DNA to accelerate DNA strand exchange: a universal step of recombinational repair." Mol Cell **11**(5): 1337-47.

Morrical, S. W. and M. M. Cox (1990). "Stabilization of recA protein-ssDNA complexes by the single-stranded DNA binding protein of Escherichia coli." Biochemistry **29**(3): 837-43.

Mu, D., M. Wakasugi, D. S. Hsu and A. Sancar (1997). "Characterization of reaction intermediates of human excision repair nuclease." J Biol Chem **272**(46): 28971-9.

Muskavitch, K. M. and S. Linn (1982). "A unified mechanism for the nuclease and unwinding activities of the recBC enzyme of Escherichia coli." J Biol Chem **257**(5): 2641-8.

Myong, S., M. M. Bruno, A. M. Pyle and T. Ha (2007). "Spring-loaded mechanism of DNA unwinding by hepatitis C virus NS3 helicase." Science **317**(5837): 513-6.

Newman, M., J. Murray-Rust, J. Lally, J. Rudolf, A. Fadden, P. P. Knowles, M. F. White and N. Q. McDonald (2005). "Structure of an XPF endonuclease with and without DNA suggests a model for substrate recognition." EMBO J **24**(5): 895-905.

Nishi, R., S. Alekseev, C. Dinant, D. Hoogstraten, A. B. Houtsmuller, J. H. Hoeijmakers, W. Vermeulen, F. Hanaoka and K. Sugasawa (2009). "UV-DDB-dependent regulation of nucleotide excision repair kinetics in living cells." DNA Repair (Amst) **8**(6): 767-776.

Nishio, S. Y. and T. Itoh (2009). "Arginine-rich RNA binding domain and protein scaffold domain of RNase E are important for degradation of RNAI but not for that of the Rep mRNA of the ColE2 plasmid." Plasmid.

Ogrunc, M., D. F. Becker, S. W. Ragsdale and A. Sancar (1998). "Nucleotide excision repair in the third kingdom." J Bacteriol **180**(21): 5796-8.

Oh, E. Y. and L. Grossman (1986). "The effect of Escherichia coli Uvr protein binding on the topology of supercoiled DNA." Nucleic Acids Res **14**(21): 8557-71.

Orren, D. K., C. P. Selby, J. E. Hearst and A. Sancar (1992). "Post-incision steps of nucleotide excision repair in Escherichia coli. Disassembly of the UvrBC-DNA complex by helicase II and DNA polymerase I." J Biol Chem **267**(2): 780-8.

Oyama, T., H. Oka, K. Mayanagi, T. Shirai, K. Matoba, R. Fujikane, Y. Ishino and K. Morikawa (2009). "Atomic structures and functional implications of the archaeal RecQ-like helicase Hjm." BMC Struct Biol **9**: 2.

Park, J. S., M. T. Marr and J. W. Roberts (2002). "E. coli Transcription repair coupling factor (Mfd protein) rescues arrested complexes by promoting forward translocation." Cell **109**(6): 757-67.

Park, J. S. and J. W. Roberts (2006). "Role of DNA bubble rewinding in enzymatic transcription termination." Proc Natl Acad Sci U S A **103**(13): 4870-5.

Patel, S. S. and K. M. Picha (2000). "Structure and function of hexameric helicases." Annu Rev Biochem **69**: 651-97.

Pfaffl, M. W. (2001). "A new mathematical model for relative quantification in real-time RT-PCR." Nucleic Acids Res **29**(9): e45.

Pieroni, L., L. Khalil, F. Charlotte, T. Poynard, A. Piton, B. Hainque and F. Imbert-Bismut (2001). "Comparison of bathophenanthroline sulfonate and ferene as chromogens in colorimetric measurement of low hepatic iron concentration." Clin Chem **47**(11): 2059-61.

Plate, I., S. C. Hallwyl, I. Shi, L. Krejci, C. Muller, L. Albertsen, P. Sung and U. H. Mortensen (2008). "Interaction with RPA is necessary for Rad52 repair center formation and for its mediator activity." J Biol Chem **283**(43): 29077-85.

Ponticelli, A. S., D. W. Schultz, A. F. Taylor and G. R. Smith (1985). "Chi-dependent DNA strand cleavage by RecBC enzyme." Cell **41**(1): 145-51.

Prisner, T., M. Rohrer and F. MacMillan (2001). "Pulsed EPR spectroscopy: biological applications." Annu Rev Phys Chem **52**: 279-313.

Pugh, R. A., M. Honda, H. Leesley, A. Thomas, Y. Lin, M. J. Nilges, I. K. Cann and M. Spies (2008). "The iron-containing domain is essential in Rad3 helicases for coupling of ATP hydrolysis to DNA translocation and for targeting the helicase to the single-stranded DNA-double-stranded DNA junction." J Biol Chem **283**(3): 1732-43.

Pyle, A. M. (2008). "Translocation and unwinding mechanisms of RNA and DNA helicases." Annu Rev Biophys **37**: 317-36.

Quaiser, A., F. Constantinesco, M. F. White, P. Forterre and C. Elie (2008). "The Mre11 protein interacts with both Rad50 and the HerA bipolar helicase and is recruited to DNA following gamma irradiation in the archaeon Sulfolobus acidocaldarius." BMC Mol Biol **9**: 25.

Radonic, A., S. Thulke, I. M. Mackay, O. Landt, W. Siegert and A. Nitsche (2004). "Guideline to reference gene selection for quantitative real-time PCR." Biochem Biophys Res Commun **313**(4): 856-62.

Rasnik, I., S. Myong, W. Cheng, T. M. Lohman and T. Ha (2004). "DNA-binding orientation and domain conformation of the E. coli rep helicase monomer bound to a partial duplex junction: single-molecule studies of fluorescently labeled enzymes." J Mol Biol **336**(2): 395-408.

Reardon, J. T. and A. Sancar (2005). "Nucleotide excision repair." Prog Nucleic Acid Res Mol Biol **79**: 183-235.

REFERENCES

Register, J. C., 3rd, G. Christiansen and J. Griffith (1987). "Electron microscopic visualization of the RecA protein-mediated pairing and branch migration phases of DNA strand exchange." J Biol Chem **262**(26): 12812-20.

Reid, S. L., D. Parry, H. H. Liu and B. A. Connolly (2001). "Binding and recognition of GATATC target sequences by the EcoRV restriction endonuclease: a study using fluorescent oligonucleotides and fluorescence polarization." Biochemistry **40**(8): 2484-94.

Ren, B., X. Duan and H. Ding (2009). "Redox Control of the DNA Damage-inducible Protein DinG Helicase Activity via Its Iron-Sulfur Cluster." J Biol Chem **284**(8): 4829-4835.

Richards, J. D., L. Cubeddu, J. Roberts, H. Liu and M. F. White (2008a). "The archaeal XPB protein is a ssDNA-dependent ATPase with a novel partner." J Mol Biol **376**(3): 634-44.

Richards, J. D., K. A. Johnson, H. Liu, A. M. McRobbie, S. McMahon, M. Oke, L. Carter, J. H. Naismith and M. F. White (2008b). "Structure of the DNA repair helicase hel308 reveals DNA binding and autoinhibitory domains." J Biol Chem **283**(8): 5118-26.

Riedl, T., F. Hanaoka and J. M. Egly (2003). "The comings and goings of nucleotide excision repair factors on damaged DNA." EMBO J **22**(19): 5293-303.

Roberts, J. A., S. D. Bell and M. F. White (2003). "An archaeal XPF repair endonuclease dependent on a heterotrimeric PCNA." Mol Microbiol **48**(2): 361-71.

Roberts, J. A. and M. F. White (2005). "An archaeal endonuclease displays key properties of both eukaryal XPF-ERCC1 and Mus81." J Biol Chem **280**(7): 5924-8.

Robertson, R. B., D. N. Moses, Y. Kwon, P. Chan, W. Zhao, P. Chi, H. Klein, P. Sung and E. C. Greene (2009). "Visualizing the disassembly of S. cerevisiae Rad51 nucleoprotein filaments." J Mol Biol **388**(4): 703-20.

Ross, S. T., S. Schwartz, T. J. Fellers and M. W. Davidson. (2007). "Total Internal Reflection Fluorescence (TIRF) Microscopy." from http://www.microscopyu.com/articles/fluorescence/tirf/tirfintro.html.

Rudolf, J., V. Makrantoni, W. J. Ingledew, M. J. Stark and M. F. White (2006). "The DNA repair helicases XPD and FancJ have essential iron-sulfur domains." Mol Cell **23**(6): 801-8.

Salerno, V., A. Napoli, M. F. White, M. Rossi and M. Ciaramella (2003). "Transcriptional response to DNA damage in the archaeon Sulfolobus solfataricus." Nucleic Acids Res **31**(21): 6127-38.

Sancar, A. (1996). "DNA excision repair." Annu Rev Biochem **65**: 43-81.

Sancar, A. and W. D. Rupp (1983). "A novel repair enzyme: UVRABC excision nuclease of Escherichia coli cuts a DNA strand on both sides of the damaged region." Cell **33**(1): 249-60.

Sandler, S. J. and K. J. Marians (2000). "Role of PriA in replication fork reactivation in Escherichia coli." J Bacteriol **182**(1): 9-13.

Sarai, N., W. Kagawa, T. Kinebuchi, A. Kagawa, K. Tanaka, K. Miyagawa, S. Ikawa, T. Shibata, H. Kurumizaka and S. Yokoyama (2006). "Stimulation of Dmc1-mediated DNA strand exchange by the human Rad54B protein." Nucleic Acids Res **34**(16): 4429-37.

Scrima, A., R. Konickova, B. K. Czyzewski, Y. Kawasaki, P. D. Jeffrey, R. Groisman, Y. Nakatani, S. Iwai, N. P. Pavletich and N. H. Thoma (2008). "Structural basis

of UV DNA-damage recognition by the DDB1-DDB2 complex." <u>Cell</u> **135**(7): 1213-23.

Seitz, E. M., J. P. Brockman, S. J. Sandler, A. J. Clark and S. C. Kowalczykowski (1998). "RadA protein is an archaeal RecA protein homolog that catalyzes DNA strand exchange." <u>Genes Dev</u> **12**(9): 1248-53.

Seitz, E. M. and S. C. Kowalczykowski (2000). "The DNA binding and pairing preferences of the archaeal RadA protein demonstrate a universal characteristic of DNA strand exchange proteins." <u>Mol Microbiol</u> **37**(3): 555-60.

Seitz, E. M. and S. C. Kowalczykowski (2006). "Human Rad51 protein displays enhanced homologous pairing of DNA sequences resembling those at genetically unstable loci." <u>Nucleic Acids Res</u> **34**(10): 2847-52.

Selby, C. P. and A. Sancar (1993). "Molecular mechanism of transcription-repair coupling." <u>Science</u> **260**(5104): 53-8.

Selby, C. P. and A. Sancar (1995). "Structure and function of transcription-repair coupling factor. II. Catalytic properties." <u>J Biol Chem</u> **270**(9): 4890-5.

Selby, C. P. and A. Sancar (1997). "Human transcription-repair coupling factor CSB/ERCC6 is a DNA-stimulated ATPase but is not a helicase and does not disrupt the ternary transcription complex of stalled RNA polymerase II." <u>J Biol Chem</u> **272**(3): 1885-90.

Selvin, P. R., Ha, T. (2008). <u>Single-molecule techniques</u>, CSHL Press.

Seroz, T., C. Perez, E. Bergmann, J. Bradsher and J. M. Egly (2000). "p44/SSL1, the regulatory subunit of the XPD/RAD3 helicase, plays a crucial role in the transcriptional activity of TFIIH." <u>J Biol Chem</u> **275**(43): 33260-6.

She, Q., R. K. Singh, F. Confalonieri, Y. Zivanovic, G. Allard, M. J. Awayez, C. C. Chan-Weiher, I. G. Clausen, B. A. Curtis, A. De Moors, G. Erauso, C. Fletcher, P. M. Gordon, I. Heikamp-de Jong, A. C. Jeffries, C. J. Kozera, N. Medina, X. Peng, H. P. Thi-Ngoc, P. Redder, M. E. Schenk, C. Theriault, N. Tolstrup, R. L. Charlebois, W. F. Doolittle, M. Duguet, T. Gaasterland, R. A. Garrett, M. A. Ragan, C. W. Sensen and J. Van der Oost (2001). "The complete genome of the crenarchaeon Sulfolobus solfataricus P2." <u>Proc Natl Acad Sci U S A</u> **98**(14): 7835-40.

Sheng, D., M. Li, J. Jiao, J. Ni and Y. Shen (2008a). "Co-expression with RadA and the characterization of stRad55B, a RadA paralog from the hyperthermophilic crenarchaea Sulfolobus tokodaii." <u>Sci China C Life Sci</u> **51**(1): 60-5.

Sheng, D., S. Zhu, T. Wei, J. Ni and Y. Shen (2008b). "The in vitro activity of a Rad55 homologue from Sulfolobus tokodaii, a candidate mediator in RadA-catalyzed homologous recombination." <u>Extremophiles</u> **12**(1): 147-57.

Shin, D. S., C. Chahwan, J. L. Huffman and J. A. Tainer (2004). "Structure and function of the double-strand break repair machinery." <u>DNA Repair (Amst)</u> **3**(8-9): 863-73.

Shin, D. S., L. Pellegrini, D. S. Daniels, B. Yelent, L. Craig, D. Bates, D. S. Yu, M. K. Shivji, C. Hitomi, A. S. Arvai, N. Volkmann, H. Tsuruta, T. L. Blundell, A. R. Venkitaraman and J. A. Tainer (2003). "Full-length archaeal Rad51 structure and mutants: mechanisms for RAD51 assembly and control by BRCA2." <u>Embo J</u> **22**(17): 4566-76.

Shrivastav, M., L. P. De Haro and J. A. Nickoloff (2008). "Regulation of DNA double-strand break repair pathway choice." <u>Cell Res</u> **18**(1): 134-47.

Sigurdsson, S., S. Van Komen, W. Bussen, D. Schild, J. S. Albala and P. Sung (2001). "Mediator function of the human Rad51B-Rad51C complex in Rad51/RPA-catalyzed DNA strand exchange." <u>Genes Dev</u> **15**(24): 3308-18.

Sijbers, A. M., W. L. de Laat, R. R. Ariza, M. Biggerstaff, Y. F. Wei, J. G. Moggs, K. C. Carter, B. K. Shell, E. Evans, M. C. de Jong, S. Rademakers, J. de Rooij, N. G. Jaspers, J. H. Hoeijmakers and R. D. Wood (1996). "Xeroderma pigmentosum group F caused by a defect in a structure-specific DNA repair endonuclease." Cell **86**(5): 811-22.

Singleton, M. R., M. S. Dillingham, M. Gaudier, S. C. Kowalczykowski and D. B. Wigley (2004). "Crystal structure of RecBCD enzyme reveals a machine for processing DNA breaks." Nature **432**(7014): 187-93.

Singleton, M. R., M. S. Dillingham and D. B. Wigley (2007). "Structure and mechanism of helicases and nucleic acid translocases." Annu Rev Biochem **76**: 23-50.

Singleton, M. R., S. Scaife and D. B. Wigley (2001). "Structural analysis of DNA replication fork reversal by RecG." Cell **107**(1): 79-89.

Singleton, M. R. and D. B. Wigley (2002). "Modularity and specialization in superfamily 1 and 2 helicases." J Bacteriol **184**(7): 1819-26.

Skorvaga, M., K. Theis, B. S. Mandavilli, C. Kisker and B. Van Houten (2002). "The beta -hairpin motif of UvrB is essential for DNA binding, damage processing, and UvrC-mediated incisions." J Biol Chem **277**(2): 1553-9.

Sleigh, M. J. (1976). "The mechanism of DNA breakage by phleomycin in vitro." Nucleic Acids Res **3**(4): 891-901.

Smith, A. J. and N. J. Savery (2008). "Effects of the bacterial transcription-repair coupling factor during transcription of DNA containing non-bulky lesions." DNA Repair (Amst) **7**(10): 1670-9.

Smith, D. R., L. A. Doucette-Stamm, C. Deloughery, H. Lee, J. Dubois, T. Aldredge, R. Bashirzadeh, D. Blakely, R. Cook, K. Gilbert, D. Harrison, L. Hoang, P. Keagle, W. Lumm, B. Pothier, D. Qiu, R. Spadafora, R. Vicaire, Y. Wang, J. Wierzbowski, R. Gibson, N. Jiwani, A. Caruso, D. Bush, J. N. Reeve and et al. (1997). "Complete genome sequence of Methanobacterium thermoautotrophicum deltaH: functional analysis and comparative genomics." J Bacteriol **179**(22): 7135-55.

Sonoda, E., M. S. Sasaki, J. M. Buerstedde, O. Bezzubova, A. Shinohara, H. Ogawa, M. Takata, Y. Yamaguchi-Iwai and S. Takeda (1998). "Rad51-deficient vertebrate cells accumulate chromosomal breaks prior to cell death." Embo J **17**(2): 598-608.

Soultanas, P. and D. B. Wigley (2001). "Unwinding the 'Gordian knot' of helicase action." Trends Biochem Sci **26**(1): 47-54.

Staresincic, L., A. F. Fagbemi, J. H. Enzlin, A. M. Gourdin, N. Wijgers, I. Dunand-Sauthier, G. Giglia-Mari, S. G. Clarkson, W. Vermeulen and O. D. Scharer (2009). "Coordination of dual incision and repair synthesis in human nucleotide excision repair." EMBO J **28**(8): 1111-20.

Story, R. M., I. T. Weber and T. A. Steitz (1992). "The structure of the E. coli recA protein monomer and polymer." Nature **355**(6358): 318-25.

Sugasawa, K. and F. Hanaoka (2007). "Sensing of DNA damage by XPC/Rad4: one protein for many lesions." Nat Struct Mol Biol **14**(10): 887-8.

Sugiyama, T., N. Kantake, Y. Wu and S. C. Kowalczykowski (2006). "Rad52-mediated DNA annealing after Rad51-mediated DNA strand exchange promotes second ssDNA capture." EMBO J **25**(23): 5539-48.

Sugiyama, T. and S. C. Kowalczykowski (2002). "Rad52 protein associates with replication protein A (RPA)-single-stranded DNA to accelerate Rad51-mediated displacement of RPA and presynaptic complex formation." J Biol Chem **277**(35): 31663-72.

Sugiyama, T., J. H. New and S. C. Kowalczykowski (1998). "DNA annealing by RAD52 protein is stimulated by specific interaction with the complex of replication protein A and single-stranded DNA." Proc Natl Acad Sci U S A **95**(11): 6049-54.

Sung, P. (1997). "Yeast Rad55 and Rad57 proteins form a heterodimer that functions with replication protein A to promote DNA strand exchange by Rad51 recombinase." Genes Dev **11**(9): 1111-21.

Sung, P. and H. Klein (2006). "Mechanism of homologous recombination: mediators and helicases take on regulatory functions." Nat Rev Mol Cell Biol **7**(10): 739-50.

Sung, P., L. Krejci, S. Van Komen and M. G. Sehorn (2003). "Rad51 recombinase and recombination mediators." J Biol Chem **278**(44): 42729-32.

Sung, P., L. Prakash, S. W. Matson and S. Prakash (1987). "RAD3 protein of Saccharomyces cerevisiae is a DNA helicase." Proc Natl Acad Sci U S A **84**(24): 8951-5.

Swagemakers, S. M., J. Essers, J. de Wit, J. H. Hoeijmakers and R. Kanaar (1998). "The human RAD54 recombinational DNA repair protein is a double-stranded DNA-dependent ATPase." J Biol Chem **273**(43): 28292-7.

Symington, L. S. (2002). "Role of RAD52 epistasis group genes in homologous recombination and double-strand break repair." Microbiol Mol Biol Rev **66**(4): 630-70, table of contents.

Takata, M., M. S. Sasaki, E. Sonoda, T. Fukushima, C. Morrison, J. S. Albala, S. M. Swagemakers, R. Kanaar, L. H. Thompson and S. Takeda (2000). "The Rad51 paralog Rad51B promotes homologous recombinational repair." Mol Cell Biol **20**(17): 6476-82.

Tanaka, N. and B. Schwer (2006). "Mutations in PRP43 that uncouple RNA-dependent NTPase activity and pre-mRNA splicing function." Biochemistry **45**(20): 6510-21.

Taylor, A. and G. R. Smith (1980). "Unwinding and rewinding of DNA by the RecBC enzyme." Cell **22**(2 Pt 2): 447-57.

Thacker, J. (2005). "The RAD51 gene family, genetic instability and cancer." Cancer Lett **219**(2): 125-35.

Theis, K., P. J. Chen, M. Skorvaga, B. Van Houten and C. Kisker (1999). "Crystal structure of UvrB, a DNA helicase adapted for nucleotide excision repair." EMBO J **18**(24): 6899-907.

Theis, K., M. Skorvaga, M. Machius, N. Nakagawa, B. Van Houten and C. Kisker (2000). "The nucleotide excision repair protein UvrB, a helicase-like enzyme with a catch." Mutat Res **460**(3-4): 277-300.

Thoma, N. H., B. K. Czyzewski, A. A. Alexeev, A. V. Mazin, S. C. Kowalczykowski and N. P. Pavletich (2005). "Structure of the SWI2/SNF2 chromatin-remodeling domain of eukaryotic Rad54." Nat Struct Mol Biol **12**(4): 350-6.

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

Torres-Ramos, C. A., R. E. Johnson, L. Prakash and S. Prakash (2000). "Evidence for the involvement of nucleotide excision repair in the removal of abasic sites in yeast." Mol Cell Biol **20**(10): 3522-8.

Troelstra, C., A. van Gool, J. de Wit, W. Vermeulen, D. Bootsma and J. H. Hoeijmakers (1992). "ERCC6, a member of a subfamily of putative helicases, is involved in

Cockayne's syndrome and preferential repair of active genes." Cell **71**(6): 939-53.

Truglio, J. J., D. L. Croteau, B. Van Houten and C. Kisker (2006a). "Prokaryotic nucleotide excision repair: the UvrABC system." Chem Rev **106**(2): 233-52.

Truglio, J. J., E. Karakas, B. Rhau, H. Wang, M. J. DellaVecchia, B. Van Houten and C. Kisker (2006b). "Structural basis for DNA recognition and processing by UvrB." Nat Struct Mol Biol **13**(4): 360-4.

Tsaneva, I. R., B. Muller and S. C. West (1992). "ATP-dependent branch migration of Holliday junctions promoted by the RuvA and RuvB proteins of E. coli." Cell **69**(7): 1171-80.

Tsvetkov, Y. D., A. D. Milov and A. G. Maryasov (2008). "Pulsed electron-electron double resonance (PELDOR) as EPR spectroscopy in nanometre range." Russian Chemical Reviews **77**(6): 487-520.

Tuteja, N. and R. Tuteja (2004). "Unraveling DNA helicases. Motif, structure, mechanism and function." Eur J Biochem **271**(10): 1849-63.

Ueda, K. and T. Komano (1984). "Sequence-specific DNA damage induced by reduced mitomycin C and 7-N-(p-hydroxyphenyl)mitomycin C." Nucleic Acids Res **12**(17): 6673-83.

Umezu, K., N. W. Chi and R. D. Kolodner (1993). "Biochemical interaction of the Escherichia coli RecF, RecO, and RecR proteins with RecA protein and single-stranded DNA binding protein." Proc Natl Acad Sci U S A **90**(9): 3875-9.

Umezu, K., K. Nakayama and H. Nakayama (1990). "Escherichia coli RecQ protein is a DNA helicase." Proc Natl Acad Sci U S A **87**(14): 5363-7.

van der Heijden, T., M. Modesti, S. Hage, R. Kanaar, C. Wyman and C. Dekker (2008). "Homologous recombination in real time: DNA strand exchange by RecA." Mol Cell **30**(4): 530-8.

Van Dyk, T. K., E. J. DeRose and G. E. Gonye (2001). "LuxArray, a high-density, genomewide transcription analysis of Escherichia coli using bioluminescent reporter strains." J Bacteriol **183**(19): 5496-505.

van Hoffen, A., W. H. Kalle, A. de Jong-Versteeg, A. R. Lehmann, A. A. van Zeeland and L. H. Mullenders (1999). "Cells from XP-D and XP-D-CS patients exhibit equally inefficient repair of UV-induced damage in transcribed genes but different capacity to recover UV-inhibited transcription." Nucleic Acids Res **27**(14): 2898-904.

Van Houten, B. (1990). "Nucleotide excision repair in Escherichia coli." Microbiol Rev **54**(1): 18-51.

Van Houten, B., D. L. Croteau, M. J. DellaVecchia, H. Wang and C. Kisker (2005). "'Close-fitting sleeves': DNA damage recognition by the UvrABC nuclease system." Mutat Res **577**(1-2): 92-117.

Veaute, X., S. Delmas, M. Selva, J. Jeusset, E. Le Cam, I. Matic, F. Fabre and M. A. Petit (2005). "UvrD helicase, unlike Rep helicase, dismantles RecA nucleoprotein filaments in Escherichia coli." EMBO J **24**(1): 180-9.

Veaute, X., J. Jeusset, C. Soustelle, S. C. Kowalczykowski, E. Le Cam and F. Fabre (2003). "The Srs2 helicase prevents recombination by disrupting Rad51 nucleoprotein filaments." Nature **423**(6937): 309-12.

Velankar, S. S., P. Soultanas, M. S. Dillingham, H. S. Subramanya and D. B. Wigley (1999). "Crystal structures of complexes of PcrA DNA helicase with a DNA substrate indicate an inchworm mechanism." Cell **97**(1): 75-84.

Verhoeven, E. E., C. Wyman, G. F. Moolenaar and N. Goosen (2002). "The presence of two UvrB subunits in the UvrAB complex ensures damage detection in both DNA strands." EMBO J **21**(15): 4196-205.

Viswanathan, M. and S. T. Lovett (1999). "Exonuclease X of Escherichia coli. A novel 3'-5' DNase and Dnaq superfamily member involved in DNA repair." J Biol Chem **274**(42): 30094-100.

Voloshin, O. N. and R. D. Camerini-Otero (2007). "The DinG protein from Escherichia coli is a structure-specific helicase." J Biol Chem **282**(25): 18437-47.

Voloshin, O. N., F. Vanevski, P. P. Khil and R. D. Camerini-Otero (2003). "Characterization of the DNA damage-inducible helicase DinG from Escherichia coli." J Biol Chem **278**(30): 28284-93.

Wadsworth, R. I. and M. F. White (2001). "Identification and properties of the crenarchaeal single-stranded DNA binding protein from Sulfolobus solfataricus." Nucleic Acids Res **29**(4): 914-20.

Waters, E., M. J. Hohn, I. Ahel, D. E. Graham, M. D. Adams, M. Barnstead, K. Y. Beeson, L. Bibbs, R. Bolanos, M. Keller, K. Kretz, X. Lin, E. Mathur, J. Ni, M. Podar, T. Richardson, G. G. Sutton, M. Simon, D. Soll, K. O. Stetter, J. M. Short and M. Noordewier (2003). "The genome of Nanoarchaeum equitans: insights into early archaeal evolution and derived parasitism." Proc Natl Acad Sci U S A **100**(22): 12984-8.

Waters, T. R., J. Eryilmaz, S. Geddes and T. E. Barrett (2006). "Damage detection by the UvrABC pathway: crystal structure of UvrB bound to fluorescein-adducted DNA." FEBS Lett **580**(27): 6423-7.

Weber, C. A., E. P. Salazar, S. A. Stewart and L. H. Thompson (1988). "Molecular cloning and biological characterization of a human gene, ERCC2, that corrects the nucleotide excision repair defect in CHO UV5 cells." Mol Cell Biol **8**(3): 1137-46.

Weil, J. A., J. R. Bolton and J. E. Wertz (1994). "Electron paramagnetic resonance: Elementary theory and practical applications." John Wiley and Sons, New York, NY, U.S.A., 568 pp.

West, S. C. and B. Connolly (1992). "Biological roles of the Escherichia coli RuvA, RuvB and RuvC proteins revealed." Mol Microbiol **6**(19): 2755-9.

White, M. F. (2003). "Archaeal DNA repair: paradigms and puzzles." Biochem Soc Trans **31**(Pt 3): 690-3.

White, M. F. (2009). "Structure, function and evolution of the XPD family of iron-sulfur-containing 5'-->3' DNA helicases." Biochem Soc Trans **37**(Pt 3): 547-51.

Wickner, S., M. Wright and J. Hurwitz (1974). "Association of DNA-dependent and -independent ribonucleoside triphosphatase activities with dnaB gene product of Escherichia coli." Proc Natl Acad Sci U S A **71**(3): 783-7.

Williams, G. J., K. Johnson, J. Rudolf, S. A. McMahon, L. Carter, M. Oke, H. Liu, G. L. Taylor, M. F. White and J. H. Naismith (2006). "Structure of the heterotrimeric PCNA from Sulfolobus solfataricus." Acta Crystallogr Sect F Struct Biol Cryst Commun **62**(Pt 10): 944-8.

Williams, R. S., G. Moncalian, J. S. Williams, Y. Yamada, O. Limbo, D. S. Shin, L. M. Groocock, D. Cahill, C. Hitomi, G. Guenther, D. Moiani, J. P. Carney, P. Russell and J. A. Tainer (2008). "Mre11 dimers coordinate DNA end bridging and nuclease processing in double-strand-break repair." Cell **135**(1): 97-109.

Woese, C. R. and G. E. Fox (1977). "Phylogenetic structure of the prokaryotic domain: the primary kingdoms." Proc Natl Acad Sci U S A **74**(11): 5088-90.

Wollenberg, M., C. Berndt, E. Bill, J. D. Schwenn and A. Seidler (2003). "A dimer of the FeS cluster biosynthesis protein IscA from cyanobacteria binds a [2Fe2S] cluster between two protomers and transfers it to [2Fe2S] and [4Fe4S] apo proteins." Eur J Biochem **270**(8): 1662-71.

Wolski, S. C., J. Kuper, P. Hanzelmann, J. J. Truglio, D. L. Croteau, B. Van Houten and C. Kisker (2008). "Crystal structure of the FeS cluster-containing nucleotide excision repair helicase XPD." PLoS Biol **6**(6): e149.

Woodman, I. L., G. S. Briggs and E. L. Bolt (2007). "Archaeal Hel308 domain V couples DNA binding to ATP hydrolysis and positions DNA for unwinding over the helicase ratchet." J Mol Biol **374**(5): 1139-44.

Wyman, C., D. Ristic and R. Kanaar (2004). "Homologous recombination-mediated double-strand break repair." DNA Repair (Amst) **3**(8-9): 827-33.

Yang, Q. and E. Jankowsky (2006). "The DEAD-box protein Ded1 unwinds RNA duplexes by a mode distinct from translocating helicases." Nat Struct Mol Biol **13**(11): 981-6.

Yang, S., X. Yu, E. M. Seitz, S. C. Kowalczykowski and E. H. Egelman (2001). "Archaeal RadA protein binds DNA as both helical filaments and octameric rings." J Mol Biol **314**(5): 1077-85.

Yodh, J. G., B. C. Stevens, R. Kanagaraj, P. Janscak and T. Ha (2009). "BLM helicase measures DNA unwound before switching strands and hRPA promotes unwinding reinitiation." EMBO J **28**(4): 405-16.

Yu, X. and E. H. Egelman (1990). "Image analysis reveals that Escherichia coli RecA protein consists of two domains." Biophys J **57**(3): 555-66.

Yu, X. and E. H. Egelman (1997). "The RecA hexamer is a structural homologue of ring helicases." Nat Struct Biol **4**(2): 101-4.

Zhang, S., T. Wei, G. Hou, C. Zhang, P. Liang, J. Ni, D. Sheng and Y. Shen (2008). "Archaeal DNA helicase HerA interacts with Mre11 homologue and unwinds blunt-ended double-stranded DNA and recombination intermediates." DNA Repair (Amst) **7**(3): 380-91.

Zittel, M. C. and J. L. Keck (2005). "Coupling DNA-binding and ATP hydrolysis in Escherichia coli RecQ: role of a highly conserved aromatic-rich sequence." Nucleic Acids Res **33**(22): 6982-91.

A1: Oligonucleotide Sequences

Table A1.1: Cloning oligonucleotide primers

MUTANT	SEQUENCE (5'-3')
Sso2452-f (Gateway)	ATGGTAAGTAGATTATCTACTGGAATGAGAATTGTCACTTTTAAAAT
Sso2452-r (Gateway)	GAACTTTATTTTCTCAACTTTAGTTTTTCTGACTCCTCCTTAACTTC
DinG-f (Gateway)	CCGAAAACCTGTATTTTCAGGGCATGGCAACCTATGCCGTTGTGGATTT
DinG-r (Gateway)	GGGGACCACTTTGTACAAGAAAGCTGGGTCCTACTTTTTCTTTTTTTGAATTTGTC
Hel308-f (Gateway)	CCGAAAACCTGTATTTTCAGGGCATGAGTTTAGAATTAGAGTGGATGTC
Hel308-r (Gateway)	GGGGACCACTTTGTACAAGAAAGCTGGGTCCTAATGAAATCTATTAAGTAATCTTG

Table A1.2: Mutagenesis oligonucleotide primers

MUTANT	SEQUENCE (5'-3')
XPD R55A-f	CCCAATCTACAGAGCCTTAACTAAAATTCG
XPD R55A-r	CGAATTTTAGTTAAGGCTCTGTAGATTGGG
XPD D66A-f	GAGATTACGATTCAAATGTACATAAAACAAAGTAAAAG
XPD D66A-r	CTTTTACTTTGTTTTATGTACATTTGAATCGTAATCTC
XPD A204Y-f	CTCTAAGTATGATATGCAAAGTGACAATATGTGG
XPD A204Y-r	CCACATATTGTCACTTTGCATATCATACTTAGAG
XPD R373Q-f	CGTGATGTAAATGGCAAGTGTAATGTATGGC
XPD R373Q-r	GCCATACATTACACTTGCCATTTACATCACG
XPD D521G-f	GTATGGCTTTTGAACAAGCGTTTCGAGTC
XPD D521G-r	GACTCGAAACGCTTGTTCAAAAGCCATAC
XPD D529N-f	CTTTTGGACAAGCAGTTCGAGTCCCTTT
XPD D529N-r	AAAGGGACTCGAACTGCTTGTCCAAAAG
XPD R531Q-f	GTATTGTTCGTAGTCGCAACACATAATGAG
XPD R531Q-r	CTCATTATGTGTTGCGACTACGAACAATAC
HEL308	CTGATGTTGAATAAATCGAATCTAAATTAGG

418STOP-f	
HEL308 418STOP-r	CCTAATTTAGATTCGATTTATTCAACATCAG
HEL308 515STOP-f	GGTCTAGAGGGATAAAAGGCTTCATGTG
HEL308 515STOP-r	CACATGAAGCCTTTTATCCCTCTAGACC
HEL308 646STOP-f	GTGAGAGATGGTATATAGGAAGAGCTA
HEL308 646STOP-r	TAGCTCTTCCTATATACCATCTCTCAC
HEL308 K86C-f	CAAATGAAAAGTATTGTACATTTAAGGATTGG
HEL308 K86C-r	CCAATCCTTAAATGTACAATACTTTTCATTTG
HEL308 K182C-f	ACAATAAGTAACTATTGCCAAATAGCCAAATGG
HEL308 K182C-r	CCATTTGGCTATTTGGCAATAGTTACTTATTGT
HEL308 L284C-f	CTGAAATTTTATGTCAATTAGATGATATTG
HEL308 L284-r	CAATATCATCTAATTGACATAAAATTTCAG
HEL308 K409C-f	GTAGATAGAGTATTTTGCAAATATGTACTATCT
HEL308 K409C-r	AGATAGTACATATTTGCAAAATACTCTATCTAC
HEL308 S302C-f	AAAGAGCTACTAAAATGTCTAATAAGTAAAGGA
HEL308 S302C-r	TCCTTTACTTATTAGACATTTTAGTAGCTCTTT
HEL308 S595C-f	GAAGATACAATCTTGTGTAAATATAACATAGGT
HEL308 S595C –r	ACCTATGTTATATTTACACAAGATTGTATCTTC

DinG 10AD12A-f	GCCGTTGTGGCTTTGGCAACAACAGGCAACC
DinG 10AD12A-r	GGTTGCCTGTTGTTGCCAAAGCCACAACGGC
DinG WalkerA-f	CTAGGCAGTGGTGCATCATTAGCATATTTAC
DinG WalkerA-r	GTAAATATGCTAATGAATGCACCACTGCCTAG
DinG Δnuclease-f	CCGAAAACCTGTATTTTCAGGGCATGGTAGATCAACTTGGCTTA ACATA
DinG Δnuclease-r	GGGGACCACTTTGTACAAGAAAGCTGGGTCCTACTTTTTCTTTTTTTGA ATTTGTC

Table A1.3: Oligonucleotides for catalytic assays

NAME	SEQUENCE (5'-3')
B25	CCTCGAGGGATCCGTCCTAGCAAGC
B50	CCTCGAGGGATCCGTCCTAGCAAGCCGCTGCTACCGGAAGCTTCTGGACC
B50-Fl	CCTCGAGGGATCCGTCCTAGCAAGCCGC[Fl~dT]GCTACCGGAAGCTTTGGACC
B50comp	GGTCCAGAAGCTTCCGGTAGCAGCGGCTTGCTAGGACGGATCCCTCGAGG
B25	CCTCGAGGGATCCGTCCTAGCAAGC
H25	GGTCCAGAAGCTTCCGGTAGCAGCG
R26-50	TCTCAACTGCAGTCTAGACTCGAGC
X26-50	GCTTGCTAGGACGGATCCCTCGAGG
X50	GCTCGAGTCTAGACTGCAGTTGAGAGCTTGCTAGGACGGATCCCTCGAGG
Lesion Fl	CGGGATCGAGCACCAGAAT(Fl-dT)CACGAGTACCTGCGG
80mer	CGGGTGTCGGGGCTGGCTTAACTATGCGGCATCAGAGCAGATTGTACTGA
Dloop	GAGTGCACCATATGCGGTGTGAAATACCGC
Bubble7	GGTCCAGAAGCTTCCGGTAGCCTACCGCTGCTAGGACGGATCCCTCGAGG

Table A1.4: Oligonucleotides for translocation assays

NAME	SEQUENCE (5'-3')
B50bio2	CC-[bio-dT]-TCGAGGGATCCGTCCTAGCAAGCCGCTGCTACCGGAAGCTT CTGGACC
B50-3'bio	CCTCGAGGGATCCGTCCTAGCAAGCCGCTGCTACCGGAAGCTTCTGGAC

	C-[BioTEG]
B50-5'bio	[BioTEG]-CCTCGAGGGATCCGTCCTAGCAAGCCGCTGCTACCGGAAGCTT CTGGACC

Table A1.5: RT-PCR oligonucleotide primers

NAME	SEQUENCE (5'-3')
radA-f	AGCAGCTGGCATTCCATTAT
radA-r	GACCCAAACTCACCGAAGAA
sso0777-f	GGACTTCCGTTTTCATCTCG
sso0777-r	GGCGATCGACCTCAAAATAA
sso1861-f	GACCGGGAACTGGTAAATCA
sso1861-r	CTTCTCCCTTTGTGCGACAT
sso2452-f	TGTGGCAGATGGGATAATCA
sso2452-r	TGCTTATCGTGATCGGTTTG
sso0961-f	CAGCAACATAGTGGAGTTCTTG
sso0961-r	GGAAAGCGAAAATAGGGGAAATG

Table A1.6: Oligonucleotides for single-molecule FRET studies

NAME	SEQUENCE (5'-3')
50T	[MMT]-TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT-[BIO3S]
sm1	[MMT]-TGGCGGACGGTCGCGAGGTTTTTTTTTTTTTTTTTTTT-[BIO3S]
sm2	[MMT]-ACCTCGCGACCGTCGCCA
sm3	TGGCGGACGGTCGCGAGGTTTTTTTTTTTTTTTTTTTT-[BIO3S]

Table A1.7: Oligonucleotides for anisotropy substrates

NAME	SEQUENCE (5'-3')
45mer-Fl	CCGAATAGCGGAATTCACGAGTACCTGCGGCCTCGAGGGA[Fl-dT]CCGT
45mer comp	CCGAATAGCGGAATTCACGAGTACCTGCGGCCTCGAGGGATCCGT
15mer-Fl	[Fl-dT]-TCGGAGTACAGTGGG
15mer comp	CCCACTGTACTCCGA

Table A1.8: Oligonucleotide for analytical gel filtration and PELDOR spectroscopy

NAME	SEQUENCE (5'-3')
34 mer	CCTCGAGGGATCCGTCCTAGCAAGCCGCTGCTAC

Table A1.9: Oligonucleotides for TargeTron^{TM} Gene Knockout System

NAME	SEQUENCE (5'-3')
IBS	AAAAAAGCTTATAATTATCCTTAACAACCGGCAACGTGCGCCCAGATAG GGTG
EBS2	TGAACGCAAGTTTCTAATTTCGATTGTTGTTCGATAGAGGAAAGTGTCT
EBS1d	CAGATTGTACAAATGTGGTGATAACAGATAAGTCGGCAACCATAACTTA CCTTTCTTTGT
EBS universal	CGAAATTAGAAACTTGCGTTCAGTAAAC
DinG_f KO	GGATTTGGAAACAACAGGCA
DinG_r KO	CTGGTAAGCGATGTGCTTCA

A2: INTENSITY CHANGE ASSOCIATED WITH XPD ANISOTROPY DATA

For all anisotropy experiments performed with SacXPD, increasing concentrations of SacXPD (WT or mutant) were titrated into a buffered solution containing 20 nM fluorescein-labelled DNA (15 nucleotides) and both anisotropy and total fluorescence intensity were recorded. Over the protein concentration range used, a decrease in total fluorescence intensity of approximately 40% was observed. Figure A2.1 shows a representative fluorescent intensity scan recorded during an anisotropy experiment performed using wild-type SacXPD.
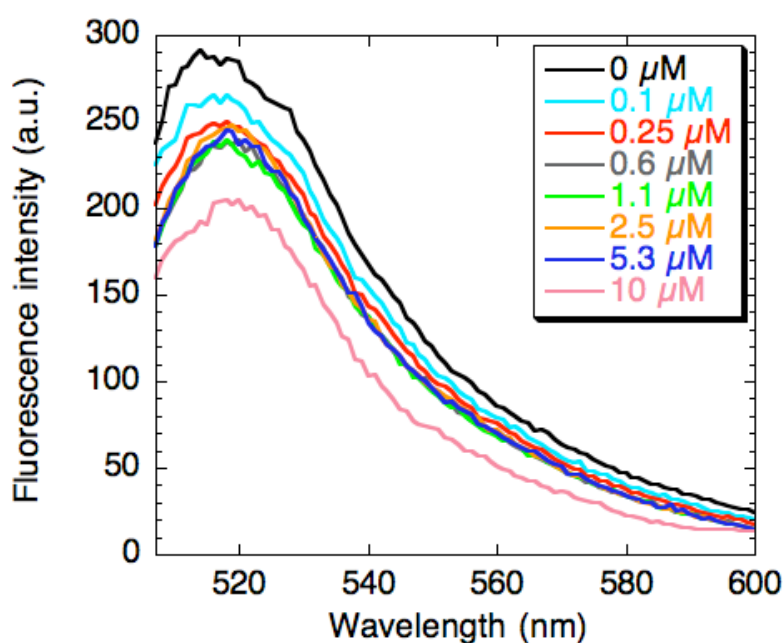

Figure A2.1: Total fluorescence intensity change during SacXPD anisotropy measurements

Wild-type SacXPD was titrated into a buffered (20 mM Hepes, pH 7.06) solution containing 20 nM fluorescein-labelled DNA (15 nucleotides) and total fluorescence intensity was recorded between 510 and 600 nm. Fluorescein has an emission maximum at 519 nm; as the concentration of XPD increases, the total fluorescence intensity of fluorescein can be seen to decrease by approximately 40%. This trace shows a representative intensity change observed for WT and all mutant XPD proteins studied.

A3: Intensity Change Associated with Sso2452 and SsoRadA Anisotropy Data

For all anisotropy experiments performed with Sso2452 and SsoRadA, increasing concentrations of protein were titrated into a buffered solution containing 20 nM fluorescein-labelled DNA (15 nucleotides or 15-base pairs) and both anisotropy and total fluorescence intensity were recorded. Over the protein concentration range used, a decrease in total fluorescence intensity of approximately 10% was observed (figure A3.1).
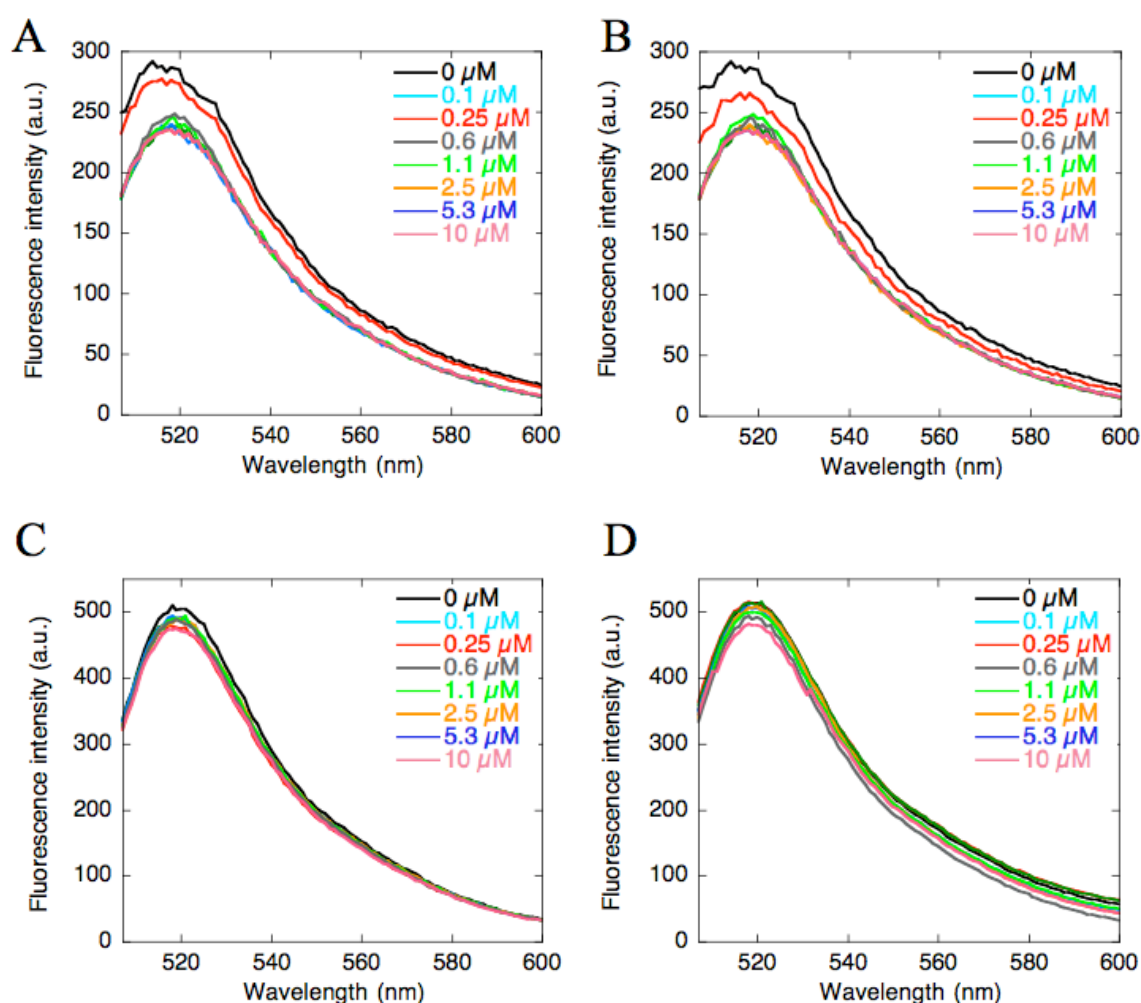

Figure A3.1: Total fluorescence intensity change during Sso2452 and SsoRadA anisotropy measurements

Total fluorescence intensity changes between 510 nm and 600 nm observed during the titration of Sso2452 or SsoRadA into a buffered (20 mM Hepes, pH 7.06) solution containing 20 nM fluorescein-labelled ssDNA (A, Sso2452; B, SsoRadA)) or dsDNA (C, Sso2452; D, SsoRadA). In all cases, a total fluorescence intensity change <10% can be observed around the emission maximum of fluorescein (519 nm).

A4: GROWTH MEDIA

A4.1 *SULFOLOBUS* MEDIA

For 1 L: 1.3 g $(NH_4)_2SO_4$; 0.28 g KH_2PO_4; 0.25 g $MgSO_4$; 0.07 g $CaCl_2$; 1 g yeast extract; 1 g tryptone; 1 mL trace elements.

pH 3 with H_2SO_4

Trace Elements

1.8 g $MnCl_2$; 4.5 g $Na_2B_4O_7$; 0.22 g $CuCl_2$; 0.03 g $NaMoO_4$; 0.03 g $VOSO_4$; 0.01 g $CoSO_4$.

Dissolve in 950 ml of water; adjust pH to 1.5 with H_2SO_4 and adjust volume to 1L before autoclaving. Inoculum should be mid logarithmic (between OD_{600} 0.35-1.0) and 5-15%. Incubate the cultures at 80 ºC and with vigorous shaking.

Storage

Liquid cultures can be stored up to 2 weeks on the bench, and still be used as inoculum; for long-term storage, however, the cells must be frozen and stored at -80 ºC. Harvest and resuspend mid-logarthimic cultures in 0.01 volume of *Sulfolobus* medium (pH 5.5) and 15 % (v/v) glycerol; freeze in liquid nitrogen.

A4.2 TSB MEDIA

For 1 L: 17 g tryptone, 3 g soya peptone, 5 g NaCl, 2.5 g K_2HPO_4, 2.5 g glucose.

Adjust to pH 7.3

A4.3 TSB AGAR

For 1 L: 15 g trypton, 5 g soya peptone, 5 g NaCl, 15 g agar.

Adjust to pH 7.3

A5: CALIBRATION OF SUPERDEX 200 10-300 ANALYTICAL GEL FILTRATION COLUMN

Proteins of known molecular weight were loaded onto a Superdex 200 10-300 analytical gel filtration column and their elution volumes (Ve) were recorded (table A3.1). The elution volume of blue dextran was considered to represent the void volume (Vo) of the column.

Table A5.1: Molecular weight markers and their respective elution volumes from the gel filtration column.

Protein marker	Molecular weight (kDa)	Log (MW)	Ve (ml)	Ve/ Vo
Blue dextran	2000	3.3	31.2	1
β-amylase	200	2.3	47.3	1.5
Alcohol dehydrogenase	150	2.2	50.2	1.6
Albumin	66	1.8	56.0	1.8
Carbonic anhydrase	29	1.5	64.5	6.1
Cytochrome C	12.4	1.1	69.5	2.2

The elution volume of each protein standard was divided by the void volume (equivalent to the elution volume (31.2 ml) of blue dextran) and plotted against the logarithm of its molecular weight.
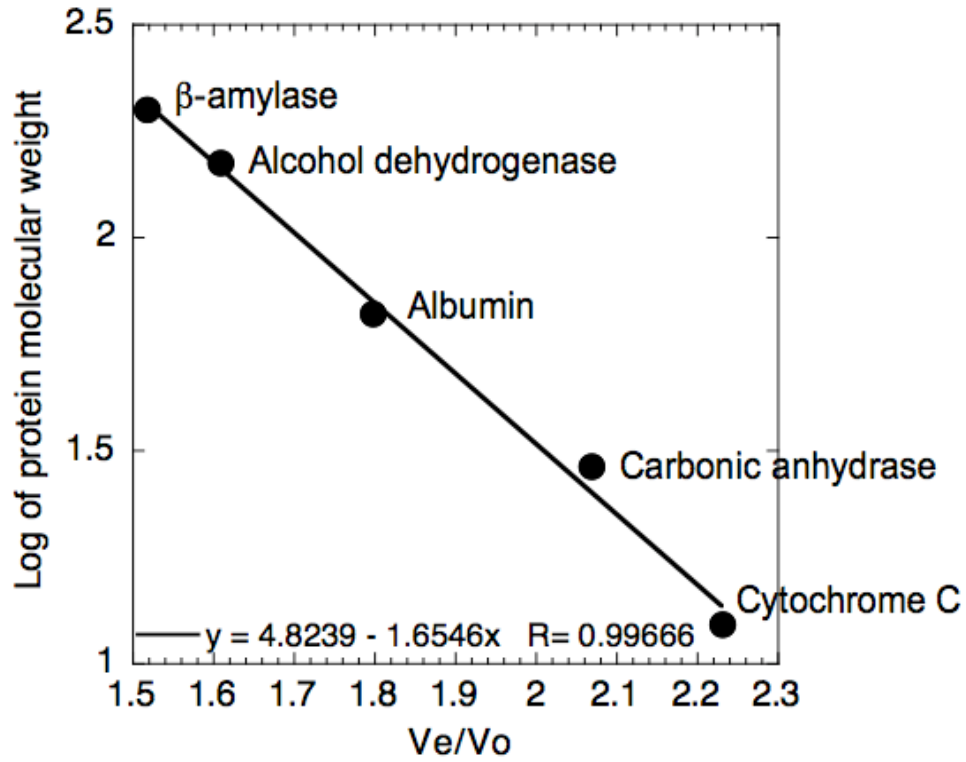
Figure A5.1: Standard curve for analytical gel filtration

The standard curve for analytical gel filtration has the following equation:
$y = 4.8239 = 1.6546x$ (R = 0.9966).

A6: PELDOR SPECTROSCOPY DATA

Table A6.1: Domain movements during Hel308 reaction cycle. Protein was incubated with MgCl2, ADP and AMP-PNP at a 1:20 ratio and with DNA at a 1:5 ratio. Increases in inter-spin distances are depicted by "+", while decreases are preceeded by "-". Significant distance increases are highlighted in blue. Errors around the distances have not been noted since distance data is generated by computer simulation rather than direct from the experiment.

MTSL site	Sample	Distance (Å)		Distance change compared to "protein only" (Å)	
		- MgCl$_2$	+ MgCl$_2$	- MgCl$_2$	+ MgCl$_2$
	Protein only	30.8	30.8	N/A	0
	+ AMP-PNP	30.8	31.0	0	+ 0.2
C182/C409	+ ADP	31.0	31.0	+ 0.2	+ 0.2
	+ DNA	31.0	N/A	+ 0.2	N/A
	Protein only	23.2	23.0	N/A	-0.2
	+ ADP	23.0	23.0	- 0.2	-0.2
C302/C595	+ AMP-PNP	23.0	23.0	- 0.2	-0.2
	+ DNA	23.0	N/A	- 0.2	N/A
	+ ADP/ DNA	N/A	23.0	N/A	- 0.2
	Protein only	28.6	28.6	N/A	0
	+ AMP-PNP	28.6	28.6	0	0
C86/C284	+ADP	N/A	27.4	N/A	- 1.2
	+ DNA	28.6	N/A	0	N/A
	+ADP/ DNA	N/A	27.4	N/A	- 1.2

A7: SMFRET ANALYSIS OF DNA TRANSLOCATION BY HEL308

The attached CD-ROM includes a movie entitled 'Hel308 translocation'. The experiment was performed using PEG-coated quartz slides containing 200 pM 5' cy3-DNA. Cy5-labeled Hel308 (10 nM) was bound to the DNA and complexes were imaged by excitation of the cy3 dye. Imaging was initiated prior to the injection of ATP/ $MgCl_2$ (in imaging buffer) onto the slide (see figure 6.23) to stimulate translocation of the protein along DNA. The attached 'Hel308 translocation' movie reveals that at the beginning of data collection, the DNA (cy3): protein (cy5) complexes were in a state of low FRET, represented by the high excitation of cy3 molecules (left panel of movie). Upon injection of ATP/ $MgCl_2$, Hel308 must translocate along the DNA, primarily in a 3'-5' direction, to account for the increase in FRET efficiency between the cy3 and cy5 molecules. This is represented by the gradual excitation of cy5 molecules (right panel of movie) and loss of excitation of cy3 molecules. At the end of data collection, the majority of DNA: protein complexes reached a state of high FRET suggesting that the majority of visualised protein molecules translocated to the 5' end of the DNA molecule during the experiment.