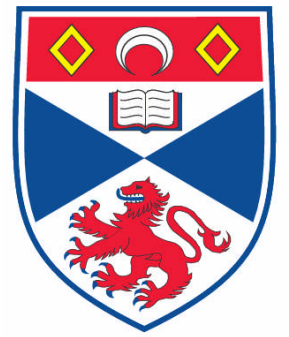
HELICASES AND DNA DEPENDENT ATPASES OF SULFOLOBUS SOLFATARICUS

Jodi Domonique Richards

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



2008

HELICASES AND DNA DEPENDENT ATPASES

OF *SULFOLOBUS SOLFATARICUS*

Jodi Domonique Richards

A thesis submitted to the
University of St Andrews for the degree of

DOCTOR OF PHILOSOPHY

June 2008

Table of contents

Figures and Tables

Abbreviations

1 U/µl	Amount of enzyme that incorporates 10 nmoles of dNTPs into acid insoluble form at 74 °C in 30 minutes under stated assay conditions
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
AAA ATPase	ATPase associated with various cellular activities
ACS	ARS consensus sequence
AEBSF	4-(2-aminoethyl)benzenesulphonyl fluoride
AP site	Apurinic/ apyrimidinic site
APS	Ammonium Persulphate
ARS	Autonomously replicating sequence
(d)ATP	(deoxyribonucleotide) adenosine triphosphate
BER	Base excision repair
BLAST	Basic local alignment and search tool
BM	Branch migration
bp	Base pair
BS	Bloom’s Syndrome
BSA	Bovine serum albumin
CD	Circular dichroism spectroscopy
Cdt1	Chromatin licensing and DNA replication factor 1
CPD	Cyclobutane pyrimidine dimer
CS	Cockayne’s Syndrome
CV	Column volume
Dam	DNA adenine methylase
DMSO	Dimethyl Sulfoxide
DOC	Deoxycholate
ds/ssDNA	Double/single stranded deoxyribonucleic acid

DRD	Damage recognition domain
D/SSB or D/SSG	Double/single strand break double/single strand gap
DTT	1,4-dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
EMSA	Electrophoretic mobility shift assay
EPR	Electron paramagnetic resonance
ERCC1/ 3	Excision repair cross complementing 1/ 3
GGR	Global genome repair
GLD	Gel loading day
HAT	Histone acetyltransferase
HD	Helicase domain
HGT	Horizontal gene transfer
HJ	Holliday junction
HR	Homologous recombination
HTH	Helix-turn-helix motif
ICPL	Isotope-coded protein label
IPTG	Isopropyl-beta-D-thiogalactopyranoside
Kd	Dissociation constant
(k)Da	(Kilo)Dalton
KOD	*Thermococcus kodakaraensis*
LB	Luria Bertani
LGT	Lateral gene transfer
MCM	Minichromosome maintenance
Mfd	Mutation frequency decline
MMR	Mismatch repair
MS	Mass spectrometry
NER	Nucleotide excision repair
NHEJ	Non homologous end joining
ntd	Nucleotide
(d)NTP	(deoxyribonucleotide) nucleotide triphosphate
OD	Optical density

ORC/B	Origin recognition complex/box
PBL 2025	Paul Bloom 2025 *Sulfolobus solfataricus* strain
PBS	Phosphate buffered saline
PCA	Perchloric acid
PCNA	Proliferating cell nuclear antigen
(RT) PCR	(Real time) polymerase chain reaction
PEG	Polyethylene Glycol
PIP box	PCNA interacting protein box
PK	Protein kinase
PNK	Polynucleotide kinase
PVDF	Polyvinylidene fluoride
RC	Replication complex
Rcf	Relative centrifugal force
RF	Replicative form
RFC	Replication factor C
RNA	Ribonucleic acid
RPA	Replication protein A
Rpm	Revolutions per minute
RTS	Rothmund Thomson Syndrome
Sac	*Sulfolobus acidocaldarius*
SDS	Sodium dodecyl sulphate
SDS-PAGE	SDS-polyacrylamide gel electrophoresis
SF1/2/3/4/5/6	Superfamily 1/2/3/4/5/6 (Helicase)
SRH	Second region homology domain
SSB	Single stranded DNA binding protein
Sso	*Sulfolobus solfataricus*
*Sso*XPB1/2	*Sulfolobus solfataricus* XPB1/2
T7g4p	Bacteriophage T7 gene 4 protein
TCA-DOC	Trichloroacetic acid-deoxycholate
TCR	Transcription coupled repair
TBE	Tris-borate EDTA

TEMED	Tetramethylethylenediamine
TEV	Tobacco Etch virus
TFIIH/E	Transcription factor II H/E
TTD	Trichothiodystrophy
(d)TTP	(deoxyribonucleotide) thymine triphosphate
UV	Ultra violet light
WS	Werner's Syndrome
WT	Wild type
XP	*xeroderma pigmentosum*
XPA/B/C/D/E/F/G	*xeroderma Pigmentosum* complementation group A/B/C/D/E/F/G

Declaration

I, Jodi Domonique Richards, hereby certify that this thesis, which is approximately 40,000 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 2004 and as a candidate for the degree of PhD in October 2005; the higher study for which this is a record was carried out in the University of St Andrews between 2004 and 2007.

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 I understand that I am giving permission for it to be made available for the 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. I also understand that the title and abstract will be published, and that a copy of the work may be made and supplied to any bona fide library or research worker.

Date …….. Signature of candidate …….

Abstract

DNA is susceptible to various types of damage as a result of normal cellular metabolism or from environmental sources. In order to maintain genome stability a number of different, partially overlapping DNA repair pathways have evolved to tackle specific lesions or distortions in the DNA. Nucleotide excision repair (NER) is highly conserved throughout eukarya, bacteria and archaea and predominantly targets lesions that result from exposure to UV light, for example cyclobutane pyrimidine dimers and 6-4 photoproducts. The majority of archaea possess homologous of the eukaryotic repair genes and this thesis describes the isolation and the characterization of two XPB homologues identified in the crenarchaeon *Sulfolobus solfataricus*, SsoXPB1 and SsoXPB2.

Human XPB is one of 10 proteins that make up the TFIIH transcription complex. The activity of XPB is tightly controlled by protein interactions, in particular with p52, which stimulates the ATPase activity of XPB. Rather than a conventional helicase, human XPB is thought to act as an ATP dependent conformational switch. Consistent with human XPB, however, the *S. solfataricus* proteins were unable to catalyse strand separation and the identification of an archaeal protein partner, Bax1, for SsoXPB2 was one of the focuses of this project.

In order to maintain genome stability, the DNA must be replicated accurately with each cell cycle. When the advancing replication fork stalls at a lesion or a DNA break, it is crucial that the fork is reset and that replication continues to completion. The helicase Hel308 is thought to clear the lagging strand template of a stalled replication fork in order for replication restart to proceed via homologous recombination (HR). Although the specific function of Hel308 is not well understood, the possibilities are described in this thesis.

Strand exchange proceeds to form a D-loop, followed by branch migration to increase regions of heterology during the synapsis stage of HR. No motors for branch migration have previously been recognised in archaea, although the identification of a possible candidate was investigated during this project.

Acknowledgements

I would like to take this opportunity to thank my supervisor Professor Malcolm White for all his advice and guidance throughout the project and for his continuous support. I would also like to thank BBSRC for funding this project and for providing the opportunity for me to attend an international conference, this thanks also extends to Professor Malcolm White.

I would also like to express my thanks to my parents, Mike and Kate Richards, my grandma, Joyce Sleep, my aunt and uncle, Lynn and Bob Sykes and my boyfriend, Mat Munro for their endless support and encouragement and for their frequent visits north of the border.

Lynn Sykes, Dr Sonia Paytubi, Dr Robert Dorazi, Dr Jana Rudolf and Professor Malcolm White receive special thanks for reading my thesis through and for providing helpful and much appreciated advice, thank you.

I should like to acknowledge Dr Catherine Botting, Dr Sally Shirran and Alex Housten in the Mass Spectrometry facility at St Andrews University for their contributions to the project.

And lastly, but certainly not least, I would like to thank my laboratory group for such a wonderful working environment and for an exciting life outside science. This thanks also extends to all the friends I have made in St Andrews over the last three years, without them this would not have been the fantastic experience it has proved to be.

CHAPTER 1: INTRODUCTION

1.1 Archaea

1.1.1 The third domain of life

Dividing the living world into either prokaryotic or eukaryotic was a fundamentally cytological concept and phylogenetic only by inference (Woese *et al.*, 1990). Organisms with eukaryotic cellular organisation were grouped according to the possession of complex properties. The prokaryotes, however, were classified based on characteristics that they lacked compared to the eukaryotic cell (Woese *et al.*, 1990). Molecular sequence comparison provided a much more accurate tool, since an organism's genome is a record of its evolutionary history. An approach to measure the degree of difference between comparable structures, therefore, was used to establish a more reliable phylogenetic relationship (Woese and Fox, 1977) .

Ribosomal RNA sequences were an excellent candidate based on its broad distribution amongst all living organisms. Any changes to its sequence occur slowly with time, meaning that relations between very distant species can be detected (Woese and Fox, 1977). Analysis of the primary structure of the 16S (18S in eukaryotes) rRNA of the small subunit of the ribosome clearly indicated that organisms clustered into distinct groups. The first comprised the typical bacteria, for example the genus *Bacillus*, and the 18S rRNA of the eukaryotic cytoplasm defined the second (Woese and Fox, 1977). Used in the phylogenetic sense, these groups correspond roughly to the categories "prokaryotic" and "eukaryotic". They do not, however, include all living organisms. Thus, the molecular criteria replaced the phenotypic criteria (Woese and Fox, 1977).

At the cytological level the archaea are prokaryotic in that they are single cells containing a circular chromosome and lacking eukaryotic organelles (Bernander, 2000). The living world, therefore, is clearly divided into three distinct groups or domains only at the molecular level; the bacteria, the archaea and the eukarya (Woese *et al.*, 1990). Characteristic bacterial, archaeal and eukaryotic versions of molecular systems exist and are shared by members of each group (Woese *et al.*, 1990).

The archaeal domain is divided into euryarchaeota (euryarchaea) and crenarchaeota (crenarchaea) (Woese *et al.*, 1990). The euryarchaeal subdomain includes extreme halophiles, sulphate reducing species, thermophiles and methanogenic lineages. The crenarchaea comprise the (extreme)thermophiles and the sulphur-dependent archaea, the

thermophilic nature of the crenarchaea is the only characteristic common to the two major archaeal subdomains (Woese *et al*., 1990). Crenarchaea are also abundant in low temperature environments such as ocean waters, terrestrial sediment and soil (Konneke *et al*., 2005). The crenarchaea (hyperthermophilic archaea) are regarded as the most ancient of the archaea and are, therefore, considered to bear the closest resemblance to the progenitor of the archaea and eukarya (Woese *et al*., 1990; Dionne *et al*., 2003a). However, the presence and comparisons of proteins, such as histones and the repair proteins, for example XPB and XPF, imply that the euryarchaea resemble the eukaryotes more closely than the crenarchaea do (White, 2003). A third archaeal phylum has been proposed based on small subunit rRNA comparison, the korarchaeota, although analysis did not clarify whether this was a separate archaeal lineage or another crenarchaeal group (Barns *et al*., 1996).

Nanoarchaeum equitans, a symbiont growing on the surface of the crenarchaeal *Ignicoccus* species, possesses a genome close to the minimum theoretical size for a living organism. Analysis of the small subunit (16S) rRNA led to the original assumption that *N. equitans* represented a novel, ancient archaeal phylum called the Nanoarchaeota (Huber *et al*., 2002). However, phylogenies based on concatenated ribosomal protein datasets and BLASTP analysis identified *N. equitans* as a very divergent euryarchaeon, with a close affinity to *Thermococcales* (Brochier *et al*., 2005).

Although the archaea are quite distinct; morphologically and metabolically they resemble the bacteria, yet they process genetic information (DNA repair, replication, transcription, translation, amino-acylation of tRNAs and tRNA splicing) much like the eukaryotes (Olsen and Woese, 1997). Archaea are, therefore, very good model organisms for the study and manipulation of important DNA transactions in the more complex and cluttered eukaryotic systems (Rossi *et al*., 2003; Fujikane *et al*., 2005).

1.1.2. *Sulfolobus solfataricus* P2

Sulfolobus is the most widely studied genus of the crenarchaeota and is considered to be a model organism for research on mechanisms such as DNA replication and transcription (She *et al*., 2001). Much of what has already been discovered and known about these organisms is due to work on this genus. *Sulfolobus solfataricus,* a sulphur-metabolising aerobic crenarchaeon, possesses a genome of 2,992,245 base pairs on a

single chromosome, encoding 2,977 proteins and a large number of RNAs. It thrives at 80 °C in an acidic (pH 2-4) sulphur-rich environment (She *et al.*, 2001).

The genome of *S. solfataricus* has been fully sequenced (She *et al.*, 2001) and the research presented in this thesis focuses on the characterisation of a number of putative helicases from *S. solfataricus* thought to be involved in repair of DNA damage and restart of stalled replication forks. The genes *sso0473*, *sso0959*, *sso1468* and *sso2462* encode these proteins. *Escherichia coli* was used as a heterologous host for the expression of recombinant proteins.

1.2 DNA damage and Nucleotide Excision Repair

1.2.1 DNA damage and Repair

DNA is susceptible to many types of damage, such as spontaneous disintegration or modification of bases, which are caused both by environmental sources and products of normal cellular metabolism (Costa *et al.*, 2003). Since no single pathway could repair all types of damage, many different, partially overlapping DNA repair pathways have evolved to tackle specific lesions or distortions in the DNA (figure 1.1). These pathways show strong evolutionary conservation and are divided into several groups: direct reversal of the DNA damage, base excision repair (BER), mismatch repair (MMR), nucleotide excision repair (NER) and damage bypass pathways (Lindahl and Wood, 1999; Seitz *et al.*, 2001).

Photoreactivation systems are able to directly reverse base damage caused by exposure to ultra violet (UV) radiation, for example cyclobutane pyrimidine dimers (CPDs). Enzymes referred to as DNA photolyases (photoreactivating enzymes) use the energy from visible light to reverse this dimerisation of pyrimidines. This is a feature of bacterial and eukaryotic organisms, with placental mammals being the exception (Lindahl and Wood, 1999). Photoreactivation has also been identified in a number of archaea, including *S. solfataricus* and *S. acidocaldarius* (Seitz *et al.*, 2001).

Moreover, certain mutagenic alkylating agents can react with DNA to form O- and N- alkylated bases, which can mispair during semiconservative DNA synthesis (Seitz *et al.*, 2001). A typical O^6-alkylguanine DNA alkyltransferase is bacterial O^6-methylguanine methyltransferase that removes the methyl group from a modified guanine

base that can mispair with thymine. The enzyme removes the methyl group and adds it to one of its own internal cysteine residues, rendering itself inactive. The whole protein molecule is, therefore, sacrificed for the sake of the cell (Lindahl and Wood, 1999). A few archaea, including *S. acidocaldarius* and *P. fulgidus*, have also been shown to possess an O^6-alkylguanine-DNA transferase activity (DiRuggiero *et al.*, 1999).

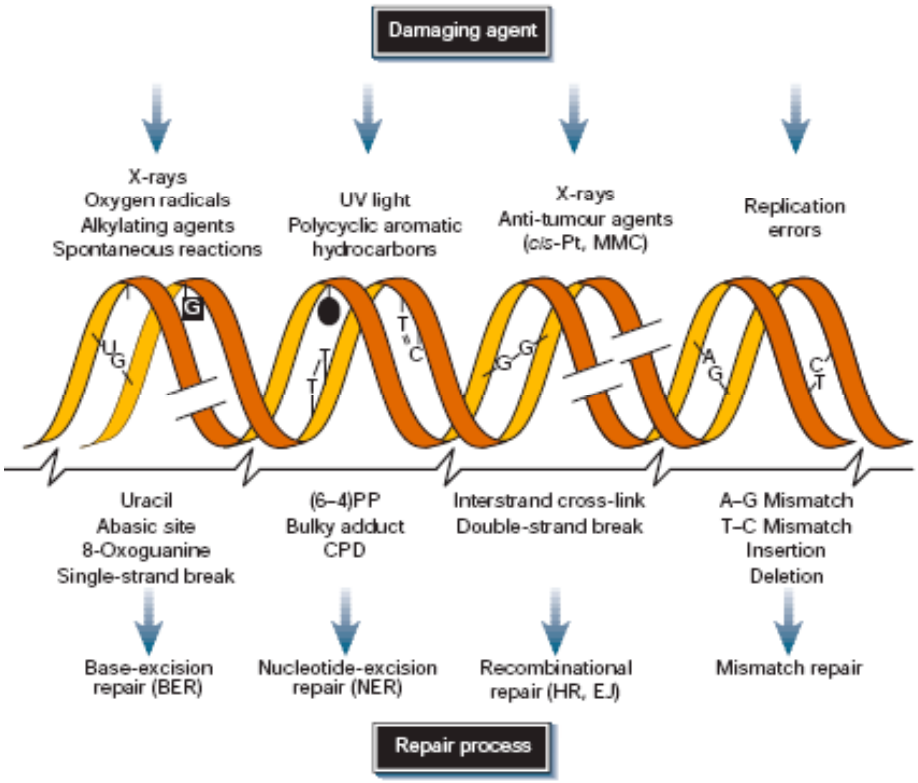


Figure 1.1. A summary of the DNA damaging agents and their respective lesions and repair mechanisms
The different sources of DNA damage are shown at the top of the diagram. Each causes a specific lesion in the DNA (middle panel) and, therefore, many repair pathways are in place to prevent deleterious mutations or cell death, as indicated at the bottom. A number of these pathways are specific to a type of lesion others, however, deal with more universal types of damage (Hoeijmakers, 2001).

The BER pathway deals with most of the endogenous damage to DNA especially oxidative damage, it also removes similar injuries caused by exogenous sources. Enzymes called DNA glycosylases remove single bases that have been chemically modified, most commonly a uracil formed as a result of cytosine deamination, to leave an apurinic/apyrimidinic (AP) site (Seitz *et al.*, 2001). The AP site is further processed by cleavage at the 5' site by an AP endonuclease or 3' by an AP lyase. A phosphodiesterase

removes the deoxyribose phosphate residue allowing the polymerase and ligase to repair the gap (Seeberg *et al*., 1995). Based on sequence homology, AP endo/exonucleases have been found in *M. thermoautotrophicus* and *M. jannaschii* (Seitz *et al*., 2001). Archaea also possess a homologue of 8-oxoguanine DNA glycosylase that is responsible for the removal of mutagenic 8-oxoguanine caused by oxidising agents or ionising radiation (Fishel and Kolodner, 1995; Seitz *et al*., 2001).

MMR tackles bases that have been misincorporated by the DNA polymerase during replication. In *E. coli*, repair of a mismatch must be immediate before Dam (DNA adenine methylase) methylates the nascent strand, rendering the two strands indistinguishable. Transient undermethylation of the nascent strand GATC sequence governs strand selection. In bacteria, the MutHLS pathway is required to produce a strand specific break to promote the recruitment of enzymes and to excise and resynthesise the strand containing the mismatch (Fishel and Kolodner, 1995). It was proposed that MutS binds the mismatch forming a sliding clamp. MutL interacts with the ATP bound MutS connecting it to the MutH endonuclease, which incises the strand at the GATC hemi methylated site. This directs the unwinding and degradation of the nascent strand by the helicase UvrD and one of four ssDNA exonuclease; exo I, exo VII, exo X or RecJ. The excision tracts cover only the stretch between the GATC site and just past the mismatch (Fishel and Kolodner, 1995; Yang, 2000; Acharya *et al*., 2003). The strand is then resynthesised by DNA Polymerase III holoenzyme and DNA ligase seals the resulting gap (Fishel and Kolodner, 1995).

Homologues of the MutL and MutS proteins have been identified in humans and a homologue of MutL was identified in *S. cerevisiae*. Archaea have only been shown to possess a single MutS homologue (Fishel and Kolodner, 1995; Seitz *et al*., 2001; Acharya *et al*., 2003).

Enzymes of the NER pathway target a number of bulky adducts and lesions that cause significant DNA distortion. Global Genome Repair (GGR) NER is the most versatile of all the repair systems, regarding damage recognition. GGR describes the repair of DNA lesions across the genome, including the untranscribed DNA strand, and is initiated by lesions that promote significant DNA distortion (Costa *et al*., 2003). The degree to which a lesion distorts the DNA determines how fast it is detected by the GGR

NER proteins; the greater the distortion the more efficiently it is recognised (Sugasawa *et al.*, 1998).

CPDs and 6-4 photoproducts (6-4 PPs) are prominent examples of lesions that cause bending in the DNA, both of which are induced by exposure to UV light. CPDs are removed at a significantly slower rate than 6-4 PPs since the former have a more subtle effect on the structure of the DNA helix. Distortion of the DNA structure alone, however, is not sufficient to induce NER. The DNA chemistry must be covalently modified in conjunction with the disruption of Watson-Crick base pairing to promote NER, consistent with a bipartite substrate discrimination mechanism (Hess *et al.*, 1997; Batty and Wood, 2000; Vasquez *et al.*, 2002).

TCR is a specialised form of NER that is responsible for removing lesions from the actively transcribed DNA strand, which cause the elongating RNA polymerase to stall (Evans *et al.*, 1997; Wood, 1997; Fousteri *et al.*, 2006) and for which GGR is too slow. In contrast to GGR, the stalled polymerase acts as a signal for damage. The NER pathway is described in detail in section 1.2.2-1.2.5.

Double strand breaks (DSBs) that can arise in the DNA are particularly harmful. There are two pathways specific for the repair of DSBs: homologous recombination (HR) and non-homologous end joining (NHEJ). HR is the mechanism of choice if a second, identical copy of the strand is available. This method is accurate since the homologous, intact strand is used as a template for resynthesis of the broken one (She *et al.*, 2001).

After DSB formation the DNA is processed by a helicase and/or an exonuclease to produce a single strand, which is bound by ssDNA binding proteins. This stage is referred to as pre-synapsis (Heyer *et al.*, 2006). The resultant nucleoprotein filament scans the DNA during synapsis and initiates strand invasion upon detecting the homologous target DNA, forming a heteroduplex DNA molecule or D loop (Seitz *et al.*, 2001; Heyer *et al.*, 2006). A four-stranded structure, a Holliday junction (HJ), is formed by the reciprocal exchange of two strands. During post-synapsis, proteins, such as those involved in strand exchange, or specialized helicases, promote branch migration to extend regions of heteroduplex DNA. RuvABC is a helicase/nuclease complex involved in branch migration and also cleavage of the HJ, which occurs in either of two orientations producing one of two different recombinant products (Seitz *et al.*, 2001).

The resulting gap is sealed by a DNA ligase (Heyer *et al.*, 2006). Bacteria repair DSBs by either the RecBCD or the RecF pathway, both of which depend on the action of RecA to promote strand exchange during synapsis. These two pathways are discussed further in section 1.3.4.

Non-Homologous End Joining (NHEJ) is the second mechanism involved in the repair of DSBs. In contrast to HR, NHEJ requires minimal base pairing; the two ends of the DSB break are aligned and ligated back together without the requirement for significant homology (Seitz *et al.*, 2001; Pitcher *et al.*, 2005a; Pitcher *et al.*, 2005b). This is a comparatively error prone pathway with the capacity to produce genetic alterations. These could range from loss or addition of nucleotides at the broken ends, to chromosome translocations (Krejci *et al.*, 2003; Pitcher *et al.*, 2005a; Pitcher *et al.*, 2005b). The three main steps to NHEJ are DNA end binding and bridging, terminal processing and ligation. The majority of ionizing radiation induced DSBs have inappropriate termini for direct joining and, therefore, after end bridging, the non-complementary ends require processing before ligation (Krejci *et al.*, 2003). This is achieved by nucleases and DNA polymerases, which trim the ends and fill any gaps respectively (Pastwa and Blasiak, 2003). In mammalian cells there is a minimal requirement for the DNA-dependent protein kinase (DNA-PK), which is composed of the DNA binding subunit, Ku70/80 heterodimer (Yku70/80 in yeast), and a catalytic subunit (DNA-PKcs). DNA-PK is recruited to the broken ends to initiate NHEJ (Krejci *et al.*, 2003; Kim *et al.*, 2005). Ku is responsible for binding and aligning the DSB ends and for recruiting the ligase IV/XRCC4 complex (Dn14/ Lif1 in yeast) and stimulating its activity (Ramsden and Gellert, 1998; Nick McElhinny *et al.*, 2000; Pitcher *et al.*, 2005a).

The Mre11/Rad50/Nbs1 nuclease complex is involved at this stage, as well as damage signaling it protects the DNA ends against degradation. It can compete with DNA-PKcs for Ku-bound DNA ends, suggesting that parallel NHEJ pathways may exist in mammalian cells (Krejci *et al.*, 2003). Finally the ligase IV/XRCC4 complex is recruited to seal the gap (Pastwa and Blasiak, 2003). In yeast the Mre11/Rad50/Xrs2 (M/R/X) complex is essential for NHEJ, it facilitates the ligation reaction and end bridging through DNA unwinding (Zhang *et al.*, 2001). The NHEJ pathway is not used by archaea to repair DSBs; homologues of the Mre11 and Rad50 protein have been

identified in archaea, but no Xrs2 or NBS1 subunits have been detected (Seitz *et al.*, 2001).

Although *E. coli* lacks NHEJ, bacteria including species of the genera *Bacillus* and *Mycobacteria* have been identified to possess putative NHEJ proteins (Doherty and Jackson, 2001; Stephanou *et al.*, 2007). In the bacterial system, the Ku-associated ligases often possess domains, in addition to the core ligase domain, that share sequence homology with eukaryotic polymerases and DNA nucleases. Specifically, the *Mycobacterial* ligase possesses a variety of activities, for example DNA dependent RNA primase, terminal transferase and DNA/RNA gap filling polymerase activities (Della *et al.*, 2004; Pitcher *et al.*, 2005a).

Alternatively, a DNA lesion may be bypassed by certain polymerases. These are able to synthesise DNA past a lesion, which is repaired after replication has occurred.

1.2.2 Bacterial Nucleotide Excision Repair

In bacteria NER is achieved by the action of the UvrABC complex in concert with the helicase UvrD (Aravind *et al.*, 1999), the "cut and paste" mechanism of which is comparable to the eukaryotic system.

1.2.2.1 Bacterial Global Genome Repair (GGR)

In bacteria the UvrABC and UvrD proteins are responsible for recognising and excising the damage.

As a dimer, UvrA binds to two UvrB molecules to form an $UvrA_2B_2$ complex (Malta *et al.*, 2006, 2007). UvrA induced dimerisation of UvrB serves as a block to premature binding of UvrC (Malta *et al.*, 2007). One UvrB monomer binds the DNA, while the second molecule remains only loosely associated (Verhoeven *et al.*, 2002; Malta *et al.*, 2006).

The DNA wraps around one of the UvrB monomers as it probes for damage. If no damage is detected this process is thought to be repeated by the second UvrB monomer, ensuring damage detection on both strands (Verhoeven *et al.*, 2002). On encountering the damage, the ATPase activity of UvrB results in the release of the UvrA dimer. A β-hairpin of UvrB is inserted into the helix locking it in place (Theis *et al.*, 2000; Mellon, 2005) and the formation of this UvrB-DNA complex results in unwinding and DNA bending/kinking (Zou and Van Houten, 1999). The contact between the UvrB monomers

is broken, releasing one of the UvrB monomers and allowing recruitment of UvrC (figure 1.2) (Verhoeven *et al.*, 2002; Malta *et al.*, 2006, 2007).

The UvrBC complex catalyses incisions on each side of the lesion; the catalytic sites responsible for the 5' and 3' incisions are located on the N- and C-terminal regions of UvrC (Mellon, 2005). The first incision occurs 3' of the lesion, between the fourth and fifth phosphodiester bonds and the second at the eighth phosphodiester bond 5' of the damaged site (Orren and Sancar, 1990; Theis *et al.*, 2000; Verhoeven *et al.*, 2000; Mellon, 2005).

UvrD loads onto the free 3' end of the damaged strand and displaces the stretch of nucleotides containing the lesion, using its 3'-5' helicase activity. UvrB remains bound to the undamaged strand even after dual incision. Since removal of UvrC and the strand containing the lesion by UvrD does not involve the undamaged strand, UvrB remains on the DNA until DNA polymerase I resynthesises the strand, displacing it (Theis *et al.*, 1999). DNA ligase I joins the 3' end of the new patch to the 5' end of the existing strand to seal the gap (Batty and Wood, 2000).

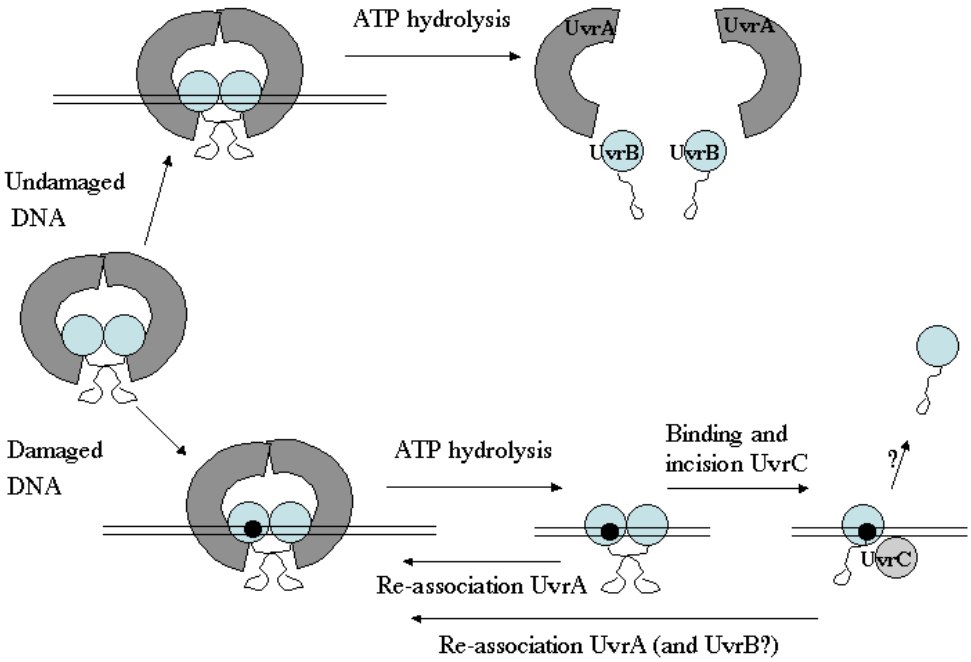


Figure 1.2. A schematic representation of the bacterial NER pathway
UvrA (dark grey) dimerises ($UvrA_2$) and binds to two monomers of UvrB (light blue) to form an $UvrA_2B_2$ complex. Upon encountering damage, the ATPase activity of UvrB results in the release of $UvrA_2$, destabilising the $UvrB_2$ dimer. One of the UvrB monomers is released and UvrC (light grey) is recruited. The figure was adapted from Malta *et al.* (2007).

1.2.2.2. Bacterial Transcription Coupled Repair (TCR)

After UV irradiation, for example, a lesion in the transcribed strand will cause stalling and backtracking of the RNA polymerase, along with the transcription bubble and the nascent mRNA strand (Park *et al*., 2002; Mellon, 2005). Mutation Frequency Decline (Mfd) protein recognises the backtracked polymerase and induces forward translocation of the elongation complex until it re-encounters or bypasses the lesion (figure 1.3) (Mellon, 2005). RNA polymerase displacement requires Mfd to interact with the DNA immediately upstream of the stalled complex. It is thought to unwind the RNA-DNA hybrids and rewind the single stranded transcription bubble resulting in destabilization of transcription complexes (Smith *et al*., 2007). Mfd interacts with UvrA, which is required to recruit UvrB and is thought to be the reason for enhanced transcription coupled DNA repair (Smith *et al*., 2007). GreA and GreB are cellular proteins that induce internal cleavage of the extruded RNA, forming a new primer for RNA synthesis (Park *et al*., 2002). The polymerase backtracks again or is released by Mfd, UvrC binds and downstream NER can continue (Park *et al*., 2002; Mellon, 2005), as described for GGR. This would suggest that coupling of NER and transcription is mediated by the correct positioning of the polymerase and the transcription bubble, rather than interactions between the proteins (Mellon, 2005)

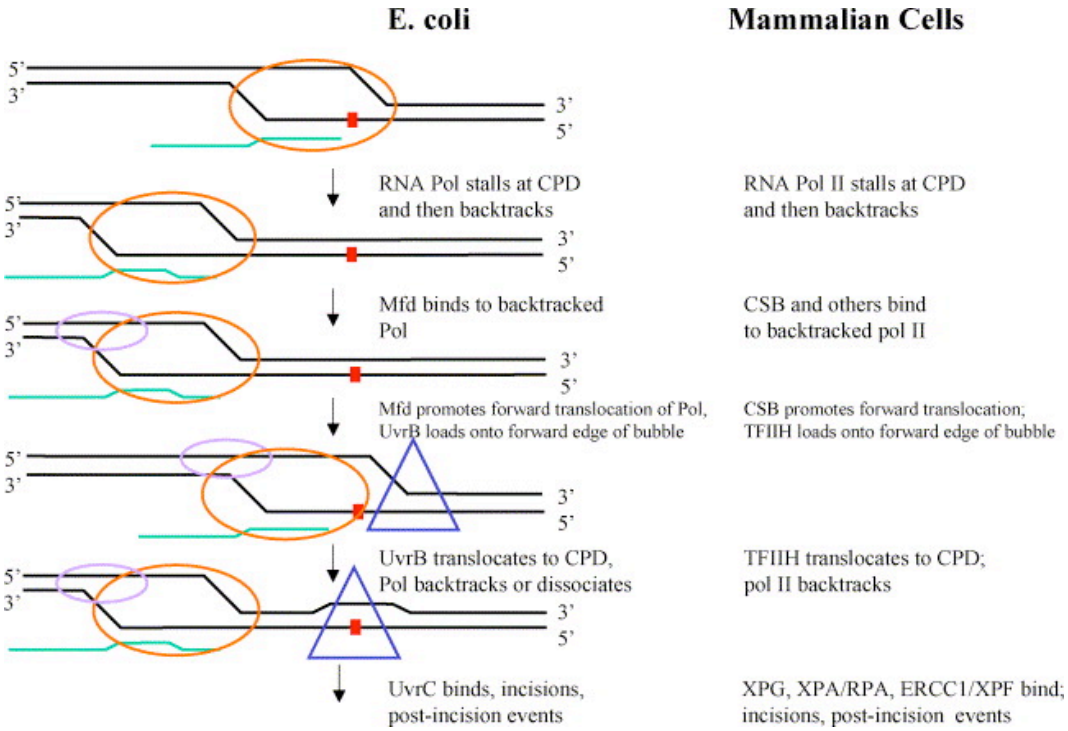


Figure 1.3. A scheme describing TCR
TCR in *E. coli*: upon encountering a lesion in the transcribed strand, the RNA polymerase (orange circle) stalls and backtracks along the DNA with the nascent mRNA strand (green line). Mfd (purple circle) recognises the backtracked polymerase and stimulates forward translocation of the complex, until it reaches or bypasses the lesion. UvrB (blue triangle) is able to load onto the DNA at the front edge of the transcription bubble. The polymerase backtracks again, or is alternatively released from Mfd and NER can continue (Mellon, 2005; Smith *et al.*, 2007). Figure taken from Mellon (2005).

1.2.3. Eukaryotic Nucleotide Excision Repair

The 'cut and paste' mechanism of NER has been strictly conserved, however, the repertoire of mammalian proteins required for NER is significantly more complex than the *E. coli* system, for example (Mellon, 2005).

1.2.3.1 Eukaryotic Global Genome Repair (GGR)

Certain lesions can induce local unwinding of a few bases around the damaged site and these are good substrates for GGR. This local unwinding energetically favours DNA bending and in turn may facilitate the NER reaction (Van Houten, 1990; van Hoffen *et al.*, 2003).

The proteins involved in NER (figure 1.4) assemble sequentially at the site of DNA damage rather than preassembling into a 'repairosome' (de Laat *et al.*, 1999). XPC-hHR23B is one of the initial damage sensors, which binds to the damage and locally

distorts the helix around the lesion. It is required to recruit the NER factors and, therefore, triggers formation of the pre-incision complex (de Laat *et al.*, 1999; Volker *et al.*, 2001; Riedl *et al.*, 2003). XPC-hHR23B can directly recognise the alterations to the DNA induced by some lesions, such as 6-4 PPs. Other lesions that are more subtle, for example CPDs, require UV DNA damage binding protein UV-DDB/ XPE to facilitate damage recognition by recruiting XPC-hHR23B and stimulating repair (de Laat *et al.*, 1999; Fitch, 2000; Kulaksiz *et al.*, 2005).

Transcription Factor II H (TFIIH) is a 10 subunit complex (Coin *et al.*, 2007), including 3'-5' and 5'-3' helicases XPB and XPD, respectively. After damage recognition, TFIIH appears at the repair site and together with XPC-hHR23B, XPB and XPD is responsible for opening the DNA to form a complex that is structurally ready to recruit other NER factors (de Laat *et al.*, 1999; Riedl *et al.*, 2003; Tapias *et al.*, 2004). This is dependent upon the ATPase activity of XPB, while the role of helicase is devoted to XPD (Coin *et al.*, 2007). The XPB protein is discussed in more detail in Chapter 3.

XPA is implicated in opening the complex further and stabilising it, acting as a wedge to retain the open structure ready for the recruitment of replication protein A (RPA), XPG and ERCC1-XPF (Riedl *et al.*, 2003; Tapias *et al.*, 2004). It interacts with core NER factors including TFIIH and plays a role in positioning factors correctly around the DNA lesion (de Laat *et al.*, 1999).

RPA is a single stranded binding protein that interacts with XPB, increasing its affinity for DNA (Riedl *et al.*, 2003). RPA binds to and protects the undamaged strand and is thought to contribute to the formation of the fully open incision complex (DE LAAT, 1999). RPA is crucial for orientating the nucleases XPG and ERCC1-XPF, which bind to the 5' and 3' sides of RPA respectively (de Laat *et al.*, 1999). RPA inhibits ERCC1-XPF incision in the undamaged strand, while stimulating incision of the damaged strand, thereby conferring strand specificity (de Laat *et al.*, 1998; de Laat *et al.*, 1999).

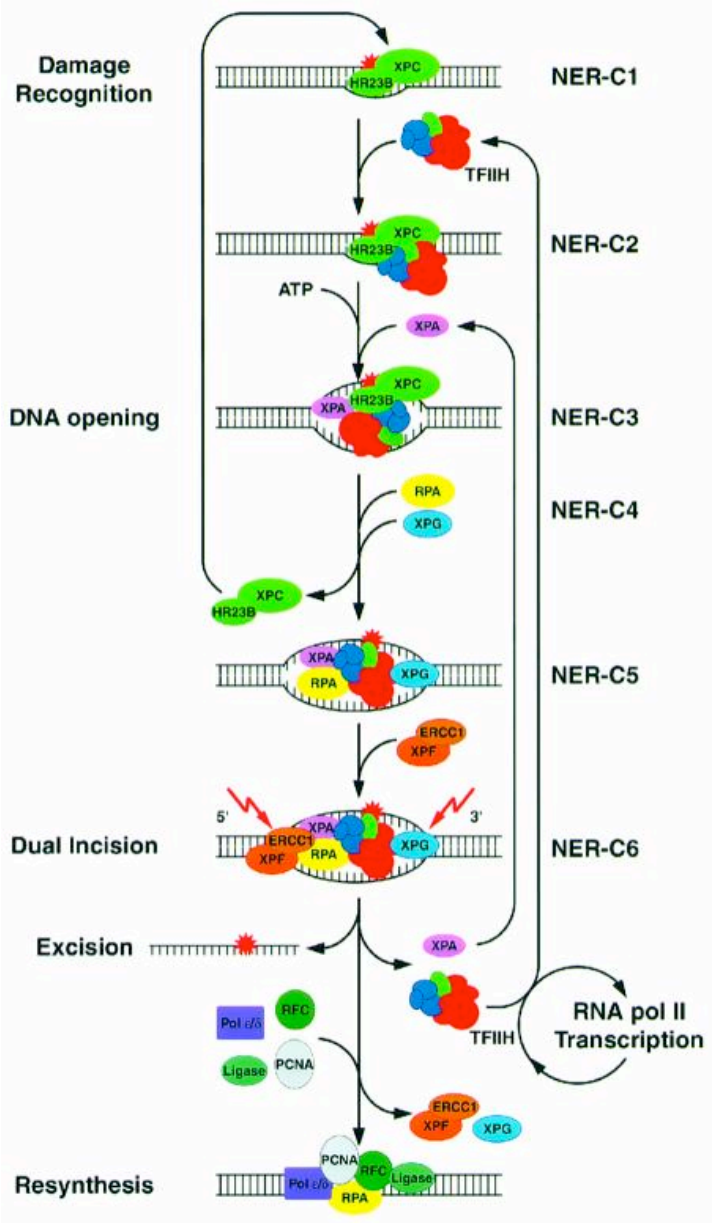


Figure 1.4. A summary of the eukaryotic NER pathway
In global genome repair (GGR) the distortion in the DNA is recognised by XPC-hHR23B (light green), which then recruits the transcription factor TFIIH (red). The helicases, XPB and XPD are part of TFIIH and catalyse DNA unwinding to form an open complex around the lesion. XPA (purple) acts as a wedge to stabilise the open structure ready to recruit RPA (yellow), XPG (blue) and ERCC1-XPF (orange). RPA protects the undamaged DNA strand and is crucial for orientating the endonucleases XPG and ERCC1-XPF. XPG makes the first incision 3' of the lesion and its presence is required for the 5' incision catalysed by ERCC1-XPF. RPA is the only factor required for both excision and resynthesis, the latter is achieved by DNA polymerases δ and ε (dark blue) in the presence of PCNA (pale blue) and RFC (dark green). The resulting nick is sealed by DNA ligase (dark green) (de Laat *et al.*, 1999; Riedl *et al.*, 2003; Tapias *et al.*, 2004). The figure was taken from Riedl *et al.* (2003).

Incisions are made asymmetrically around the damaged site; XPG catalyses strand cleavage 2-8 nucleotides (Svoboda *et al.*, 1993) away on the 3' side, whereas ERCC1-XPF incises the DNA 15-24 nucleotides (Svoboda *et al.*, 1993) 5' of the lesion. The 3' incision must be complete and XPG must be present in the complex for ERCC1-XPF to complete the incision, however, its catalytic activity is not required at this point (de Laat *et al.*, 1999). The 5' incision reaction leaves a free hydroxyl-group, which acts as a primer for resynthesis and, therefore, no further modifications are required before gap-filling DNA synthesis (Sijbers *et al.*, 1996).

Most of the NER factors will disassociate before DNA synthesis occurs, except RPA, which is required for gap filling and protecting the DNA against nucleases (de Laat *et al.*, 1999). DNA synthesis is achieved by DNA polymerase δ or ε, in the presence of processivity factor Proliferating Cell Nuclear Antigen (PCNA). Replication factor C (RFC) is also involved in facilitating polymerase assembly, it binds to the 3' termini of the priming strand and helps to load PCNA. Finally, the new strand is ligated to the original sequence by DNA ligase I (Costa *et al.*, 2003).

1.2.3.2 Eukaryotic Transcription Coupled Repair (TCR)

In eukaryotic cells, TCR and GGR differ at the damage recognition stage. The stalling of RNA polymerase II at a lesion is more than sufficient to initiate NER (Fousteri *et al.*, 2006) and, therefore, the XPC-hHR23B complex is dispensable for TCR. Two other proteins, CSA and CSB, are required to prevent the degradation of the stalled RNA polymerase complex and to recruit proteins involved in damage repair and transcriptional restart. CSB is a member of the DNA-dependent ATPase SWI/SNF family, involved in chromatin remodelling (Costa *et al.*, 2003). It is required to assemble NER factors and histone acetyltransferase (HAT) p300 to the stalled polymerase II, which is known to remain associated with its template. CSA contains 5 WD-40 repeats important in protein-protein interactions and it interacts with CSB and a subunit of TFIIH, implicating both CSA and CSB in transcription (Henning *et al.*, 1995; Costa *et al.*, 2003). Proteins containing WD-repeats appear to play a regulatory role, since they do not seem to posses a catalytic function. CSA and CSB may be involved in controlling the chromatin structure during transcription and subsequently facilitating DNA repair (Henning *et al.*, 1995).

Assembly of the TCR complex does not require the stalled RNA polymerase II to dissociate from the DNA as it is able to backtrack along the DNA with its associated transcription bubble (Fousteri *et al*., 2006), similar to the bacterial mechanism described in section 1.2.2.2 (figure 1.3). CSB and XPG interact with the active polymerase, however, upon encountering a block the interaction between the polymerase and CSB becomes more stable. This promotes forward translocation of the RNA polymerase and allows for CSB to fulfil its role as a coupling factor by recruiting a number of proteins including HAT p300, CSA and the NER factors (Mellon, 2005; Fousteri *et al*., 2006). TFIIH loads onto the forward edge of the bubble and translocates to the lesion; the RNA polymerase backtracks once again (figure 1.3) (Mellon, 2005). CSA has no role in targeting the NER factors to the TCR site, instead it is involved in the recruitment of the scaffolding protein XAB2, the nucleosome binding protein HMGN1 and TFIIS. TFIIS cleaves the 3' end of the mRNA strand to reposition it in the active centre of the polymerase in order for transcription to restart (Tornaletti *et al*., 1999; Fousteri *et al*., 2006). The NER reaction itself continues as described for GGR in section 1.2.3.1.

1.2.4. Human Disease

Mutations in the NER proteins are associated with *xeroderma pigmentosum* (XP), Cockayne Syndrome (CS) and trichothiodystrophy (TTD). The common feature of these three syndromes is photosensitivity, although it manifests at varying degrees. The other clinical manifestations differ substantially, ranging from a high predisposition to cancer to aging symptoms (Costa *et al*., 2003). Severe photosensitivity and a high incidence of skin cancers induced by sunlight are characteristic of XP (Volker *et al*., 2001) as a result of mutations in any of the XP genes (*xpa-xpg*). The frequency of internal tumours also increases and some patients exhibit accelerated neurodegeneration. CS is a TCR disorder caused by mutations in the *csa* or *csb* and is very different from XP. No predisposition to cancer is seen because the CS cells protect against tumourigenesis by a particular sensitivity to lesion–induced apoptosis (Hoeijmakers, 2001). Although TTD and CS share many symptoms, scaly skin and brittle hair and nails are typical of TTD patients.

Since *xpb* and *xpd* genes encode proteins with major roles in transcription and nucleotide excision repair, mutations in either can give rise to all three of these

syndromes (Hoeijmakers, 2001). The consequences of mutations to the *xpb* gene are discussed further in chapter 3.

1.2.5 Nucleotide Excision Repair in the third kingdom

The NER incision pattern of *Methanothermobacter thermoautotrophicus* resembled that of the bacterial system (Ogrunc *et al.*, 1998), leading to the assumption that archaeal NER was homologous to the bacterial pathway. However, outside the bacteria only a handful of organisms (mesophilic methanogens and halophiles), including *M. thermoautotrophicus* and *Halobacterium sp. NRC-1*, have been identified with the UvrABC exicision system (Costa *et al.*, 2003; White, 2003; Crowley *et al.*, 2006). The *M. thermoautotrophicus* genome contains a complete operon including *uvrA*, *B* and *C* genes, which is likely the result of lateral gene transfer from bacteria (Aravind *et al.*, 1999).

The majority of archaea possess homologues of the helicases XPD and XPB and the nucleases XPF and XPG, which implies that archaeal NER is eukaryotic in character (White, 2003). The damage recognition proteins, XPA, XPC and XPE do not seem to have any archaeal homologues (White, 2003), a pattern seen in plants and also in the genome of the human malaria parasite, *Plasmodium falciparum* (Gardner *et al.*, 2002). This would suggest that NER has undergone increasing evolutionary refinement and the proteins have been acquired in the higher eukaryotes, possibly in order to be able to deal with the more organized packaging of chromosomes into chromatin (Costa *et al.*, 2003). *S. solfataricus* possesses a single stranded DNA binding protein (SSB). It comprises an OB fold, similar to that found in eukaryotic RPA, linked to a flexible C-terminal tail reminiscent of the *E. coli* homologue (Wadsworth and White, 2001; Kerr *et al.*, 2003; Cubeddu and White, 2005). *S. solfataricus* SSB is able to melt dsDNA, to discriminate between undamaged and damaged DNA strands and to interact with putative repair proteins. This, therefore, implicates SSB in damage recognition (Cubeddu and White, 2005).

The eukaryotic and archaeal 5'-3' helicase XPD contains a conserved domain between Walker A and B boxes, identified by studies carried out in *S. acidocaldarius*. This domain comprises four cysteine residues, which form an Fe-S cluster binding motif. Mutants lacking this cluster no longer exhibit helicase activity, implicating the Fe-S

cluster domain in strand displacement. Such a destabilisation of the cluster has severe clinical relevance in humans (Rudolf *et al.*, 2006). It was proposed that this cluster has a structural role in stabilising the domain responsible for duplex disruption.

The *S. solfataricus* XPB homologues, XPB1 and XPB2, were the original focus of this project, the purification and characterisation of which are detailed in chapters 3 and 4.

Eukaryal and euryarchaeal XPF show a similar domain organization, both possess an N-terminal helicase and a C-terminal nuclease domain. The crenarchaeal homologue consists only of the C-terminal nuclease domain and has an absolute requirement for the functional interaction with the heterotrimeric sliding clamp PCNA (Roberts *et al.*, 2003). All XPF homologues show obligate dimerisation, for example, eukaryotic XPF forms a heterodimer with ERCC1 (Park *et al.*, 1995) whereas archaeal XPF exists as a homodimer (Roberts and White, 2005). Archaeal and eukaryotic XPF both show sequence preference for cleavage at pyrimidines and both cut a 3' flap DNA structure with the utmost specificity (Roberts *et al.*, 2003; Roberts and White, 2005).

Archaea possess an orthologue of the eukaryotic Fen-1 protein. Putative archaeal Fen-1 proteins contain sequences homologous to the N-terminal and central regions of eukaryotic Fen-1. The C-terminal region of archaeal Fen-1 contains conserved residues that mediate interactions with heterotrimeric PCNA (DiRuggiero *et al.*, 1999). Archaeal Fen-1 proteins show specificity for 5'-3' ssDNA flap substrates (DiRuggiero *et al.*, 1999). A double flap DNA substrate shown to be optimal for human Fen-1 *in vivo* was also cleaved by archaeal Fen-1 (Friedrich-Heineken and Hubscher, 2004). There is no evidence, as yet, for TCR in archaea, it is plausible, however, that lesion-induced stalling of the archaeal RNA polymerase could signal damage as described for the eukaryotes. Moreover, no archaeal homologues of bacterial Mfd or eukaryotic CS proteins have been identified.

1.3 DNA replication

1.3.1. DNA replication in a bacterial system

The bacterial origin of replication, *oriC*, is usually positioned within AT rich regions of the genome (Mott and Berger, 2007). The replication origin of *E. coli* contains

multiple DnaA boxes (9 bp consensus sequence) to which the AAA+ (**A**TPase **a**ssociated with diverse cellular **a**ctivities) ATPase DnaA binds, inducing bending in the DNA. This causes local unwinding of *oriC*, which exposes high affinity DnaA-ATP binding sites on the ssDNA. Binding of DnaA-ATP to these sites stablises the melted state (Robinson and Bell, 2005). The DnaA gene is located next to the origin of replication and it contains high affinity DnaA boxes, binding of DnaA-ATP to which reduces its expression, maintaining a tight control of this initiator protein. DnaB is a replicative hexameric helicase that loads onto the DNA at the melted origin. A second AAA+ ATPase, DnaC, controls the helicase activity of DnaB. DnaC-ATP binds tightly to the DNA and effectively stalls DnaB. ATP hydrolysis releases this block (Robinson and Bell, 2005).

The DNA polymerase III holoenzyme (figure 1.5) consists of the α polymerase, the 3'-5' proofreading exonuclease, ε, and the θ subunits. It is an asymmetric dimer, with both a leading and lagging strand polymerase (McHenry, 2003). Progression of the core itself is slow and, therefore, to become an efficient replicase it requires the β-clamp. The clamp encircles the DNA and is pulled behind the core as the DNA is extended (Johnson and O'Donnell, 2005). The closed β-clamp requires an ATP powered 'clamp loader' complex to open it and lock it onto the DNA. The clamp loader is a heptamer consisting of χ, ψ and the DnaX subunit, which comprises five subunits (τ,τ,γ,δ and δ'). The τ protein is responsible for interacting with DnaB at the replication fork and the α subunit of the holoenzymes to hold the two core polymerases together in order for them to function in a coordinated fashion (Flores *et al.*, 2001; McHenry, 2003).

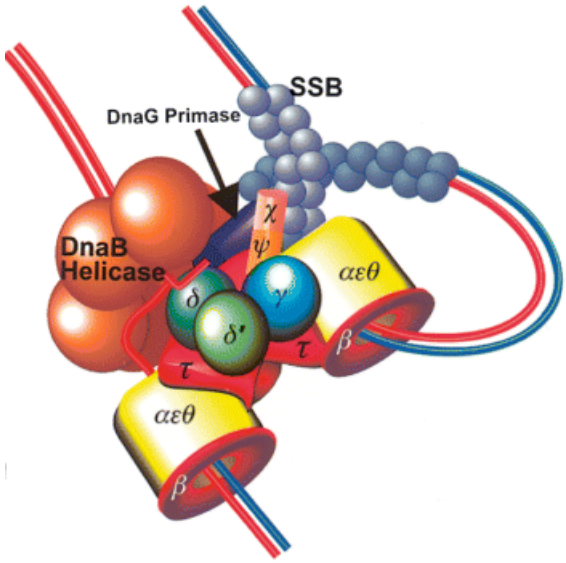


Figure 1.5. A model of the Polymerase III holoenzyme
The α polymerase, the 3'-5' proofreading exonuclease ε, and θ subunits constitute the polymerase III holoenzyme. It is an asymmetric dimer with both a leading and lagging strand polymerase, which requires the β-clamp for it to become an efficient replicase. The closed β-clamp requires an ATP powered 'clamp loader' complex (τ, τ, γ, δ, δ', χ and ψ) to open it and lock it onto the DNA. The replicative helicase DnaB loads onto the lagging strand and is coupled to DNA synthesis by the τ protein subunit of the clamp loader. τ interacts with both α subunits of the holoenzyme to ensure that the two polymerases function in a coordinated fashion (McHenry, 2003).

Lagging strand synthesis is a discontinuous process, which is achieved by the production and maturation of Okazaki fragments. Each fragment is initiated by the action of the primase DnaG, which requires interaction with the replication fork helicase DnaB (LeBowitz and McMacken, 1986; Zechner *et al.*, 1992; Hiasa and Marians, 1994). DnaB is able to migrate along the lagging strand template to unwind duplex DNA, ahead of the leading strand, and to initiate RNA primer synthesis for the nascent Okazaki fragment (LeBowitz and McMacken, 1986). DnaG is required to couple leading and lagging strand synthesis from *oriC*. Without the interaction between DnaB helicase, DnaG primase and the polymerase III holoenzyme complex, rolling circle-type replication dominates (Hiasa and Marians, 1994). Contact between DnaG and the lagging strand polymerase is crucial to limit primer synthesis and to control primer length. DnaG remains associated with the 3' nascent end of the primer until the polymerase is available to extend it. Signals to the polymerase to terminate synthesis of the previous and to commence the next Okazaki fragment are also achieved through DnaG-polymerase

interactions (Zechner *et al.*, 1992; Hiasa and Marians, 1994). DnaG and DnaB also interact with SSB to maintain their position at the RNA primed site (Johnson and O'Donnell, 2005), this ensures DnaB does not unwind the DNA before replication can occur (Hiasa and Marians, 1994). An interaction between DnaA and DnaG would position the primase to also prevent DnaB unwinding the template DNA before replication fork formation. The χ subunit of the clamp loader competes for SSB binding, causing DnaG to dissociate and allowing the clamp loader to position the β-clamp on the lagging strand primer for replication to ensue. The two polymerases, linked by the β-clamp, move in opposite directions; as the lagging strand polymerase extends the primer a loop of DNA is formed. Once the Okazaki fragment is complete, the core is released in order for it to extend the next primer (Johnson and O'Donnell, 2005). Prior formation of a multisubunit complex, containing γ, δ, β and ATP, at the primed template is required for rapid binding of the polymerase III holoenzyme to initiate elongation of the primer (O'Donnell *et al.*, 1987). The polymerase on the leading strand extends the DNA continuously as the replisome advances (McHenry, 2003).

1.3.2. Eukaryotic DNA replication

The core replicative components of the eukarya are functionally and structurally similar to those of the bacteria, although the complexity is far greater. Eukaryotes contain a much greater network of proteins involved in propagation and regulation of DNA replication (Johnson and O'Donnell, 2005).

Chromosomal DNA is replicated, through multiple origins, during S-phase of each cell cycle. Control of replication, to ensure that it does not occur more than once per cell cycle, is achieved by the alternation between two cellular states. The first promotes and the second inhibits the assembly of the protein complexes at the origin (Dutta and Bell, 1997). Eukaryotic origins, however, are difficult to identify since they rarely contain obvious sequence motifs and replication can initiate from many origins (Robinson and Bell, 2005).

The origins of replication often overlap sequences recognised by initiator proteins. These are referred to as replicators and contain a number of short sequence elements, including the ACS (ARS (autonomously replicating sequence) consensus sequence),

which is conserved at sequence level, as well as the additional and more divergent B-elements (Dutta and Bell, 1997).

The six subunit eukaryotic origin recognition complex (ORC) acts as the initiator protein. It interacts with both the ARS and B elements (Bell and Dutta, 2002) and recruits additional proteins forming the pre-replicative complex (pre-RC). ATP binding maintains a stable ORC-DNA interaction (Dutta and Bell, 1997) and double stranded DNA at the origin inhibits ATP hydrolysis, retaining ORC in the ATP bound state.

Cdc6, a member of the AAA+ ATPase family, plays a key role in assembling proteins for replication initiation. It displays a requirement for ORC-ATP in order to associate with the DNA (Bell and Dutta, 2002). Chromatin licensing and DNA replication factor (Cdt1) associates with the C-terminus of Cdc6 and together they promote recruitment of the Minichromosome Maintenance (MCM) proteins. Cdt1 is also the target of replication inhibitor geminin, which is involved in the prevention of re-replication (Bell and Dutta, 2002; Tada, 2007). Since ORC and Cdc6 are dispensable upon origin firing and can then be removed from the complex, it seems that the priority of the pre-RC is to load the putative MCM complex (Mcm 2-7) (Perkins and Diffley, 1998; Fujita *et al*., 1999; Weinreich *et al*., 1999; Bell and Dutta, 2002). The MCM proteins are loaded onto the chromatin in an ORC/Cdc6 dependent manner. In order to ensure that the genome is replicated one per cell cycle the binding of the MCM complex to the chromatin is negatively controlled by Cdc2 kinase (Perkins and Diffley, 1998; Fujita *et al*., 1999).

Single stranded DNA, formed as a result of a downstream event, stimulates ATPase activity of ORC and release of any bound replication factors. Such events could assist the transition from replication initiation to elongation (Bell and Dutta, 2002). MCM proteins are members of the AAA+ family of ATPases and are thought to possess the helicase activity needed for initiation and elongation. Subunits 4, 6 and 7 are the catalytic domains whereas subunits 2, 3 and 5 fulfil a more regulatory role (Chen *et al*., 2005b). The associated factors, Cdc45 and GINS, are fundamental for the replication apparatus to bind to the origin (Branzei and Foiani, 2007). GINS interact with MCM and Cdc45 and are responsible for including MCM in the replisome (Perkins and Diffley, 1998; Fujita *et al*., 1999; Weinreich *et al*., 1999; Barry and Bell, 2006).

MCM proteins are the targets of the kinase Cdc7/Dbvf4p (DDK), the catalytic subunit of DNA polymerase α/primase and Cdc45p. DDK is the trigger of DNA replication initiation (Bell and Dutta, 2002).

There are three polymerases involved in eukaryotic replication. DNA polymerase α/primase couples primase and polymerase activities (Barry and Bell, 2006). It binds to the ssDNA template, catalyses primer formation and then extends the primer by approximately 20 nucleotides (Garg and Burgers, 2005). A stable ATP driven replication factor C (RFC)-PCNA complex on the ssDNA has been proposed to mediate a switch from polymerase α to the lagging strand polymerase δ, by dissociating the priming polymerase. ATP hydrolysis causes RFC to exit the complex leaving PCNA, bound to DNA as a closed ring (Johnson and O'Donnell, 2005), to stabilise the elongation complex. Synthesis of the Okazaki fragments occurs very quickly by polymerase δ (Garg and Burgers, 2005). Nicks that can be ligated by ligase I must be maintained by polymerase δ and Fen-1. When polymerase δ arrives at a dsDNA region, efficient nick translation occurs, which is terminated by a ligation reaction (Garg and Burgers, 2005).

The third polymerase is the leading strand enzyme, polymerase ε, which loads onto the DNA before primer synthesis employing a non-polymerase function. Despite its high processivity, it does interact with PCNA, although this does not have much effect on the activity of polymerase ε *in vitro*. The interaction may be more important for the recruitment of polymerase ε to sites of DNA damage rather than required for a polymerase processivity during DNA replication (Dua *et al*., 2002; Garg and Burgers, 2005). The minimal set of proteins required for DNA replication is shown in figure 1.6.

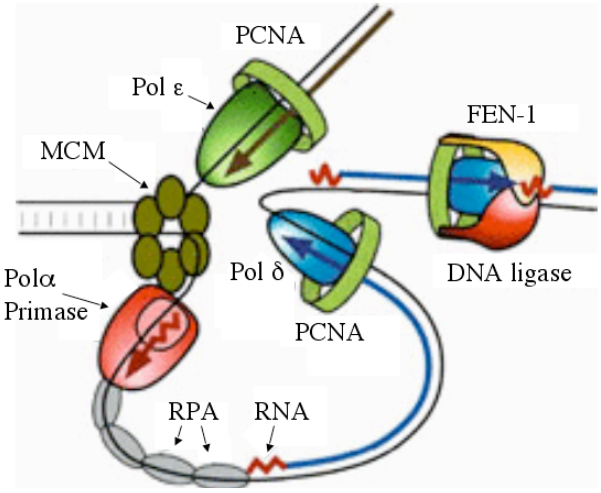


Figure 1.6 A cartoon representation of the eukaryotic replication fork
ORC, Cdc6 and Cdt1 assemble the MCM helicases onto the DNA, ORC and Cdc6 are then removed. Polymerase α/primase catalyses primer formation and extends the primer by approximately 20 nts. The switch from pol α/primase to the lagging strand polymerase δ is mediated by the RFC-PCNA complex. ATP hydrolysis by RFC releases it from the DNA, leaving a stable initiation complex to extend the Okazaki fragment. Polδ and Fen-1 maintain the nicks at the end of each fragment, which are sealed by DNA ligase. Leading strand replication is achieved by the continuous progression of polymerase ε (Garg and Burgers, 2005).

1.3.3. Archaeal DNA replication

The archaeal replication machinery is a less complicated version of the eukaryotic system with a much simpler organisation of a number of proteins homologous to essential eukaryotic factors (Cann and Ishino, 1999; Barry and Bell, 2006). Archaea have multiple origins; *S. solfataricus*, for example, has three replication origins, which are fired together (Kelman and White, 2005; Barry and Bell, 2006). They contain inverted repeat sequence elements, origin recognition boxes (ORB), that are specifically bound by Cdc6 proteins. Some archaea possess one Orc1/Cdc6 homologue, whereas others have multiple Orc1/Cdc6 homologues possibly with individual roles at initiation (Dionne *et al.*, 2003b). *S. solfataricus*, for example, has three Orc1/Cdc6 homologues referred to as Cdc1-3. All three show sequence similarity to both Orc1 and Cdc6, which indicates that the *S. solfataricus* proteins possess a combination of functions (Duggin and Bell, 2006). The MCM proteins are the primary replicative helicases, the activities of which are inhibited by archaeal Cdc6 proteins. This is reminiscent of bacterial DnaC and DnaB and, therefore, Cdc6 could function as the helicase loader (Kelman and White, 2005). Many archaea, in contrast to the eukaryotes, possess only one homologue of the MCM

proteins. The MCM complex exists as a homomultimeric complex, predominantly forming hexamers and double hexamers in solution (Chen *et al*., 2005b; Gomez-Llorente *et al*., 2005; Duggin and Bell, 2006).

The eukaryotic primase consists of a small catalytic subunit, PriS, and a large non-catalytic, regulatory subunit, PriL (Barry and Bell, 2006). Although *S. solfataricus* possesses a homologue of the eukaryotic primase, most archaea only contain the core PriS and PriL subunits (Duggin and Bell, 2006). Together the PriSL heterodimer possesses two regions that partake in regulating the length of the RNA primer and controlling its overall catalytic activity (Duggin and Bell, 2006). Both the euryarchaea and the crenarchaea posses B family polymerases (replicative polymerases), but euryarchaea also contain D family polymerases. So far this family seems to be unique to euryarchaea and is not very well characterised (Burgers *et al*., 2001; Barry and Bell, 2006). *S. solfataricus* has three B-family polymerases, B1-B3, which may be involved in various aspects of replication. The structure of polymerase B1 presents features likely to be specific to hyperthermophiles. The N-terminus has the archaeal template-scanning motif, this is able to recognise and pause at deaminated cytosine bases 3-4 nucleotides ahead in the template strand using a so-called 'read ahead' mechanism (Gruz *et al*., 2003; Shuttleworth *et al*., 2004; Duggin and Bell, 2006). The *Sulfolobus* species also possess a Y family polymerase, Y1. Polymerase B1 arrests upon encountering a lesion and polymerase switching enables polymerase Y1 to resume synthesis. Polymerase Y1 is specialized in replication, it is able to bypass the lesions in an error free manner (Gruz *et al*., 2003). It is particularly important for the hyperthermophiles to efficiently remove uracil from the DNA template before replication since elevated temperatures greatly increase cytosine deamination (Shuttleworth *et al*., 2004).

PCNA interacts with the DNA polymerase tethering it to the DNA to enhance its processivity. Archaeal PCNA is a heterotrimer, each subunit interacts with different proteins which allows coordination of multiple factors simultaneously (Dionne *et al*., 2003a). DNA ligase I and Fen-1 both interact with PCNA and participate in replacing the RNA primers in the lagging strand with DNA (Duggin and Bell, 2006). Consistent with the eukaryotic system, PCNA requires a clamp loader. RFC consists of a single large

subunit (RFC_L) and four subunits (RFC_S) that are identical to each other, and is responsible for loading PCNA onto the DNA.

1.3.4 Stalled replication fork restart

It is very common for replication forks to collapse at sites of UV damage, but they also routinely stall under normal growth conditions (Cox *et al.*, 2000). DNA lesions or gaps in the DNA backbone; direct chemical damage to the DNA template; protein complexes on the DNA, such as transcription apparatus; nucleosomes and DNA secondary structures are all responsible for blocking the progression of the replication fork (McGlynn and Lloyd, 2002). The stalled fork must be reset and the damage repaired for the replication process to proceed and ensure accurate transmission of the genetic code (McGlynn and Lloyd, 2002). Progression of the replicating polymerase through a single strand break results in a DSB (Morimatsu and Kowalczykowski, 2003).

In *E. coli*, replication restart can be accomplished by a variety of pathways; the initial step involves regression of the stalled fork by the helicase RecG to form a HJ intermediate (McGlynn *et al.*, 2001). This requires unwinding of the newly synthesised strands from their templates, annealing of the nascent strands together and reannealing of the parental duplex (McGlynn and Lloyd, 2002). Single strand lesions can be overcome without cleavage, however, if the lesion affects both strands of the template DNA, then HJ cleavage and homologous recombination is necessary for replication restart (McGlynn and Lloyd, 2000, 2002). Formation of a new primosome by the PriA or PriC pathway is an essential process for replication restart (Gregg *et al.*, 2002; Heller and Marians, 2005).

Lesions on the lagging strand template do not pose too much of a problem, since the Okazaki fragment can terminate early and extension from the downstream primer can continue; the resulting gap is repaired by RecA-dependent mechanisms. Damage to the leading strand, however, poses a serious threat to replication. Unlike the lagging strand, priming by the 3' end of the leading strand is required for its continuous replication and cannot be initiated downstream of the lesion (Gregg *et al.*, 2002).

RecA dependent HR is also achieved via the RecF pathway. RecFOR functions to promote RecA nucleation onto SSB coated gapped DNA, either by changing the conformation of the SSB-DNA complex or by replacing SSB with RecA (Umezu *et al.*, 1993; Morimatsu and Kowalczykowski, 2003). The RecF pathway can also function on

DSBs if RecG is inactive. RecFOR requires a base paired 5' terminus at the ss-dsDNA junction. RecQ helicase acts both to initiate HR via this RecFOR pathway and to suppress illegitimate recombination (Bennett and Keck, 2004). A common feature of this family of RecQ proteins is the combined action of the helicase with a nuclease; in this case the RecQ helicase works in concert with the RecJ exonuclease to process the DNA end to form a 3' ssDNA tail appropriate for the loading of RecFOR (Morimatsu and Kowalczykowski, 2003; Bennett and Keck, 2004).

In eukaryotes Rad54 is a core protein of the *Rad52* epistasis group; these genes encode proteins that are essential and form the core machinery for HR (Heyer *et al.*, 2006). Rad54 is a member of the Swi2/Snf2 chromatin remodeling family of proteins that possesses dsDNA dependent ATPase activity. Proteins of the Swi/Snf2 family are unable to separate duplex DNA strands, rather they are regarded as motor proteins that translocate on duplex DNA remodeling protein-DNA complexes (Heyer *et al.*, 2006). Rad54 has been implicated in all three stages of HR: pre-synapsis, synapsis and post-synapsis (described in section 1.2.1), although no bacterial homologue is known (Mazina and Mazin, 2004; Heyer *et al.*, 2006). During pre-synapsis Rad54 promotes the formation of a Rad51 nucleofilament on ssDNA bound by RPA and forms a co-complex with the filament to stabilise it. This is independent of its ATPase activities (Heyer *et al.*, 2006). Rad54 stimulates DNA strand exchange and Rad51-mediated joint molecule (D-loop) formation during synapsis. The ATPase activity of Rad54 is important at this stage; ATP dependent translocation of Rad54 on duplex DNA is thought to be vital for searching the DNA for homologous regions. The events after D-loop formation are referred to as post synapsis. During this stage, Rad54 is involved in migrating the HJ, formed as a result of strand exchange, in order to extend the heterologous DNA regions (Bugreev *et al.*, 2006; Heyer *et al.*, 2006). In addition, Rad54 functions to dissociate Rad51 from the DNA. In contrast to the role in pre-synapsis, branch migration and Rad51 turnover are catalysed by Rad54 in an ATP dependent manner (Heyer *et al.*, 2006).

Two Rad51/RecA homologues were identified in *P. furiosus*, RadA and RadB. RadA is a DNA dependent ATPase and possesses DNA pairing and strand exchange activities. RadB, however, is a weak ATPase and exhibits little strand exchange activity.

Archaeal RadB may, therefore, be the functional counterpart of eukaryotic Rad55-Rad57, a complex that promotes RPA-associated strand exchange achieved by Rad51 (Komori *et al*., 2000). RadB interacts with and inhibits the archaeal-specific Holliday junction resolvase Hjc, ATP dependent release of RadB activates Hjc and results in HJ resolution (Komori *et al*., 2000).

RecA, Rad51 and RadA are highly conserved in regions that bind to and hydrolyse ATP (Guy *et al*., 2006). *P. furiosus* RadB, however, is confined to the euryarchaea and is more accurately described as a RadA or Rad51 paralogue.

S. solfataricus does not possess RadB; a RadA protein and a less conserved RadA paralogue have been identified. Putative RadA proteins are more similar to Rad51 family (approximately 40 % amino acid sequence identitiy) than RecA family (approximately 20 % identity at the amino acid sequence level) proteins (Sandler *et al*., 1996; Abella *et al*., 2007). These RadA proteins are involved in homologous recombination at the strand-pairing phase (DiRuggiero *et al*., 1999). *S. solfataricus* possesses two HJ resolving enzymes, Hjc and Hje (Kvaratskhelia and White, 2000), although no apparent motor for branch migration has been recognised (Roberts *et al*., 2003). Two hypothetical proteins from *S. solfataricus* (Sso1289 and Sso1468) were the subjects of a BLASTP analysis, which identified members of AAA ATPase protein family as the closest matches to both Sso1289 and Sso1468. The identification of a possible branch migration protein is discussed further in chapter 6.

S. solfatricus possesses a homologue of Hel308, a helicase involved in replication restart. Possible mechanisms by which *S. solfataricus*, strain PBL2025, Hel308 clears the lagging strand at a stalled replication fork are proposed in chapter 5.

1.4 Identification and organisation of helicases

Helicases have a major involvement in DNA metabolism, including transcription, recombination and repair, and have been linked to a number of human genetic disorders. In humans, defects in the RecQ family helicases WRN, BLM and RecQ4 are characterised by genomic instability and an increased incidence of cancer. They are associated with the rare autosomal disorders Werner's Syndrome (WS), Bloom's Syndrome (BS) and Rothmund Thomson Syndrome (RTS), respectively (Bennett and

Keck, 2004; Opresko *et al.*, 2004). Helicases function to catalyse nucleic acid duplex separation and this action is dependent upon the hydrolysis of a nucleoside 5'-triphosphate (NTP) (Hall and Matson, 1999). A helicase will usually bind to single stranded DNA or to a ssDNA/dsDNA junction and will translocate unidirectionally along the strand it has loaded onto. On the basis of the direction of translocation they can be divided into 2 types, 5'-3' or 3'-5' helicases. The protein sequences contain conserved motifs, found clustered in the 'core' region of the three-dimensional protein, that act as the engine to power this unwinding activity (Hall and Matson, 1999).

The presence of these motifs, however, is not sufficient to identify helicases since only a minority of these proteins actually possess true helicase activity (Singleton and Wigley, 2002). This may be because these proteins require oligomerisation, protein interactions or activation such as phosphorylation, for example. It is also true that these motifs are characteristic of translocases as the molecular motors themselves may not be enough to achieve strand separation and this may be the function of additional domains (Singleton and Wigley, 2002). Proteins involved in chromatin remodelling have been shown to contain the helicase motifs in their sequence as they possess the ability to disrupt the interactions between DNA and histones rather than unwind duplex DNA. It is, therefore, widely accepted that these motifs are typical of NTPases rather than helicases specifically.

1.4.1 The organisation of helicases into families

The helicases were originally classified into three superfamilies SF1, SF2, SF3 and two smaller families F4 and F5, based on the similarity and organisation of conserved 'signature motifs' and primary structure analysis (Gorbalenya and Koonin, 1993; Hall and Matson, 1999; Tuteja and Tuteja, 2004). More recently it has become apparent that not all of the putative helicases catalyse strand separation. Instead, they couple NTP hydrolysis to translocation along nucleic acids. In addition, it has been suggested that the AAA+ ATPase proteins are defined as Superfamily 6 (SF6) since many nucleic acid motors are members of this family (Singleton *et al.*, 2007). Consequently, the helicases are defined by the six superfamilies shown in figure 1.7. The N- and C-terminal regions show much sequence and length variability, for this reason these regions are thought to

be responsible for the individual functions exhibited by the helicases (Tuteja and Tuteja, 2004).

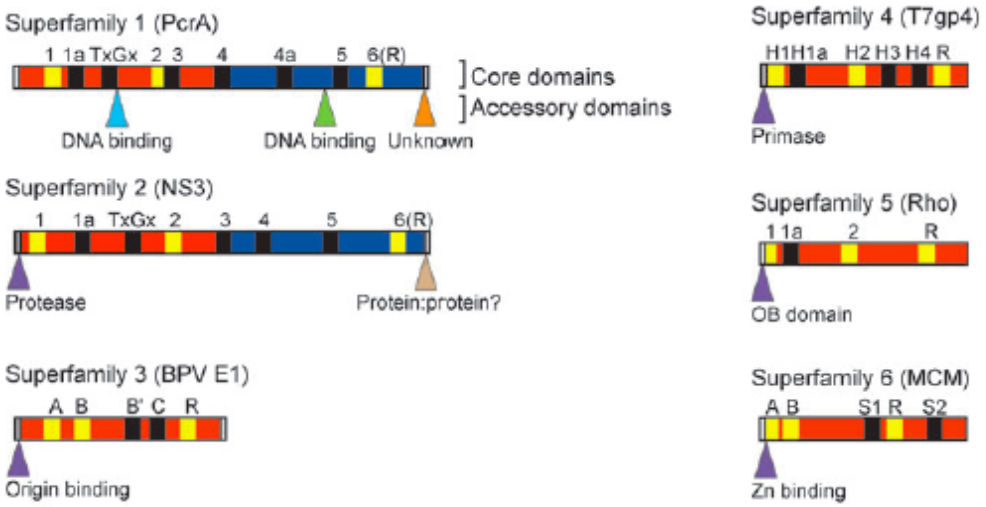


Figure 1.7. Classification of NTPases into six Superfamilies
Nucleic acid motors are classified into six Superfamilies based on specific core motifs. These core domains and the signature motifs for each family are shown, however, the precise position is based on the example shown and is representative of the whole family. The positions and functions of the accessory domains are shown but these are specific for this particular example, their location, function and even presence may vary between members of the same family. Universal structural elements common to all helicases are shown in yellow (Singleton *et al.*, 2007).

Superfamilies 1 and 2 are the largest of the families and are closely related (figure 1.7). Helicases from both families contain the seven typical short amino acid sequence fingerprints, which are very similar in sequence, arrangement and secondary structure (Hall and Matson, 1999). These conserved helicase motifs are usually clustered in a core region of 200-700 residues in the amino acid sequence (Hall and Matson, 1999; Tuteja and Tuteja, 2004). The tandem RecA like folds found either in the same polypeptide or between subunits are formed from core domains. The Walker A and B boxes, involved in nucleotide binding and hydrolysis, and an arginine finger that plays a role in energy coupling are universal features of these core domains (Singleton *et al.*, 2007). Many ATPases share this common core RecA like fold, which contains the ATP binding site. This fold was initially identified in *E. coli* RecA protein and is, therefore, referred to as the RecA-like fold or RecA motor domain (Ye *et al.*, 2004; Bell, 2005). Two RecA-like domains are found in the SF1 and SF2 helicases but this number varies among different ATPases (Story *et al.*, 1992; Ye *et al.*, 2004; Singleton *et al.*, 2007). SF3-6 members

form hexameric or double hexameric rings and, therefore, possess six or twelve RecA-like folds (Singleton *et al.*, 2007).

The proximity of all of the motifs to each other and the NTP implicates them all in nucleotide binding and hydrolysis and possibly the coupling of energy from hydrolysis to DNA unwinding, despite not all motifs making physical contact with the bound nucleotide (Hall and Matson, 1999). Structural data for the two SF1 helicases, *B. stearothermophilus* PcrA and *E. coli* Rep, has shown that the seven motifs occupy positions in domains 1A and 2A of the four-domain proteins and that these two domains are located at the base of the enzymes forming a nucleotide binding pocket (Hall and Matson, 1999). Motifs I, Ia, II, V and VI of the SF2 helicase NS3 appear to occupy the same relative positions compared to the SF1 helicases and they fulfil similar functions in both superfamilies (Hall and Matson, 1999). Only motifs III and IV show major differences between the NS3 and the two SF1 proteins Rep and PcrA. The motifs are, therefore, well conserved across the two superfamilies to act together as the engine to power the enzyme. Both SF1 and SF2 contain ssDNA 3'-5' and 5'-3' motors, however, no dsDNA motors are found in SF1. The specific function of each helicase is achieved by a degree of variation between proteins (Hall and Matson, 1999). The XPB proteins and Hel308 from *S. solfataricus* are members of SF2 and are discussed in detail in chapters 3 and 5, respectively.

The helicases categorised into SF3 are limited to those found in DNA and RNA viruses, this group is much smaller than SF1 and 2. They form hexameric helicases that translocate along DNA (3'-5') threading the ssDNA through the central channel of the ring (Singleton *et al.*, 2007). The members of SF3 contain fewer conserved motifs in their peptide sequence; the Walker A and B boxes, the B' box and the sensor 1 motif. Based on data obtained for AAV type 2 (AAV2) Rep40, it was proposed that the protein must oligomerise in order to form a catalytic nucleotide-binding site (James *et al.*, 2004). Therefore, ATP binding acts to regulate the oligomeric state of this helicase. This may be a characteristic of SF3 helicases (James *et al.*, 2004). The B' motif, found only in SF3 helicases, is important for the helicase reaction since it encompasses residues involved in direct interactions with the nucleotide at the oligomeric interface and has been proposed to bind ssDNA necessary for helicase activity (Koonin, 1993; Yoon-Robarts *et al.*, 2004).

Motif B' couples DNA binding and ATP hydrolysis and possesses residues with a direct involvement in DNA interactions (James *et al.*, 2004).

The helicases of the fourth family are related to bacterial DnaB. They are associated with a primase, both functionally and physically, either as two distinct proteins that interact or as two domains of the same protein (Gorbalenya and Koonin, 1993). They contain only five conserved motifs (H1, H1a, H2, H3 and H4), including the Walker A (H1) and B (H2) boxes, and unwind DNA by threading the single stranded portion through the centre of a hexameric ring structure (5'-3') (Caruthers and McKay, 2002; Donmez and Patel, 2006). The most extensively studied SF4 helicase is bacteriophage T7 gene 4 protein (T7gp4) the helicase mechanism of which is discussed in section 1.5.3.

There is also a fifth group that contains helicases (5'-3') such as Rho, the bacterial transcription terminator factor (Singleton and Wigley, 2002). Rho is a DNA-RNA helicase with significant sequence similarity to proton translocating ATPase, rather than other helicases (Gorbalenya and Koonin, 1993). SF6 includes proteins such as the MCMs that contain the AAA+ fold but do not fall into SF3 (Singleton *et al.*, 2007).

1.4.2 The helicase 'signature' motifs of SF1 and SF2 helicases

There is strict conservation of the Walker A and Walker B box sequences across the helicases families, motif VI is the third most conserved, whereas motifs Ia, III and V show much more sequence variation among different families (Gorbalenya *et al.*, 1989; Tuteja and Tuteja, 2004).

The Walker A box (motif I) is identified by the consensus sequences AxxGxGKT (SF1) and GxxxxGKT/S (SF2); the final three residues are the minimal requirement (Gorbalenya *et al.*, 1989). The side chain of the invariant lysine contacts the β-phosphate of the NTP upon binding to stabilise the transition state during the hydrolysis reaction (Gorbalenya *et al.*, 1989; Hall and Matson, 1999; Caruthers and McKay, 2002; Tuteja and Tuteja, 2004). The Walker B box (Motif II) generally takes the form of DExx across the two major superfamilies. Proteins in SF2 with this motif are also referred to as the DEAD-box proteins, as their signature motif consists of the consensus Asp-Glu-Ala-Asp (Gorbalenya *et al.*, 1989; Pause and Sonenberg, 1992). The Walker B box is thought to activate the attacking water molecule and, more specifically, the carboxyl portion of the

aspartic acid coordinates the Mg^{2+} ion in order to do this. The glutamic acid is important for ATP hydrolysis (Hall and Matson, 1999; Caruthers and McKay, 2002).

The sequence of motif VI exhibits covariation with the Walker B box; proteins with the sequence DEAD in motif II have the signature 'HxxGRxxR' in motif VI and those with 'DExH' consensus have 'QxxGRxxR' (Gorbalenya *et al.*, 1989). Motif VI is responsible for mediating nucleotide binding induced conformational changes.

1.5 Mechanisms of strand separation

The mechanism of strand displacement can be classed as either passive or active. This depends on whether the helicase directly participates in the unwinding event or merely stabilises the resulting ssDNA.

The passive mechanism describes the indirect DNA unwinding achieved by binding of the helicase to the ssDNA that is formed by transient fraying at a ss/dsDNA junction due to thermal fluctuations (figure 1.8) (Chen *et al.*, 1992). At this junction, the helicase would trap the ssDNA and net unwinding would, therefore, be coupled to translocation with a step size of one base, or a few base pairs at most (Lohman and Bjornson, 1996).

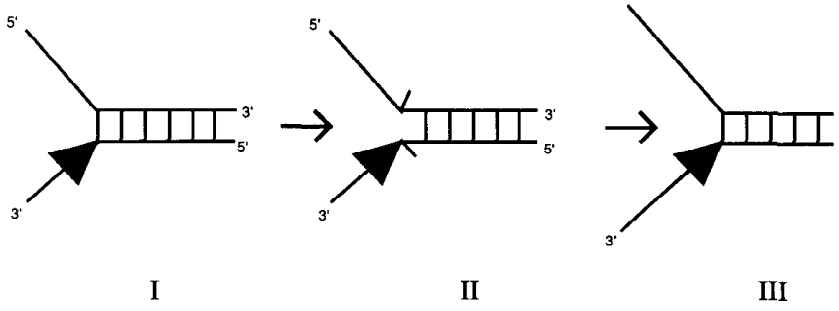


Figure 1.8 A model describing a passive mechanism of helicase unwinding
The helicase translocates along the ssDNA to a ss/dsDNA junction. Here the helicase is unable to bind the dsDNA but it can trap the ssDNA formed by transient unwinding induced by thermal fluctuations. Net unwinding is coupled to translocation of a few base pairs at most (Lohman and Bjornson, 1996).

Helicases that have a direct role in destabilising the duplex use an active mechanism; conformational changes actively promote destabilisation of the helix resulting in strand separation (Vindigni, 2007). Active mechanisms include the 'inchworm' and 'active rolling' models and the model for hexameric helicases, these are described in the following sections.

1.5.1 The 'Inchworm' model

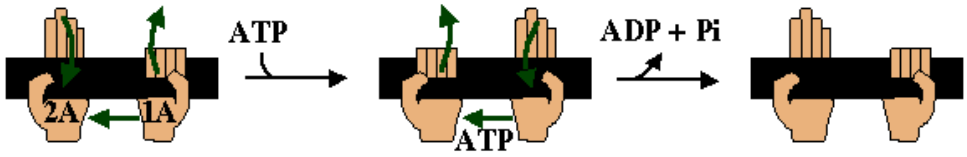
Studies of *E. coli* Rep protein led to the first proposal of the 'inchworm' mechanism of strand unwinding (Yarranton and Gefter, 1979). This model, however, has been refined since Rep was identified as a homodimer upon DNA binding (Wong *et al.*, 1992; Lohman and Bjornson, 1996).

The 'inchworm' model requires coordinated alternate binding of the nucleic acid at two different sites within the functional unit of the helicase. One site is required to bind ssDNA (tail site) and the other to bind both ssDNA and duplex DNA (leading site) (Lohman and Bjornson, 1996). This can be accomplished by a monomer containing two binding sites, or a dimer with a single binding site per subunit. Translocation along the ssDNA is coupled to ATP binding, whereas the unwinding reaction is driven by ATP hydrolysis (Tuteja and Tuteja, 2004).

Structural characterisation of *Bacillus subtilis* PcrA has led to a more detailed mechanistic view of DNA unwinding by the 'inchworm' model (Mackintosh and Raney, 2006). The helicase reaction can be divided into two processes, the translocation of PcrA along ssDNA and duplex destabilisation. It is the coupling of these that constitutes helicase activity (Velankar *et al.*, 1999).

Initially ssDNA is bound to both domains 1A and 2A. Upon ATP binding the conformation of PcrA changes and it adopts a 'closed' arrangement. At this stage domain 2A is bound to the ssDNA and 1A has loosened its grip to slide along the strand. Hydrolysis of the ATP causes domain 1A to grip the DNA tightly, now 2A releases the strand and translocation is affective across it. The protein is said to be in an open conformation (Velankar *et al.*, 1999). This process by which tight DNA binding alternates between the two domains is described in cartoon form in figure 1.9A.

A



B

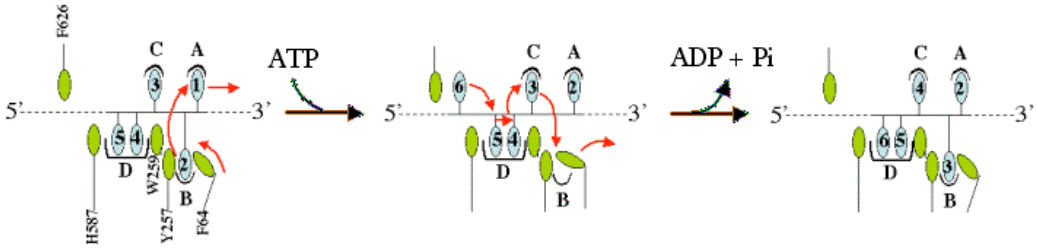


Figure 1.9 Proposed models for the translocation of PcrA along ssDNA
A) The affinity of each of the domains of PcrA, 1A and 2A, alternates during translocation and is dependent upon ATP binding or hydrolysis. ATP binding induces a closed cleft conformation, which allows 2A to bind tightly to the DNA. Upon ATP hydrolysis, the cleft opens. 2A loosens its grip on the DNA and PcrA springs back to its original conformation. 1A retains a tight grip on the DNA at this point pulling the DNA across domain 2A.
B) A cartoon representation of the base flipping events that occur during translocation of PcrA along ssDNA. Domain 1A is tightly bound to the DNA when there is a base in binding pocket B. The side chain of F64 moves into this pocket displacing the base to pocket A, which in turn displaces the adjacent base from pocket A to the outside of the protein. ATP hydrolysis opens the cleft, now F64 is released and the base in pocket C can move to pocket B. Cleft opening now forces the base from the stacked pair in pocket D to flip to pocket C, the second of this pair moves along and the next base can join it. Both figures were adapted from Velankar, *et al*. (1999).

When domain 1A has a tight grip on the DNA, bases are in each of the acceptor pockets of 1A (figure 1.9B). When PcrA adopts the 'closed' conformation, domain 1A releases its grip on the DNA by displacement of bases in its binding pockets; the result is translation across domain 1A. Flipping of the adjacent bases into the binding pockets of domain 1A promotes tight binding of domain 1A once again to the ssDNA strand and the DNA is pulled over domain 2A (Velankar *et al.*, 1999). Step sizes of only a few to as little as one base pair are characteristic of the 'inchworm' unwinding mechanism (Velankar *et al.*, 1999).

SF2 helicases are mechanistically distinct from other helicases. In contrast to SF1 helicase that rely on the bases as the primary recognition determinant, SF2 helicases can track along the continuous ribose-phosphate backbone of the loading strand without

interacting with the bases. RNA helicase NPH-II is an example of a helicase that peels the complementary strand off during translocation (Kawaoka *et al*., 2004).

1.5.1.1 'Cooperative Inchworm' model

Some proteins, despite being monomeric, exhibit enhanced unwinding when multiple monomers are allowed to function cooperatively. The basis of the 'cooperative inchworm' model describes independently translocating monomers that cooperate upon encountering a 'challenge', for example, duplex DNA or a protein block. Multiple monomers increase the chances of overcoming such an obstacle and the more difficult the challenge, the more important this enzyme cooperativity becomes (Mackintosh and Raney, 2006). The bacteriophage T4 Dda helicase is an example of a helicase that exhibits greater helicase activity when the single stranded region of the DNA is long enough to accommodate at least two monomers of Dda. The increased activity, however, is most likely not the result of protein-protein interactions but instead simply due to the presence of other monomers, translocating in the same direction, that can prevent the first monomer slipping backwards upon encountering a block (Byrd and Raney, 2004, 2006).

1.5.1.2 'Quantum Inchworm' model

The 'quantum inchworm' model describes an unwinding mechanism by which the leading domain (L) of the helicase binds one strand of the duplex DNA ahead of the trailing domain (T), anchoring the protein to the duplex. The trailing domain advances, unwinding the DNA in small steps. As the trailing helicase domain approaches the anchored leading domain it transmits a signal, which leads to dissociation and rebinding of the leading domain to the duplex DNA ahead to restart the process (Bianco and Kowalczykowski, 2000). *E. coli* RecBC helicase is an example of an enzyme that catalyses DNA unwinding according to this model (figure 1.10)

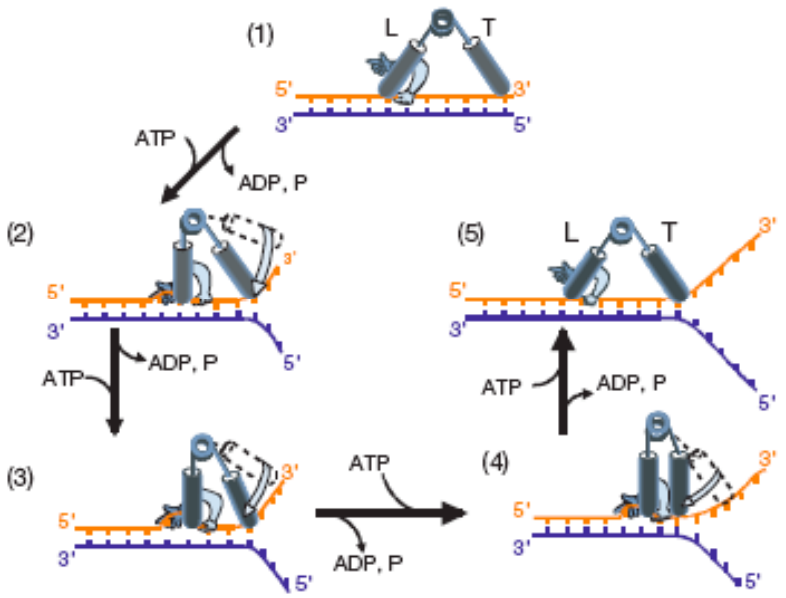


Figure 1.10 A cartoon representation of the 'quantum inchworm' mechanism employed by *E. coli* RecBC helicase
The leading domain (L) binds the DNA ~23 nts ahead of the trailing domain (T), anchoring the protein. The trailing domain then unwinds the duplex DNA in a number of small steps (2-5 nts). As T approaches, L dissociates from the strand and rebinds the DNA ~ 23 nts ahead and the process is repeated (Bianco and Kowalczykowski, 2000).

1.5.2 The 'Active Rolling' mechanism

The 'active rolling' model was proposed to describe the mechanism by which *E. coli* Rep helicase unwinds duplex DNA (Wong *et al.*, 1992). Two identical DNA binding sites are required for the 'active rolling' mechanism (figure 1.11), both of which must be able to bind ssDNA and dsDNA. At any one time during the cycle, at least one of the Rep subunits must be bound to the 3' ssDNA at the fork; the other is either bound to the same ssDNA or to the duplex beyond the fork. The alternation of these two states is controlled by the binding and hydrolysis of ATP (Lohman and Bjornson, 1996).

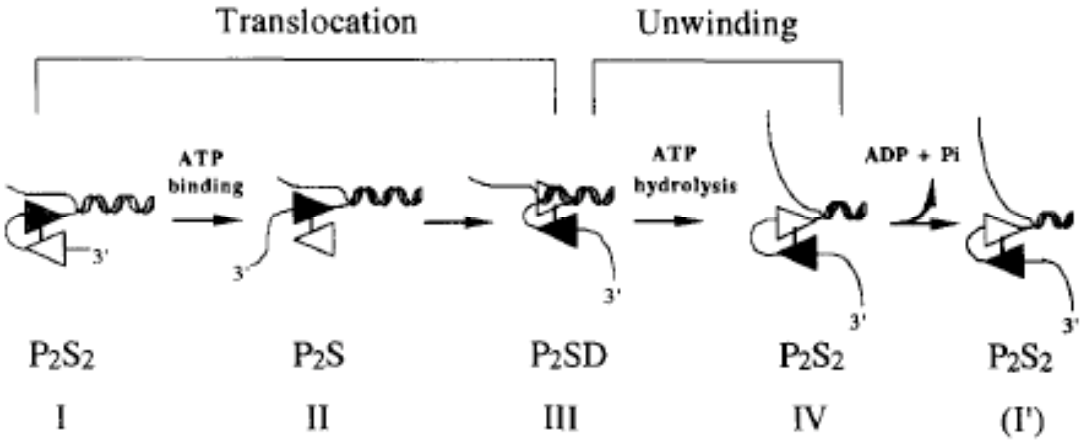


Figure 1.11 A schematic representation of the 'active rolling' mechanism
Both subunits are bound to the 3' single strand until ATP binds (P_2S_2); this causes one of the subunits to loosen its grip on the ssDNA (P_2S) and its affinity for dsDNA to increase (P_2SD). ATP hydrolysis induces conformational changes in the enzyme, which destabilise multiple base pairs in the duplex resulting in displacement of the 5' strand. All the while the second subunit is still bound to the ssDNA (P_2S_2) (Lohman and Bjornson, 1996). With each cycle the dsDNA binding role alternates between subunits.

Initially both subunits are bound to the 3' ssDNA (P_2S_2). ATP binding decreases the affinity of one Rep subunit for ssDNA, forming the P_2S intermediate, and increases its affinity for duplex DNA (Wong *et al.*, 1992). The result is a P_2SD intermediate (Lohman and Bjornson, 1996) where one subunit is bound to the 3' ssDNA and the other to the duplex DNA. Disruption of the duplex DNA is coupled to ATP hydrolysis (Bjornson *et al.*, 1996). Conformational changes in the enzyme cause duplex destabilisation and displacement of the 5' strand. Rep remains bound to the single strand, reforming the P_2S_2 complex. Release of ADP and inorganic phosphate completes this cycle (Bjornson *et al.*, 1996; Lohman and Bjornson, 1996). This process repeats itself, however, the dsDNA binding alternates between subunits with each cycle.

1.5.3 Hexameric helicases

Bacteriophage T7 (gene 4 protein) (T7gp4) is a SF4 5'-3' helicase/primase involved in T7 DNA replication (Hacker and Johnson, 1997). Upon nucleotide binding, T7gp4 helicase forms a hexamer that is able to bind ssDNA (Donmez and Patel, 2006).

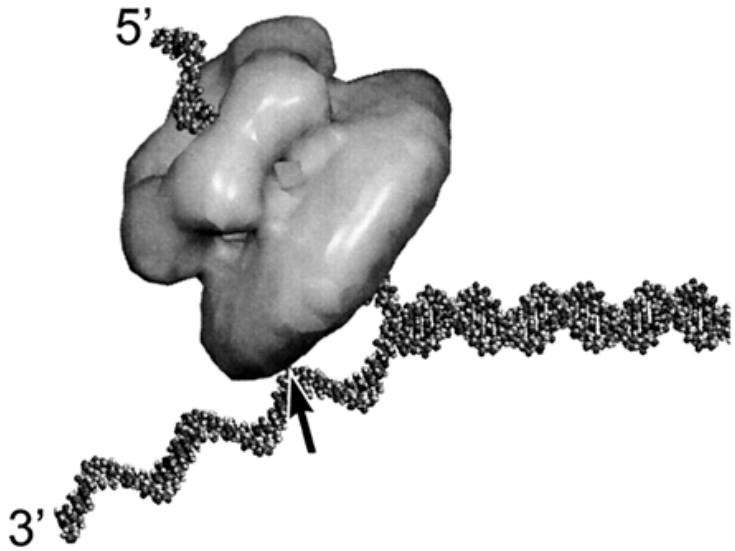


Figure 1.12 A model for T7 DNA helicase binding and unwinding DNA
The hexameric protein loads onto and encircles the 5' strand of the fork and the 3' strand is excluded from the central channel of the oligomer. Its exclusion is required for the initiation of DNA unwinding. The 3' tail interacts with a separate site on the helicase, possibly in the front of or outside the ring. The arrow designates a possible binding region on the outer surface of the hexamer (Ahnert and Patel, 1997).

The T7gp4 is structurally very flexible; a more compact conformation accompanies DNA and dTTP binding (Yong and Romano, 1995). This flexibility lends itself to a ring-opening mechanism for self-assembly around the DNA strand (Donmez and Patel, 2006). T7gp4 requires both the 3' and the 5' non-complementary ends of a DNA fork to load onto the DNA. The 5' tail binds the helicase through the central pore, whereas the 3' is excluded from the channel (Hacker and Johnson, 1997; Jeong *et al*., 2004). The 3' end, however, is thought to interact with the helicase at a separate site, possibly in front of or outside the hexameric ring (figure 1.12) (Ahnert and Patel, 1997). The T7gp4 hexamer contains six nucleotide-binding sites; docking of dTTP is the power stroke resulting in translocation (Liao *et al*., 2005). Hydrolysis of dTTP cycles through the subunits and the DNA is passed from one subunit to the next. This allows time for each subunit to recharge while waiting its turn (Donmez and Patel, 2006). The hexamer threads one strand of the duplex through the central channel of the ring excluding the second (figure 1.12) and, therefore, unwinding is described as a strand exclusion or wedge model (Donmez and Patel, 2006). However, translocation and rotation of the helicase would cause one strand to rotate around the other subsequently unwinding the dsDNA by torque on both strands (Ahnert and Patel, 1997).

Papillomavirus E1 protein (AAA+ family) is a SF3 hexameric helicase involved in melting dsDNA at origins of replication (Castella *et al.*, 2006). E1 is either loaded onto the DNA strand as an open ring or, sequentially, as monomers until the hexameric structure is formed and, in contrast to the T7gp4, translocates along the DNA using an escort mechanism (Enemark and Joshua-Tor, 2006). The hairpins responsible for protein-DNA interactions protrude into the interior of the central channel, they resemble a spiral 'staircase' with a right-handed helical arrangement (Enemark and Joshua-Tor, 2006).

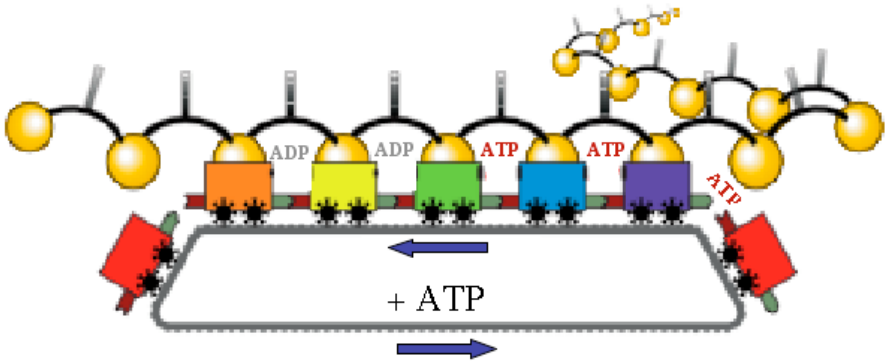


Figure 1.13 The escort mechanism by which E1 hexameric helicase translocates along DNA
Each subunit of E1 binds to a ssDNA phosphate with which it maintains contact. The subunit migrates down the 'staircase' hydrolysing ATP and subsequently releasing ADP and the ssDNA phosphate when it reaches the bottom of the 'staircase'. The subunit binds a new ATP molecule and migrates back to the top position, binding to the next available ssDNA phosphate (Enemark and Joshua-Tor, 2006).

In the model describing translocation of E1 protein, each DNA binding hairpin remains in contact with the same nucleotide throughout the cycle and the entire unit migrates with the DNA molecule (figure 1.13). The subunit binds ATP and DNA when it is located at the top of the 'staircase' (Enemark and Joshua-Tor, 2006). ATP hydrolysis drives the progression of the hairpin downwards towards the bottom of the 'staircase'. The subunit releases its associated ssDNA phosphate and ADP concluding the passage of the DNA nucleotide through the channel (Enemark and Joshua-Tor, 2006). ATP binding initiates the cycle again and the subunit moves back to the top to bind the next available ssDNA phosphate. Strand separation occurs as a result of the force produced from directional translocation and neglects the need for a base pair melting mechanism (Enemark and Joshua-Tor, 2006).

1.5.4 Helicase autoinhibition

Autoinhibition is a regulatory strategy that is widespread and it serves as a reversible barrier to prevent unnecessary activation of biological pathways and as a response to various signalling pathways (Pufall and Graves, 2002; Wang *et al*., 2006). Proteins that are known to possess such autoinhibitory domains have diverse functions, however, there is a common mechanistic theme. These proteins have an intramolecular interaction that interferes with the function of a second domain either directly or allosterically. This results in the target domain being either indirectly constrained in a non-functional conformation, or occluded from ligand interactions directly (Pufall and Graves, 2002). Autoinhibition represses an activity and only when the appropriate activation signal is sufficiently strong is the block removed (Pufall and Graves, 2002; Wang *et al*., 2006). Autoinhibitory domains are on site repressors that achieve a secure 'off' state (Pufall and Graves, 2002; Wang *et al*., 2006). They are often linked to the protein via a flexible region, which is required for the transition from the inhibited to the activated state (figure 1.14) (Pufall and Graves, 2002; Wang *et al*., 2006). In each case, except where the block is removed by proteolysis, the activation is reversible which provides a resetting switch for the next activation event. This is important for the control of the 'off' and 'on' states of proteins involved in regulatory pathways (Pufall and Graves, 2002).

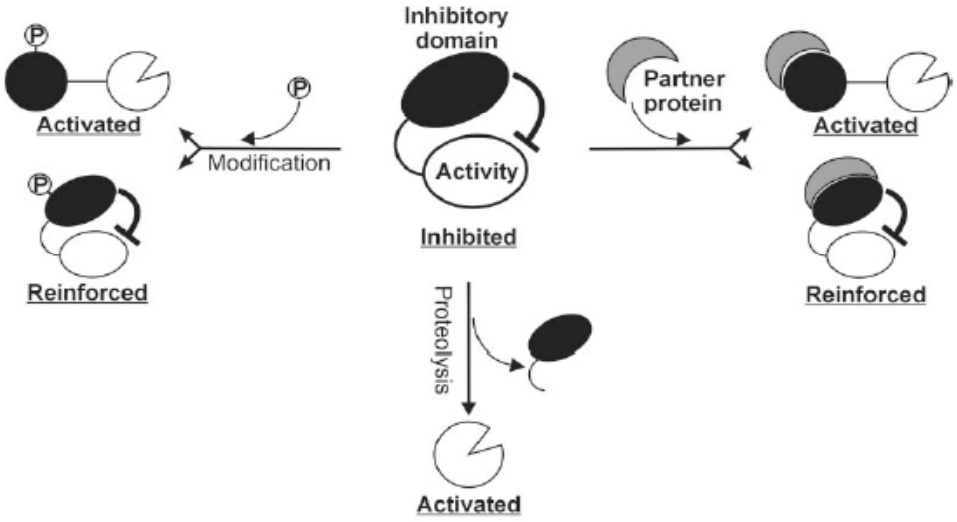


Figure 1.14. A schematic representation of the mechanism of autoinhibitory relief
Autoinhibition is the modulation of a protein domain by a second, separable domain (centre). This block can be released or reinforced by protein modification, for example phosphorylation (left); by interaction with another molecule, such as a protein partner (right); and the block can also relieved by a proteolysis reaction to remove the inhibitory domain (below) (Pufall and Graves, 2002).

Autoinhibition is relieved by a number of mechanisms, including proteolysis of the inhibitory domain, post-translational modifications such as phosphorylation, interactions with a second protein, or binding to small molecules that alter the inhibitory domain (Wang *et al.*, 2006). Interestingly, both protein modification and protein partner interactions could actually reinforce the inhibitory effect (figure 1.14). Differences in protein levels as the cell progresses through the cell cycle could regulate helicases that rely on self-association or hetero-interactions with other proteins to relieve the autoinhibitory block (Brendza *et al.*, 2005). Autoinhibitory domains have been identified in a number of proteins, including UvrB, *E. coli* Rep, Mfd protein and possibly RecB (Brendza *et al.*, 2005; Wang *et al.*, 2006; Smith *et al.*, 2007).

This negative regulatory mechanism is discussed in chapter 5 with regard to the helicase Hel308.

CHAPTER 2: MATERIALS AND METHODS

2.1 Cloning

All plasmids were sequenced at The Sequencing Service, School of Life Sciences, Dundee. The *Sulfolobus solfataricus* P2 genomic DNA was isolated from cells according to the animal tissue protocol of the Qiagen DNeasy® Tissue Kit. Restriction enzymes were supplied by Fermentas and oligonucleotides (including primers) were supplied by Operon, unless otherwise stated.

2.1.1 Cloning of *S. solfataricus* XPB1

The two *S. solfataricus xpb1* constructs, pET28c *xpb1* and pET28c *xpb1 k192a*, were constructed by Dr Jennifer Roberts. The genes were amplified from *S. solfataricus* (Sso) P2 genomic DNA using Pfu polymerase (Promega). The sequences of the primers used are shown in appendix 1.

The products of PCR amplification were cloned into the 5'*Nco*I/3'*Bam*HI restriction sites of the pET28c vector (Novagen) for native expression of the protein.

2.1.2 Cloning of *S. solfataricus* XPB2

The *xpb2* gene was amplified from *S. solfataricus* P2 genomic DNA by KOD polymerase (Novagen) using the primers shown in appendix 1. The PCR product was cloned into the *Nco*I/*Bam*HI sites of the pEHISTEV vector (Supplied by Dr Huanting Liu, Biomolecular Sciences, St Andrews University) for polyhistidine-tagged protein expression. This construct was made and donated to the project by Dr Liza Cubeddu. The *xpb2* gene was also amplified, using the same primers and KOD polymerase (94 °C - 2' x 1 cycle; 94 °C – 30", 52 °C – 30", 68 °C – 2' x 30 cycles; 68 °C – 10' x 1 cycle; 4 °C hold), for cloning into pETDuet with *sso0475* (*bax1*). The reactions were prepared according to the polymerase instruction manual.

2.1.3 Cloning of *S. solfataricus sso0475* (*bax1*)

The *sso0475* gene was amplified by PCR form *S. solfataricus* P2 genomic DNA template for cloning into pET28c and pETDuet (for coexpression with XPB2). The primer sequences for both are listed in appendix 1. KOD polymerase was used for the PCR reaction (94 °C - 2' x 1 cycle; 94 °C – 30", 56 °C – 30", 68 °C – 2' x 30 cycles; 68 °C – 10' x 1 cycle; 4 °C hold) according to the instruction manual.

2.1.4 Gateway® Cloning

In order to amplify *sso1468*, *Sulfolobus acidocaldarius (sac)1656*, and PBL2025 *hel308* genes the Gateway® Cloning system (Invitrogen) was used. Primers were designed to incorporate a TEV cleavage site and a recombination site (*att*B) at the 5' and 3' ends of the gene respectively. A third primer, the mega primer was designed to introduce, from the 5' to the 3' end, an *att*B site, a ribosomal binding site, a polyhistidine tag and a TEV cleavage site. The TEV cleaveage site is common to the two 5' primers in order for the mega primer to anneal to the first PCR products and incorporate a histidine tag. The primer sequences are shown in appendix 1.

Pfu Polymerase was used in the PCR reaction (1 x Buffer, 200 μM dNTPs, 0.5 μM 5' primer, 1 μM 3' primer, 0.05 μM megaprimer, 4 % DMSO (Sigma), 20 ng template genomic DNA, 0.5 U/μl DNA polymerase) using the following protocol: 95 °C - 7' x 1 cycle; 95 °C – 1', 55 °C – 1', 72 °C – 1'30" x 5 cycles; 95 °C – 1', 60 °C – 1', 72 °C – 1'30" x 25 cycles 72 °C – 7' x 1 cycle; 4 °C hold.

The PCR products were gel extracted using the Millipore clean up kit, according to the manufacturer's instructions and cloned into a donor vector pDONR211 by incubation with a Clonase™ enzyme mix (Invitrogen) at 25 °C for 1 hour (5 μl PCR product, 1 μl donor vector, 3.5 μl 5 x Clonase™ Reaction Buffer, 7 μl TE (pH 8.0), 2 μl BP Clonase™ enzyme mix). The reaction was then incubated for a further 10 minutes at 37 °C with proteinase K. These entry clones were then transformed into competent *Escherichia coli* DH5α cells (Novagen). The plasmids were purified from colonies grown overnight, according to the Qiagen Miniprep instruction manual, and subjected to a second recombination reaction (LR reaction) between this entry clone and a destination vector, pDEST14, to yield an expression clone. The entry clone was incubated with LR Clonase™ enzyme mix for 1 hour at 25 °C (2 μl entry clone, 1 μl destination vector, 2 μl 5 x LR Clonase™ Reaction Buffer, 3 μl TE (pH 8.0), 2 μl LR Clonase™ enzyme mix). The reaction was incubated with1 μl of proteinase K at 37 °C for 10 minutes and the expression clones were transformed into *E. coli* DH5α. The colonies were grown up and pure plasmid was purified using the Miniprep kit (Qiagen). The plasmid was then transformed into an expression strain, *E. coli* C43 (Imaxio), for over-expression and purification for further biochemical analysis of Sso1289, Sso1468, Sac1656 and Hel308.

Cloning reactions were carried out by Dr Huanting Liu, Biomolecular Sciences, St Andrews University.

2.1.5 Site-directed mutagenesis

In order to obtain specific mutants of the *S. solfataricus xpb1, xpb2* and Sso*1468* genes and the *S. solfataricus* PBL2025 *hel308* gene, site-directed mutagenesis was carried out using the WT plasmid construct as the template. PCR reactions were carried out according to the Stratagene® XL QuikChange™ Site-Directed Mutagenesis Kit (unless otherwise stated): 1 x Reaction Buffer, 0.2 μM dNTPs, 0.5 μM each primer, 3 % DMSO, 10 ng template DNA, 0.05 U/μl DNA Pfu Turbo polymerase; 95 °C – 30" x 1 cycle; 95 °C – 30", 55 °C – 1', 68 °C – 10' x 18 cycles; 68 °C – 7' x 1 cycle. Details of changes made to this protocol for specific PCR reactions can be found in the relevant results sections. The 5' primer sequences are shown in appendix 1; the 3' primer sequences were the complement of the 5' sequence.

The PCR reactions were incubated for 1 hour at 37 °C with restriction enzyme *Dpn*I (0.2 U/μl), the PCR product was purified and concentrated using a PCR cleanup kit (Millipore) and transformed into the cloning strain, *E. coli* DH5α. DNA sequencing identified the mutations and mass spectrometry confirmed the amino acid substitution by comparing the molecular weight of the peptide sequence containing the mutant to the expected value.

The Companion™ pET Directional TOPO® Expression Kit was used to clone domain 5 of Hel308, which was achieved according to the instructions. The PCR product was gel purified and incubated for 5 minutes at room temperature with the pET151/D TOPO® vector.

2.2 Overexpression of recombinant proteins

2.2.1 Overexpression of recombinant *S. solfataricus* XPB1

Recombinant *S. solfataricus* XPB1 (section 2.1.1) was expressed from a pET28c construct in *E. coli* Rosetta (DE3) pLysS cells (Novagen). The cultures were grown at 37 °C to an OD_{600} of approximately 0.7 in Luria Bertani (LB) broth containing 35 μg/ml kanamycin (Melford) and protein expression was induced by addition of 0.2 mM IPTG (Melford). The cells were incubated overnight at 20 °C before being harvested (15

minutes, 6,000 rpm, Beckman rotor JLA 8.1000) and frozen until required. The XPB1 K192A Walker A box mutant was expressed according to this method.

2.2.2 Overexpression of recombinant *S. solfataricus* XPB2

Recombinant *S. solfataricus* XPB2 (section 2.1.2) was expressed with a polyhistidine-tag from a pEHISTEV construct (supplied by Dr Huanting Liu, Biomolecular Sciences, St Andrews University) in *E. coli* C43 cells (Imaxio). The cultures were grown at 37 °C to an OD_{600} of approximately 0.8 in LB broth containing 35 μg/ml kanamycin and protein expression was induced by addition of 0.1 mM IPTG. The cells were incubated for a further 3 hours at 30 °C before being harvested (15 minutes, 6,000 rpm, Beckman rotor JLA 8.1000) and frozen until required.

2.2.3 Overexpression of recombinant hypothetical protein Sac1656

The pDEST14 *sac1656* gene construct was provided by the Scottish Structural Proteomics Facility (SSPF), St Andrews. *S. acidocaldarius* 1656 protein was expressed in *E. coli* Rosetta (DE3) pLysS cells by auto induction (Studier, 2005) (see appendix 3). In this case the overnight cultures were grown in 100 ml of LB broth and supplemented with ampicillin (Melford) (100 μg/ml). After harvesting the cells (15 minutes, 4,000 rpm, Heraeus rotor #7591), the pellets were resuspended in 2 ml PBS, 1 ml of which was added to 500 ml of ZY5052 minimal media. For more effective aeration, and therefore growth and expression, these cultures were incubated at 37 °C in 2000 ml flasks and shaken at 300 rpm for 6 hours. The temperature was then dropped to 20 °C and the cultures were incubated for a further 18 hours, before being harvested (15 minutes, 6,000 rpm, Beckman rotor JLA 8.1000) and frozen.

2.2.4 Overexpression of recombinant *S. solfataricus* PBL2025 Hel308

The pDEST14 *sso hel308* construct was cloned and supplied by Dr Huanting Liu. Recombinant *S. solfataricus* PBL2025 Hel308 was expressed with a polyhistidine-tag in *E. coli* C43 cells (Imaxio) cultured in LB containing ampicillin at a final concentration of 100 μg/ml, otherwise conditions were the same as for XPB2 expression (section 2.2.2).

2.2.5 Overexpression of recombinant hypothetical proteins Sso1468 and Sso1289

The pDEST14 *sso1468* construct was transformed into *E. coli* C43 cells, and cultures were grown in LB broth containing ampicillin (100 μg/ml) at 37 °C. Overexpression was achieved by induction at an OD_{600} of 0.6 – 0.8 with 0.1 mM IPTG.

The cultures were incubated at 30 °C for a further 3-4 hours before being harvested (15 minutes, 6,000 rpm, Beckman rotor JLA 8.1000) and frozen. Sso1468 K53A was expressed according to this protocol. This protocol was followed for Sso1289, also.

2.3 Protein purification

2.3.1 Purification of recombinant *S. solfataricus* XPB1

The bacterial pellet was thawed and resuspended in 50 ml of lysis buffer (20 mM MES pH 6.0 (Sigma), 1 mM EDTA (Sigma), 0.5 mM DTT (Melford), 0.5 M NaCl (Fisher Scientific), 1 mM benzamidine (Sigma), 0.4 mM AEBSF (Melford), one Complete EDTA-free protease inhibitor tablet (Roche)) and sonicated (Soniprep 150, MSE (UK) Ltd) for 2 minutes, three times on ice. The cell lysate was centrifuged for 20 minutes at 4 °C at 48,000 rpm (Beckman, Ti70 rotor) or for 30 minutes at 4 °C at 20,000 rpm (Beckman, JA 25.50 rotor) and the supernatant was then subjected to a 20 minute 70 °C heat step. The lysate was centrifuged for 15 minutes at 4,000 rpm (Heraeus rotor #7591) and the soluble proteins in the supernatant were filtered (Acrodisc® 0.45 μm syringe filter) and diluted two-fold with buffer A (20 mM MES pH 6.0, 1 mM EDTA, 0.5 mM DTT, 1 mM benzamidine, 0.4 mM AEBSF). The clear lysate was loaded onto an equilibrated 10 ml HiTrap™ Heparin HP column (GE Healthcare) and the protein was eluted off the column over a 195 ml 0 – 1 M NaCl linear gradient.

The fractions containing recombinant XPB1 were identified by SDS-PAGE (Invitrogen) analysis, pooled and concentrated to a volume of 7 ml. The sample was then loaded onto an equilibrated HiLoad® 26/60 Superdex® 200 Gel filtration column (GE Healthcare) (20 mM MES pH 6.0, 1 mM EDTA, 0.5 mM DTT, 0.5 M NaCl, 10 % glycerol (Riedel-de Haèn), 1 mM benzamidine, 0.4 mM AEBSF). The purified protein was identified by SDS-PAGE and the relevant fractions were concentrated. The protein concentration was measured on the Nanodrop Spectrophotometer (NanoDrop Tecnologies) and was calculated using the Beer Lambert Law, $A = \varepsilon cl$, where A is the absorbance of the recombinant protein sample at 280 nm, ε is the extinction coefficient, c is the protein concentration and l is the path length, which is taken as 1.

The extinction coefficient is calculated based on the amino acid sequence and is dependent upon the presence of aromatic residues within the protein sequence

(http://www.expasy.ch/tools/protparam.html). Protein stocks were frozen in liquid nitrogen and stored at –80 °C.

The Walker A box mutant XPB1 K192A was purified using the same method.

2.3.2 Purification of recombinant polyhistidine-tagged *S. solfataricus* XPB2

The bacterial pellet was thawed and resuspended in 50 ml lysis buffer (20 mM Tris-HCl (Sigma) pH 8.0, 0.5 M NaCl, 0.1% Triton-X-100 (Sigma), 1 mM $MgCl_2$ (Fisher Scientific), one Complete EDTA-free protease inhibitor tablet). Resuspended cells were then sonicated on ice three times for 2 minutes. The cell lysate was centrifuged for 20 minutes at 4 °C at 48,000 rpm (Beckman, Ti70 rotor) or for 30 minutes at 4 °C at 20,000 rpm (Beckman, JA 25.50 rotor). The resulting supernatant was filtered (Acrodisc® 0.45 µm syringe filter) and diluted two-fold with buffer A (20 mM Tris-HCl pH 8.0, 0.5 M NaCl, 30 mM imidazole (Sigma), 30 mM NaH_2PO_4 (Fisher Scientific)). The clear lysate was loaded onto an equilibrated 5 ml HiTrap™ Chelating HP Nickel column (GE Healthcare) and the polyhistidine-tagged protein was eluted off the column over a 115 ml 30-500 mM imidazole linear gradient.

The fractions containing recombinant XPB2 were identified by SDS-PAGE analysis and pooled. XPB2 was loaded onto a 10 ml HiTrap™ Heparin HP column (GE Healthcare) equilibrated with buffer A (20 mM MES pH 6.0, 1 mM EDTA, 0.5 mM DTT, 0.5 M NaCl) and eluted manually off the column in 15 ml of buffer B (20 mM MES pH 6.0, 1 mM EDTA, 0.5 mM DTT, 1 M NaCl, 0.1% Triton-X-100, 30% glycerol), in order to prevent XPB2 precipitating. The protein was diluted two-fold in 0 M NaCl buffer B to reduce the NaCl concentration to 0.5 M. The XPB2 K96A Walker A box mutant was expressed and purified in the same way.

10 μg of pure XPB2 was run through a HiPrep™16/60 Sephacryl™S-300 HR column (GE Healthcare) calibrated with protein standards of known molecular weight markers to identify the oligomeric state of the protein.

2.3.3 Purification of the hypothetical protein Sac1656

The frozen bacterial pellet was thawed in a water bath at 37 °C and resuspended in the lysis buffer used for the XPB proteins (section 2.3.1). The cells were sonicated on ice three times for 2 minutes and centrifuged at 48,000 rpm for 20 minutes at 4 °C (Beckman, Ti70 rotor) or for 30 minutes at 4 °C at 20,000 rpm (Beckman, JA 25.50

rotor). The supernatant was filtered (Acrodisc® 0.45 μm syringe filter) before loading it onto the equilibrated 5 ml HiTrap™ Chelating HP Nickel column (GE Healthcare). Polyhistidine tagged Sac1656 was eluted off the column over a 54 ml 30-500 mM imidazole gradient, the buffers used were the same as those used to purify XPB2.

The relevant fractions, identified by SDS-PAGE, were pooled, concentrated down to 12 ml and further purified on an equilibrated HiLoad® 26/60 Superdex® 200 Gel filtration column (GE Healthcare) (20 mM MES pH 6.0, 1 mM EDTA, 0.5 mM DTT, 500 mM NaCl). Sac1656 was concentrated in a Vivaspin column (Vivascience) to 1 ml, the concentration was measured as described in seciton 2.3.1. and the protein was frozen in liquid nitrogen and stored at –80 °C.

2.3.4 Purification of *S. solfataricus* Hel308

Clear lysate containing polyhistidine tagged *S. solfataricus* PBL2025 Hel308 was heat treated for 10 minutes at 70 °C before applying it to the 5 ml HiTrap™ Chelating HP Nickel column (GE Healthcare). *S. solfataricus,* strain PBL2025, Hel308 was further purified on the HiLoad® 26/60 Superdex® 200 Gel filtration column (GE Healthcare); details of these purification methods can be found in sections 2.3.1 and 2.3.2 respectively. Each mutant was purified according to the WT protein protocol.

2.3.5 Purification of the hypothetical protein Sso1468

The soluble cell lysate was subjected to a 10 minute 70 °C heat treatment in order to precipitate a large number of the *E. coli* proteins, which were removed by centrifugation for 15 minutes at 4,000 rpm (Heraeus rotor #7591). This heat step was not carried out with the walker A box mutant protein. Otherwise Sso1468 was purified in the same way as Sac1656 (section 2.3.3); pure protein was then concentrated in dialysis tubing to 0.5-1 ml using PEG 8000 (Promega), frozen in liquid nitrogen and stored at –80 °C. Sso1468 K53A was purified as WT.

2.4 DNA binding and catalytic assays

2.4.1 Purification of oligonucleotides

The oligonucleotides were denatured in 40 % formamide (Promega) at 80 °C for 10 minutes and immediately transferred onto ice. A denaturing polyacrylamide: TBE gel (12 % acrylamide (Flowgen Bioscience), 1xTBE, 7 M urea (Melford), 0.05 % APS,

0.15% TEMED (Fluka BioChemika)) was pre-run for 1 hour at 130 V and the wells were thoroughly rinsed. The samples were loaded and the gel was run at 130 V for a further 90 minutes. The relevant bands were excised and the DNA eluted from the gel pieces overnight at 37 °C in TE buffer (10 mM Tris pH 7.5, 1 mM EDTA). The DNA was purified by ethanol precipitation.

2.4.2 Assembly and purification of DNA Substrates

The oligonucleotide was 5' [γ-^{32}P] ATP (Amersham) end labelled using polynucleotide kinase (PNK) (Fermentas) and the substrates were annealed slowly overnight from 86 °C to room temperature. Gel loading dye, GLD (30% ficoll (Sigma), 0.25% bromophenol blue (Promega), 0.1% xylene cyanole (Sigma)), was added to the samples and the substrates were run and purified from a 12% native polyacrylamide: TBE (12 % acrylamide, 1xTBE, 0.05 % APS, 0.15% TEMED) gels in 1 x TBE buffer and eluted overnight from gel slices at 37 °C in TE buffer (10 mM Tris pH 7.5, 1 mM EDTA). The DNA was ethanol precipitated, resuspended in water and the concentration was measured.

In order to construct a mobile Holliday junction, the oligonucleotides (11M, 10, 22, 21) were purified from a 12 % denaturing polyacrylamide gel (made with 7 M urea). The sequences of the oligonucleotides are shown in appendix 2; a mismatch (highlighted in bold) was incorporated into the sequence to reduce spontaneous branch migration (BUGREEV, 2006).

Phospho-labelled oligonucleotides 11M and 10 were annealed to oligonucleotides 22 and 21 respectively, by slow cooling overnight from 95 °C, to form two splayed duplexes. ^{32}P radiolabelled 11M was also annealed to 22 and to 10 in order to construct size markers. The two splayed duplexes were then annealed together overnight by slow cooling from 50 °C to 4 °C (BUGREEV, 2006). The substrate was purified on a 12 % polyacrylamide: TBE gel at 4 °C at 130 V for 4 hours, the gel was run in 1 x TBE buffer. The DNA duplex was gel excised and incubated for 36 hours at 4 °C in 1xTE buffer with 2 mM $MgCl_2$, in order to prevent spontaneous branch migration, and ethanol precipitated.

2.4.3 Electrophoretic Mobility Shift Assay (EMSA)

The DNA binding affinity was determined by incubating protein at varying concentrations for 30 minutes at room temperature in buffer containing 20 mM HEPES

(Sigma) pH 7.0, 0.1 mg/ml BSA (Sigma), 2 mM DTT, 20 mM NaCl, and 10 nM DNA radiolabelled substrate (section 2.4.2) in a final volume of 15 μl. GLD was added to the samples and they were loaded onto a 12% native polyacrylamide: TBE gel and run for 2 hours at 130 V in 1 x TBE buffer. The gels were exposed overnight and images were visualised by phosphoimaging (Fuji FLA5000).

2.4.4 Anisotropy

DNA binding was also investigated using anisotropy, by either direct titration of protein or by titration of a DNA competitor. All experiments were carried out at 20 °C.
In the case of direct titration, a 15 nucleotide ssDNA strand (the sequence is shown in appendix 2) with a 5' fluorescein label was used at a final concentration of 20 nM; 3 μl of DNA (1μM) was added to 150 μl of anisotropy buffer (20 mM HEPES pH7.6, 100 mM NaCl, 1 mM DTT and 1 drop of Triton-X-100). The first measurement was taken prior to the addition of protein and was subtracted from the data as a blank. The protein concentration was increased cumulatively and further readings were taken after each addition. Data was fitted, using Kaleidagraph, to the standard equation, presented in Reid *et al*. 2001, describing the equilibrium between free DNA and protein and the DNA-protein complex:

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

Where A is the measured anisotropy; E is the total protein concentration; and D represents the total DNA concentration. A_{min} (minimum anisotropy) is the anisotropy of free DNA; A_{max} (maximum anisotropy) is the anisotropy of the DNA-protein complex; and Kd is the dissociation constant (Reid *et al*., 2001).

Measurements of the fluorescence intensities were also recorded. In order to eradicate the effects of anisotropy on fluorescence intensity, "magic angle" conditions were used (Reid *et al*., 2001). Measurements were also taken for a double stranded 15mer (same sequence as ssDNA).

For the competition assays, 1μM protein was incubated with 100 nM fluorescently labelled DNA in anisotropy buffer to a final volume of 150 μl and the first anisotropy reading was taken. Readings were taken after each addition of unlabelled DNA, including fluorescence intensity measurements ("magic angle" conditions). Unlabelled DNA was added until the anisotropy value was constant. The unlabelled DNA

competitor used was the same nucleotide sequence as the fluorescein labelled substrate. In all cases subtle changes in concentrations, due to slight increases in volume, were taken into account. G (grating) factors were automatically calculated by the fluorometer (Cary Eclipse, Varian) for each measurement and used to correct differences in the response to vertical and horizontal polarised light (Reid *et al*., 2001).

The titration curves were plotted in Kaleidagraph and were fitted to the equation described in Reid *et al* (2001), assuming a 1:1 interaction between the protein and DNA molecules (Hey *et al*., 2001; Reid *et al*., 2001):

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

Where A is the measured anisotropy; E is the total protein concentration; and D represents the total DNA concentration. A_{min} (minimum anisotropy) is the anisotropy of free DNA; A_{max} (maximum anisotropy) is the anisotropy of the DNA-protein complex; and Kd is the dissociation constant (Reid *et al*., 2001).

2.4.5 Helicase assay

The protein was assayed for helicase activity over a concentration range; these reactions were set up containing 1 x helicase buffer (20 mM MES pH 6.0, 1 mM DTT, 100 mM NaCl), 0.1 mg/ml BSA, and 10 nM ^{32}P-labelled DNA substrates. The assays were carried out at 45 °C (unless otherwise stated), reactions were incubated for 1 minute and started by the addition of 1 mM ATP (Sigma)-$MgCl_2$ to give a total volume of 15 μl. 10 µl samples were taken after 30 minutes and added to 20 μl chilled STOP Solution (1 x STOP solution buffer (100 mM Tris pH8.0, 50 mM EDTA), 0.5% SDS, 1 mg/ml Proteinase K (New England Biolabs), 5 µM competitor DNA to bind to the displaced strand and prevent reannealing) and GLD was added after 15 minutes. DNA controls were prepared and incubated at 45 °C to ensure DNA substrate stability at the assay temperature and were also incubated at 95 °C to show substrate melting. The samples were loaded onto a 12% native polyacrylamide: TBE gel and run in 1 x TBE buffer for 2 hours at 130 V, the results were visualised as described for the electrophoretic mobility shift assays. The data was quantified using Image Gauge (Fuji); the amount of unwound product was plotted as a percentage of the total DNA against time, using Kaleidagraph software.

2.4.6 ATPase assay

Malachite Green Phosphate Assay Kit standard curves were produced according to the BioAssay Systems protocol. Reaction mixtures consisted of 1 x helicase buffer, 0.2 μM protein, 0.1 mg/ml BSA, and DNA substrates at a concentration of 5.4 μM nts (PhiX174 ss and ds plasmid DNA (New England Biolabs) was used at a concentration of 10 nM) unless stated otherwise to a final volume of 240 μl. The reactions were run at 45 °C and started with addition of a ATP -$MgCl_2$ mixture, to a final concentration of 1 μM. 40 μl samples were taken at specific time points and added to 40 μl of 0.3 M perchloric acid (PCA), in a 96 well plate, to stop the reaction. The conditions described here were used for all experiments, unless described otherwise. Levels of free phosphate were measured after 12 minutes incubation with 20 μl malachite green, on a plate reader at 650 nm (BIO TEK). Each reaction was carried out with the relevant controls, the assays containing only DNA substrate were regarded as background and the numerical data was corrected for this.

2.4.7 Streptavidin displacement assay

10 nM ^{32}P-labelled DNA substrates, with a biotin label on either the 3' or the 5' end, were incubated with 300 nM streptavidin (Sigma), 1xhelicase buffer, 0.1 mg/ml BSA, 1 mM ATP-$MgCl_2$ mix at 37 °C for 3 minutes to allow the streptavidin to bind the biotin (Sigma). 6 μM free biotin was added as a streptavidin trap, the reactions were incubated at 45 °C and 0.5 μM enzyme was added to start the reaction, the final volume of the reaction was 60 μl. At relevant time points 10 μl samples were taken and added to 10 μl STOP solution (1 M NaCl, 200 mM EDTA pH 8.0, 10 μM unbiotinylated DNA to bind to the protein and reduce the effect of band shifting). The samples were separated on a 12% native polyacrylamide: TBE gel and run in 1 x TBE buffer for 2 hours at 130 V, the gels were visualised as described for the EMSA assays (Morris and Raney, 1999). The data was quantified using Image Gauge (Fuji); the streptavidin-free DNA was plotted as a percentage of the total DNA against time, using Kaleidagraph software.

2.4.8 Branch migration assay

Sso1468 was incubated, at a concentration of 0.5 μM, in 1x reaction buffer (25 mM Tris-acetate pH8.0, 10 mM $Mg(CH_3COO)_2$ (Sigma), 50 mM NaCl, 0,5 mM DTT, 10 % glycerol, 1 mg/ml BSA) and 1 mM ATP. The reactions (final volume 70 μl) were

incubated at 30 °C for 1 minute before addition of the Holliday junction, to a final concentration of 10 nM. 10 µl samples were taken at relevant time points and added to 2 µl of STOP solution (1 mg/ml proteinase K, 1.25 % SDS, 10 nM Tris-HCl, 2 mM $MgCl_2$) to terminate the reaction. The samples were incubated on ice in STOP solution, after 5 minutes 5 µl of GLD was added and the samples were run on an 8 % native polyacrylamide: TBE gel at 4 °C for 3 hours in 1 x TBE buffer. The gel pictures were visualised as described for the DNA binding assays (Dennis *et al*., 2004; Bugreev *et al*., 2006).

2.4.9 Auto-phosphorylation detection

2.4.9.1 Radioactive assay

0.2 μM Sso1468 was incubated for 30 minutes at room temperature with 10 μCi [γ-^{32}P] ATP. A control reaction was prepared without protein. The samples were run on a 12 % polyacrylamide: TBE gel for 2 hours at 130 V in 1 x TBE buffer.

2.4.9.2 Intact molecular mass analysis

10 μM Sso1468 was incubated with 10 μM ATP for 30 minutes at room temperature and the intact molecular mass was analysed by mass spectrometry to 0.1 Da. The control sample was prepared without ATP.

2.5 Protein Folding

2.5.1 Circular Dichroism Spectroscopy

The main absorbance is due to the peptide bond although there is a weak, broad absorbance at 210 nm due to the *n to* π^* transition and an intense π *to* π^* transition around 190 nm (Kelly and Price, 1997). The absorbance of the proteins, therefore, was measured in the Far UV Spectra (260 – 180 nm) to determine more quantitative information about their secondary structures. The proteins were added to 300-400 µl of sodium-phosphate buffer (10 mM sodium-phosphate (100 mM disodium hydrogen phosphate (BDH AnalaR®);100 mM sodium dihydrogen phosphate (Fisher Scientific)), 500 mM NaF (Fisher Scientific) and 0.5 mM DTT in the case of XPB2) to give a final concentration of 1 mg/ml, and the absorbance was measured on the Circular Dichroism (CD) Spectrometer (Jasco J-810 Spectropolarimeter). The absorbance was measured a total of six times and automatically averaged.

2.6 Protein-protein interactions

2.6.1 Gel filtration column: identification of protein interactions

A HiPrep™ 16/60 Sephacryl™ S-300 HR column (GE Healthcare) was equilibrated (50 mM Tris pH7.5, 100 mM KCl (Fisher Scientific)) and calibrated using known molecular weight standards (blue dextran, ß-amylase, alcohol dehydrogenase, albumin, carbonic anhydrase and cytochrome C (Sigma)). The proteins of interest were run through and eluted from the calibrated column separately and in combination (flow rate of 1ml/min). Samples from the centre of each peak were analysed by SDS-PAGE (Invitrogen).

2.6.2 Affi-Gel column: identification of protein interactions

2.6.2.1 Preparation of the Affi-Gel columns

A 10 ml disposable column (Pierce) was loaded with 1 ml of well-mixed Affi-gel beads (BioRad) and the isopropanol (Fisher Scientific) in which the beads were stored was discarded. The column was washed with 4 column volumes (CV) of distilled water followed by 4 CV of binding buffer (20 mM HEPES pH 7.0, 1 mM DTT, 1 mM EDTA, 10 % glycerol). The column was incubated for 1-2 hours with at least 3-4 mg of Sac1656 bait protein at room temperature, with mixing. A control column was incubated with the equivalent volume of buffer.

The flow through was discarded and the columns were washed with 10 CV of binding buffer. The remaining binding site esters were blocked by incubating the columns with 1 M ethanolamine (BDH AnalaR®) pH 8.0 for 2 hours at room temperature.

The ethanolamine was removed by washing the columns with 10 CV of binding buffer and a sample of the beads was taken to ensure the protein was bound to the beads. The columns were stored overnight at room temperature, with mixing.

2.6.2.2 Binding and elution of target protein

The columns were equilibrated with 10 CV binding buffer (0.1 M NaCl) and incubated for 1 minute with 3-4 mg *S. solfataricus* XPB2 target protein. The flow through was collected and the columns were then washed again with 40-50 CV of binding buffer (containing 0.1 M NaCl). The first 2 ml of the wash were collected as two 1 ml samples and the rest was discarded. The columns were washed until the protein was

no longer detected in the flow through; this was confirmed using the Bradford Reagent (BioRad).

The proteins were eluted from the column in 40-50 CV of elution buffer (20 mM HEPES pH 7.0, 1 mM DTT, 1 mM EDTA, 10 % glycerol) containing 0.25 M NaCl. Two 1 ml fractions of each elution were collected. This was then repeated with 40-50 CV of 0.5 M NaCl and finally 40-50 CV of 1 M NaCl elution buffer.

The columns were stored at room temperature in 0.1 M NaCl binding buffer until required again.

2.6.2.3 TCA-DOC protein *precipitation*

This method was used to precipitate the proteins collected at very low concentrations for detection. To each 1 ml sample of eluate collected (section 2.6.2.2), 1/100 volume of 2 % deoxycholate (DOC) (Sigma) was added. The samples were vortexed and incubated on ice for 30 minutes. A further 1/10 volume of TCA (Sigma) 100% (100 % TCA: 454 ml H_2O/ Kg TCA, maintain at 4 °C in the dark) was added to each sample, they were vortexed and incubated on ice for a further 30 minutes. The precipitated proteins were centrifuged for 10 minutes at 4 °C at 13,000 rpm (Beckman rotor JA 25.50), the supernatant was discarded and the pellets were dried by inversion on tissue paper. The pellets were washed with 1 volume of cold acetone (stored at –20 °C), vortexed and centrifuged as before. The supernatant was discarded and this wash repeated. The pellets were dried in the speed vacuum (Savant) for 20 minutes at 45 °C. The protein pellets were resuspended in 2 x SDS-PAGE GLD (Invitrogen) and separated on an SDS-PAGE gel. The proteins were visualised by western blot.

2.7 Western Blot

2.7.1 Protein transfer

In order to identify proteins using antibodies, the samples were first run on an SDS-PAGE gel. A PVDF membrane (Hybond-P, Amersham Biosciences) was activated in methanol. The SDS-PAGE gel, the filter paper and the membrane were soaked for 15 minutes in transfer buffer (38.6 mM glycine (Sigma), 47.9 mM Tris pH 8.0, 0.04 % SDS (Sigma), 20 % MeOH (Fisher Scientific)). The proteins from the gel were blotted onto the membrane using the wet blotter (Hoefer SemiPhor semi-dry transfer unit) for 1 hour

at 100 volts/ 300 mAmps. Ponceau stain (Sigma) was used to ensure the proteins had completely transferred and then it was washed off with water before blocking the membrane in blocking buffer (0.05 % TWEEN-20 (Sigma), 5 % milk powder (Marvel) prepared in PBS) for 30 minutes at room temperature (or overnight at 4 °C).

2.7.2 Antibody incubation and detection

The membranes were incubated in fresh blocking buffer with a 1/1,000 dilution of the primary antibody (sheep antiserum containing the polyclonal antibodies against Sso XPB2, Alba Biosciences). After a 1 hour (or overnight at 4 °C) incubation at room temperature, the membrane was washed three times with fresh blocking buffer and then incubated for a further hour with secondary antibody (Immuno Pure® Antibody, Pierce), rabbit anti-goat, used as a 1/10,000 dilution. The washing step was repeated before the membranes were ready to be developed. The membranes were developed using a SuperSignal® West Pico Chemiluminescent Substrate or Femto Maximum Sensitivity Substrate Kit (Pierce) and visualised on the Fuji Luminescent Image Analyser LAS-1000 and Image Reader Software.

2.8 Analysis of protein levels in *Sulfolobus solfataricus* cells

2.8.1 Culturing *S. solfataricus*, Western Blot analysis

S. solfataricus were inoculated to an OD_{600} 0.08 and grown at 80 °C in acidic media (for ingredients see appendix 3) for two days until the OD_{600} value reached 0.84. A sample of the culture was lysed (lysis buffer, see section 2.3.2) and run on an SDS-PAGE gel (Invitrogen) along with a serial dilution of pure recombinant XPB1 or XPB2. The levels of the XPB proteins in the lysate were analysed by western blot (see section 2.7), using antibodies against *S. solfataricus* XPB1 or *S. solfataricus* XPB2.

2.8.2 Real Time (RT) PCR

RNA was extracted (RNeasy kit Qiagen; Proteinase K replaced lysozyme for more efficient lysis of the *S. solfataricus* cells) from *S. solfataricus* cells grown to OD 0.84 and was used at a concentration of 1.5 ng/ µl (1/100 dilution). The real time (RT) PCR reactions were set up according to the BioRad iScript™ One-Step RT-PCR Kit with SYBR® Green kit. The amplifications of the reverse transcribed cDNA were analysed in triplicate using a BioRad iQ5 RT-PCR system (iCycler iQ Optical System Software

Version 3.1). The *xpb1* and *xpb2* gene-specific primers used for amplification are shown in appendix 1. The efficiency of these primers was determined by gene amplification from 132.7 ng, 12.1 ng (1/10 dilution), 1.4 ng (1/100 dilution), and 0.3 ng (1/1000 dilution) of RNA template (measurements to 1 dp).

CHAPTER 3: PURIFICATION AND CHARACTERISATION OF *SULFOLOBUS SOLFATARICUS* XPB1 AND XPB2

3.1 Introduction

The eukaryotic transcription complex TFIIH is required for both promoter opening and escape at transcription initiation from RNA polymerase II promoters and for DNA opening around a lesion during NER (Dvir *et al.*, 1997; Moreland *et al.*, 1999; Laine *et al.*, 2006). The two largest subunits of TFIIH, XPB and XPD, are ATP-dependent helicases with distinct roles in repair and transcription initiation. XPB is a weak helicase and has a major involvement in promoter opening, a role which is indispensable (Coin *et al.*, 1998; Laine *et al.*, 2006; Coin *et al.*, 2007). Contrary to this, XPD does not contribute to transcription initiation. It is a more active helicase, which would account for the net 5'-3' helicase activity of TFIIH, with a role devoted to NER (Coin *et al.*, 1998; Laine *et al.*, 2006; Coin *et al.*, 2007) (Winkler *et al.*, 2000).

TFIIH and TFIIE are both essential for ATP-dependent formation of the open complex around the promoter and they both function to suppress premature arrest of early elongation complexes (Dvir *et al.*, 1997). TFIIE does not posses catalytic activity, but it is capable of recruiting TFIIH to the transcription start site and stabilising its interaction with the polymerase or the preinitiation complex (Dvir *et al.*, 1997; Lin and Gralla, 2005). The α- and β-subunits of TFIIE interact with human ERCC3 (Excision Repair Cross Complementing 3)-XPB inhibiting and stimulating its ATPase activity respectively (Lin and Gralla, 2005). This tight regulation of XPB accomplishes a controlled entry into the elongation phase of transcription of the RNA polymerase bound at the promoter. The activities of XPB and XPD are also stimulated by interactions with partner proteins, p52 and p44 respectively (chapter 4),which are components of the human transcription complex (Coin *et al.*, 1998; Jawhari *et al.*, 2002; Coin *et al.*, 2007).

In contrast to XPD, cross-linking studies have shown that XPB binds the DNA downstream of the transcription bubble and, therefore, is not positioned close enough to the melted DNA to act as a conventional helicase (Kim *et al.*, 2000; Lin and Gralla, 2005). Moreover, addition of ATP to the system promotes alterations in protein-DNA interactions in and downstream of the transcription bubble but not upstream of this region. Based on these discoveries, an alternative model describing the mechanism of promoter opening was proposed (Kim *et al.*, 2000). In the absence of ATP, TBP, TFIIB and possibly TFIIF rotationally fix the promoter in the region upstream of the

transcription bubble by sequence specific protein-DNA interactions. Downstream of the transcription bubble, a number of the RNA polymerase II subunits surround but do not fix this downstream region (Kim *et al.*, 2000). ERCC3-XPB interacts with the remaining accessible DNA site in this region and, upon addition of ATP, rotates the DNA by one turn relative to the upstream fixed DNA. This induces DNA melting between the site bound by ERCC3-XPB and the fixed position, resulting in a ssDNA transcription bubble (Kim *et al.*, 2000). The subunits of RNA polymerase II may act to prevent ERCC3-XPB from making one full rotation at once, completing the process in a number of smaller increments instead. ERCC3-XPB, therefore, acts as a "ratchet wrench" to generate torque in the DNA to form and stabilise the nascent transcription bubble. This rotating mechanism can also facilitate downstream extension of the bubble region suggesting that this mechanism is common to both promoter opening and escape (Kim *et al.*, 2000).

In NER, XPB is thought to bind and wrap around the DNA using the energy produced from ATP hydrolysis to induce local unwinding around the lesion. XPB acts as a 'wedge' to maintain this open state for correct positioning of XPD (Coin *et al.*, 2007). Phosphorylation of XPB during NER inhibits the repair activity of TFIIH, leaving its transcription function intact (Coin *et al.*, 2004). More specifically, the phosphorylation of XPB prevents the ERCC1-XPF catalysed 5' incision reaction without any modification to open complex formation or the XPG mediated 3' incision reaction (Evans *et al.*, 1997; Coin *et al.*, 2004). This phosphorylation reaction is reversible and the kinases and phosphatases may be required to add and remove phosphate groups from XPB to regulate its involvement in NER. It would seem, then, that XPB acts as an ATPase in both transcription and NER (Coin *et al.*, 2007).

Mutations in human XPB and XPD result in a range of defects in transcription and NER and give rise to the three rare inherited disorders *xeroderma pigmentosum* (XP), Cockayne Syndrome (CS) and trichothiodystrophy (TTD) (Hoeijmakers and Bootsma, 1994; Coin *et al.*, 1999; Winkler *et al.*, 2000). XP is characterized by hyperpigmentation of the skin upon exposure to the sun and a predisposition to skin cancers, although a number of clinical features including neurological abnormalities and developmental retardation cannot be attributed to NER defects. These are most likely the result of defective transcription. CS does not show a predisposition to cancer but instead is

associated with neurological and developmental abnormalities. TTD presents repair defects similar to those described for XP. Additionally TTD patients present facial abnormalities, mental retardation and brittle hair but, in contrast to XP, no predisposition to cancer (Coin *et al.*, 1999; Winkler *et al.*, 2000).

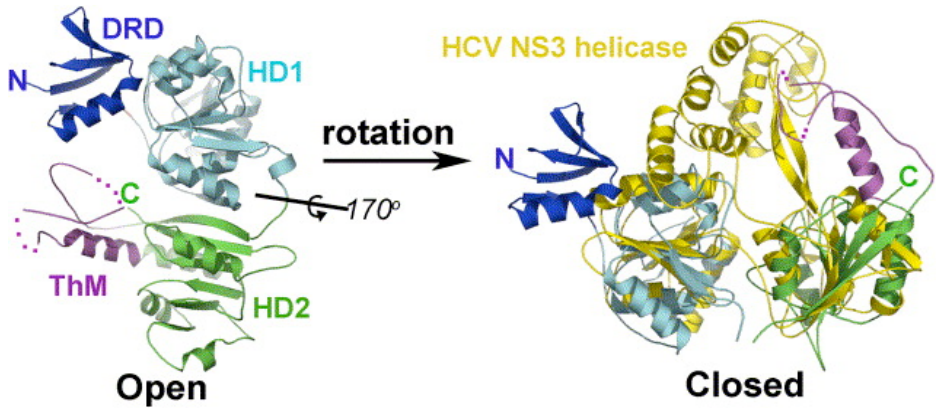
Patients (XP11BE) belonging to the XP-B complementation group exhibit clinical symptoms of both CS and XP (Weeda *et al.*, 1990). An insertion mutation in the ERCC3- XPB gene results in improper splicing of the mRNA and the incorporation of 4 bp into the adjacent exon. The resulting frameshift renders the protein inactive in repair (Weeda *et al.*, 1990). In contrast to the symptoms displayed by XP11BE patients, XP-F patients present mild clinical features with a slight sensitivity of skin to sunlight and a late onset of skin cancer. They do not have neurological abnormalities.

Interestingly, XP-F patients exhibit the same DNA repair defect as XP-B patients, both are unable to catalyse the 5' strand incision (Coin *et al.*, 2004). It has been suggested that sun sensitivity of XP-B patients is the hallmark of XPF to incise the damaged strand, whereas most XP-B features, such as the early onset of skin cancers, are a direct result of a failure to transcribe specific genes (Evans *et al.*, 1997; Coin *et al.*, 2004).

Other mutations that disrupt the interaction between XPB and p52, and similarly for XPD-p44, manifest as transcriptional abnormalities associated with XP, CS and TTD (Coin *et al.*, 1999).

The structure of *A. fulgidus* XPB revealed a damage recognition domain (DRD), two helicase/motor domains (HD1 and HD2), a thumb (ThM) and a RED motif (Fan *et al.*, 2006). In contrast to eukaryotic XPB, biochemical analysis suggested that the DRD binds preferentially to damaged dsDNA. This suggests that XPB recognizes damage-induced distortions in the DNA and that XPB possesses helicase activity that is modulated by DNA lesions (Fan *et al.*, 2006).

A



B

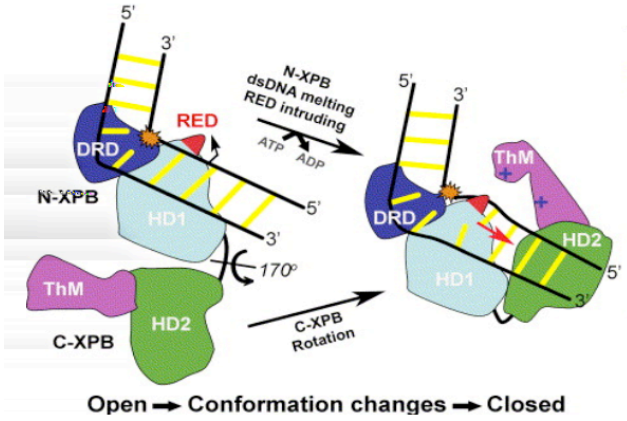


Figure 3.1. The structure of Afu XPB and the proposed mechanism of action
A) The crystal structure of *Archaeoglobus fulgidus* XPB represents the enzyme in an'open' conformation; the helicase motifs adopt a 'stretched out' arrangement. The HD1 domain of *A. fulgidus* XPB was superimposed onto the equivalent domain in the crystal structure of the HCV NS3 (yellow) in complex with DNA. In order for *A. fulgidus* XPB HD2 to adopt the 'closed' conformation, it would have to rotate through an angle of 170 ° (Fan *et al.*, 2006).
B) XPB consists of a damage recognition domain (DRD, dark blue), two helicase domains (HD1, light blue and HD2, green), and a thumb motif (ThM, purple). It was proposed that the DRD initially recognises duplex distortion and binds damaged dsDNA in the 'open' conformation. Upon DNA binding and ATP hydrolysis (HD1), the C-terminal portion (HD2 and ThM) rotates 170 ° to form the 'closed' DNA bound conformation. Local DNA melting allows the RED motif to intrude between the strands and 'unzip' the base pairs (Fan *et al.*, 2006).

The relative orientation of the RecA motor domain was particularly intriguing for Fan *et al.* (2006). In contrast to previously published structures of helicases in complex with DNA, for example PcrA (Velankar *et al.*, 1999) and HCV NS3 (Kim *et al.*, 1998), the seven motifs of *A. fulgidus* XPB were in a "stretched out" arrangement rather than in a cluster around the cleft between the two helicase domains. By superimposing the HD1

domain of *A. fulgidus* XPB on the equivalent domain of HCV NS3 associated with DNA, it was shown that in order for HD2 of XPB to position itself so that the seven motifs formed a corresponding cluster, HD2 would have to rotate 170 ° (figure 3.1A). The *A. fulgidus* XPB crystal structure was, therefore, thought to represent an 'open' conformation and that DNA binding would induce this large conformational change forming the 'closed' structure. This type of domain movement is common in helicases upon substrate binding and ATP hydrolysis (Korolev *et al.*, 1997).

The model proposed by Fan *et al.* (2006) describing the mechanism by which XPB unwinds duplex DNA containing a lesion is shown in figure 3.1B. Structural and biochemical data suggested that XPB would bind to a distorted duplex much like bacterial MutS mismatch repair protein. The crucial base stacking phenylalanine residue, however, is missing from the XPB sequence and, therefore, it is more plausible that XPB recognises the distortion rather than the adduct itself, a mechanism that is more appropriate for NER (Obmolova *et al.*, 2000; Fan *et al.*, 2006). The DRD is targeted to the lesion and binds to the duplex in the 'open' conformation. DNA binding induces a conformational change dependent on the ATPase activity of HD1. The rotation of HD2 (170 °) relative to HD1 forms an active, 'closed' conformation. This movement positions the RED motif in the helicase active site and the side chains of RED are able to protrude between the distorted strands (Fan *et al.*, 2006). One of the strands of the DNA bubble threads through the channel between the RED and ThM motifs, whereas the other strand occupies the groove on the other side of the RED motif. The RED motif, therefore, aids strand separation by unzipping the duplex. This is a similar mechanism to that proposed for the β-hairpin of the bacterial NER protein UvrB. This 'unzipping' mechanism was anticipated to be an addition to the inchworm model that describes helicase movement along one DNA strand (Fan *et al.*, 2006).

The *S. solfataricus* genome contains two copies of the *xpb* gene (chapter 4), *sso0959* and *sso0473*; the protein products are referred to as XPB1 and XPB2, respectively. The purification and characterisation of both SsoXPB1 and SsoXPB2 are described in this chapter, as well as some possible explanations for the limited activity exhibited by these two putative helicases. SsoXPB1 and SsoXPB2 are referred to as XPB1 and XPB2, unless stated otherwise.

3.2 Expression and Purification of *S. solfataricus* XPB1

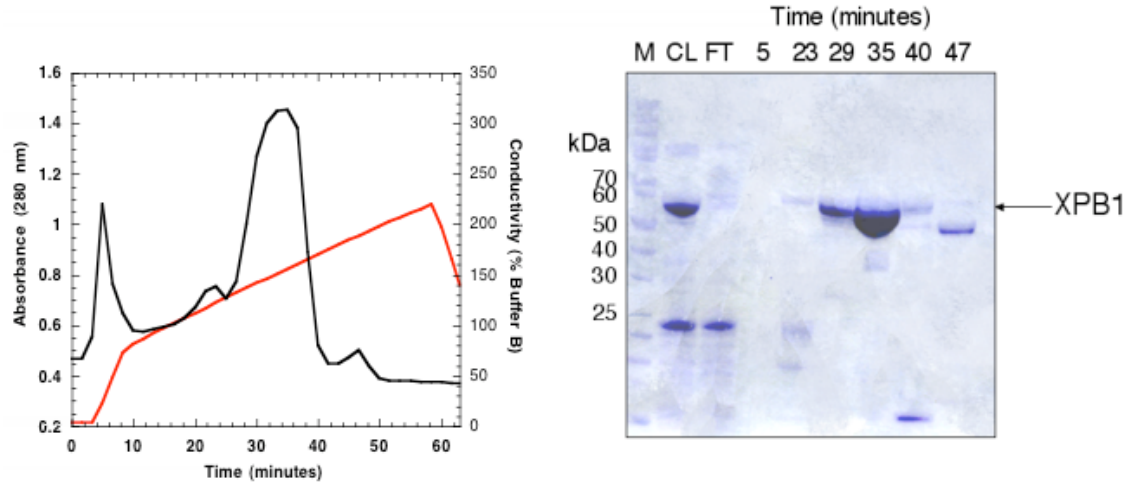
The *S. solfataricus xpb1* gene (*sso0959*) was PCR amplified and cloned into a pET28c vector for expression of the recombinant protein, XPB1 (provided by Dr Jen Roberts). Protein purification was achieved by a combination of heat treatment to precipitate the *E. coli* proteins, ion exchange and size exclusion column chromatography. The protein was eluted from an equilibrated heparin column across a 0 – 1 M NaCl gradient. The peak fractions were analysed by SDS-PAGE, and XPB1 was purified further by gel filtration (figure 3.2).

The Walker A box mutant K192A was constructed by site directed mutagenesis and was provided by Dr Jen Roberts, to be used as a negative control for activity. Expression and purification of K192A was achieved according to the WT XPB1 protocol.

XPB1 eluted from the gel filtration column across 3 peaks (figure 3.2B). XPB1 was not detected in the first peak, according to SDS-PAGE gel analysis. The second and third peaks were pooled separately and the DNA content was measured on the nano-drop spectrophotometer. The DNA (260 nm):protein (280 nm) ratio of the second and third peaks were 1.19 and 0.58, respectively and it was, therefore, assumed that the second peak represented XPB1 in complex with DNA. XPB1 from peak 3 (DNA-free) was consequently used in all subsequent investigations.

The identification of XPB1 WT and K192A was confirmed by mass spectrometry.

A



B

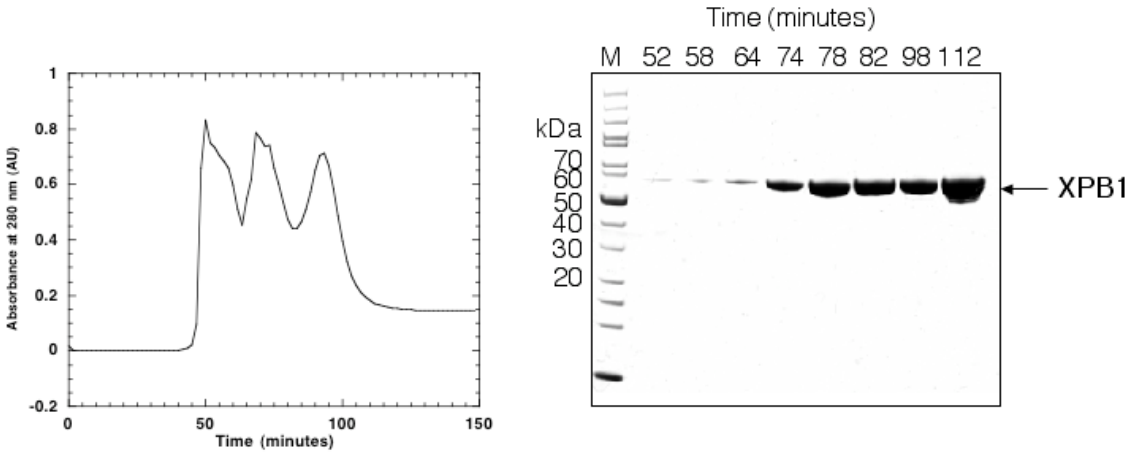


Figure 3.2. Purification of recombinant XPB1
XPB1 was purified by a combination of heat treatment and column chromatography. A) XPB1 was eluted from the heparin column across a 0 – 1 M NaCl gradient, the peak fractions identified on the absorbance trace were analysed by SDS-PAGE. CL represents the clear lysate (soluble) and FT is the column flow through. B) XPB1 was passed through a gel filtration column, the peak fractions indicated on the absorbance trace were analysed by SDS-PAGE. M represents the size marker.

3.3 *S. solfataricus* XPB2, expression and purification

The *S. solfataricus xpb2* gene (*sso0473*) was PCR amplified and cloned into a pEHISTEV vector by Dr Huanting Liu (and donated to the project by Dr Liza Cubeddu) for expression of the protein with an N-terminal polyhistidine-tag to increase soluble

expression and efficient purification. *E. coli* clear lysate containing XPB2 could not be subjected to heat treatment as soluble XPB2 was lost above 50 °C (figure 3.3).

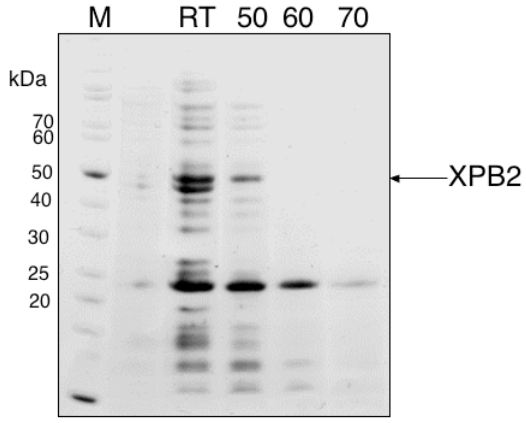


Figure 3.3. Heat treatment of XPB2
SDS-PAGE gel analysis showing the effect raising the temperature has on soluble XPB2. Soluble cell lysate from an *E. coli* expression culture was incubated for 10 minutes at the temperatures indicated, the soluble lysate was analysed by SDS-PAGE. Temperatures are shown in °C, RT represents room temperature and M is the size marker.

XPB2 was immobilised on a chelating nickel column and eluted across a 30 mM – 500 mM imidazol gradient (figure 3.4A). The peak fractions were separated on an SDS-PAGE gel and the fractions containing XPB2 were collected and loaded onto an equilibrated heparin column in 0.5 M NaCl. XPB2 was eluted from the ion exchange column in 1 M NaCl (figure 3.4B), identified by SDS-PAGE analysis and concentrated. XPB2 was not the subject of size exclusion chromatography since preliminary trials resulted in high levels of protein precipitation at 4 °C. XPB2 proved to be quite unstable, so the shorter the purification process the higher the yield of XPB2.

Since the polyhistidine tag removal reaction was only 50 % efficient (figure 3.4C) and resulted in a significant level of protein precipitation, further experiments were carried out with polyhistidine tagged XPB2.

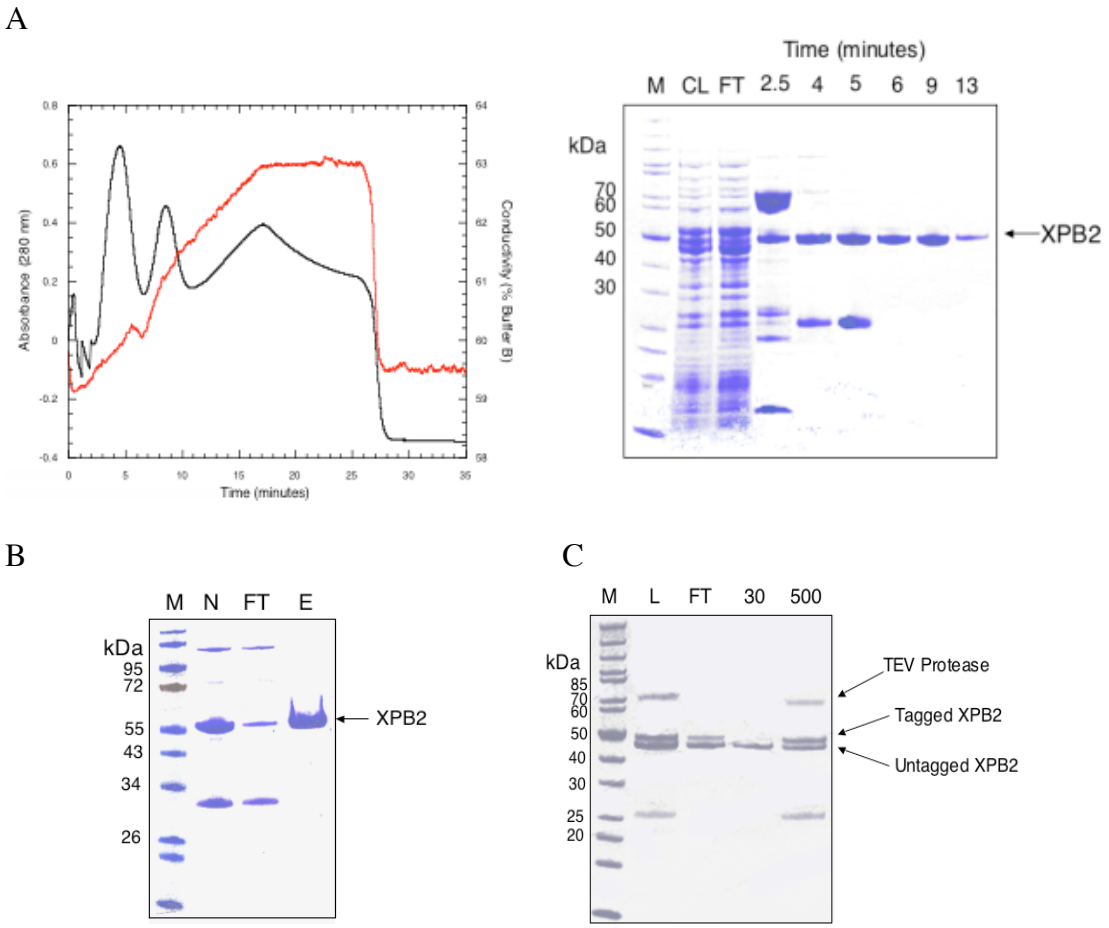


Figure 3.4. Purification of recombinant, polyhistidine tagged XPB2
Purification of XPB2 was achieved by a combination of column chromatography. A) XPB2 was immobilized on the nickel column. The peak fractions, as identified on the absorbance spectrum, were analysed by SDS-PAGE. B) The relevant fractions were loaded onto the heparin column in 0.5 M NaCl and eluted in 1 M NaCl. XPB2 was identified by SDS-PAGE gel analysis. N represents the pooled fractions from the nickel column; FT is the heparin column flow through and E represents XPB2 eluted from the column in 1M NaCl buffer. C) An SDS-PAGE gel of XPB2 eluted from the nickel column following treatment with TEV protease. L is the sample loaded onto the column, FT is the flow through; 30 represents the 30 mM imidazol wash and the 500 mM imidazol elution is shown as 500. On each gel M represents the size marker.

As a negative control for enzyme activity, the Walker A box mutant XPB2 K96A was constructed by site-directed mutagenesis of the pEHISTEV XPB2 template plasmid using Phusion High Fidelity DNA polymerase (New England Biolabs). The PCR reaction is described in the materials and methods, although the annealing and elongation stages required optimisation for this particular gene. The annealing temperature was dropped to 50 °C for 30 seconds and the elongation time was extended to 3 minutes at 72

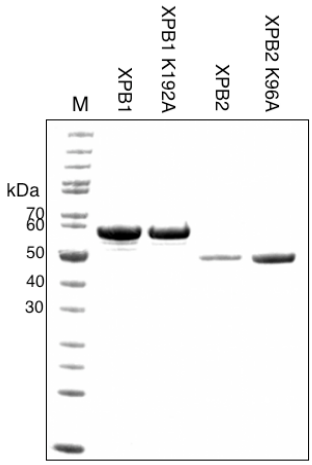
°C. XPB2 K96A was expressed and purified according to the WT XPB2 purification protocol.

The identities of XPB2 and XPB2 K96A were confirmed by mass spectrometry.

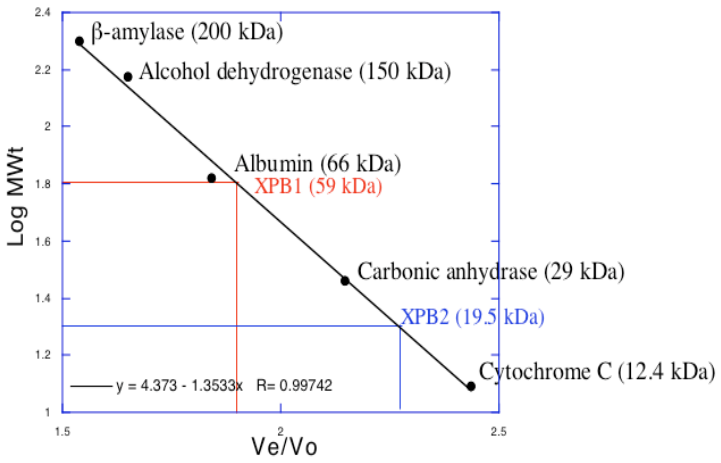
3.4 Size determination of XPB1 and XPB2

XPB1, K192A, XPB2 and K96A were analysed on an SDS-PAGE gel (figure 3.5A) as conformation of the estimated molecular weights of the monomeric proteins. Samples of XPB1 and XPB2 (50 μg/ml) were passed down a calibrated gel filtration column in order to obtain an estimate of their sizes and, therefore, determine the oligomeric state of the proteins. XPB1 eluted with a calculated mass of 59 kDa, the expected mass of XPB1 is 64,175.1 Da so it was clear that this protein eluted as a monomer. The calculated mass of XPB2, 19.5 kDa, was not consistent with the expected mass of 50,817.2 Da. This was repeated a number of times and each time XPB2 eluted from the column with an estimated molecular weight between 11- 22 kDa. Possible explanations for the late elution of XPB2 include the interaction of XPB2 with the column material or tight folding of the XPB2 monomer, enabling it to occupy smaller volumes within the matrix than expected (figure 3.5B) (Morris *et al.*, 2001). The intact molecular weights of both XPB1 and XPB2 were analysed by mass spectrometry. The masses were 64,034.5 Da and 53,539.4 Da respectively (figure 3.5C and D), and were in good agreement with the expected values. The intact mass for XPB2 was slightly higher than expected due to an N-terminal polyhistidine tag and a TEV protease cleavage site.

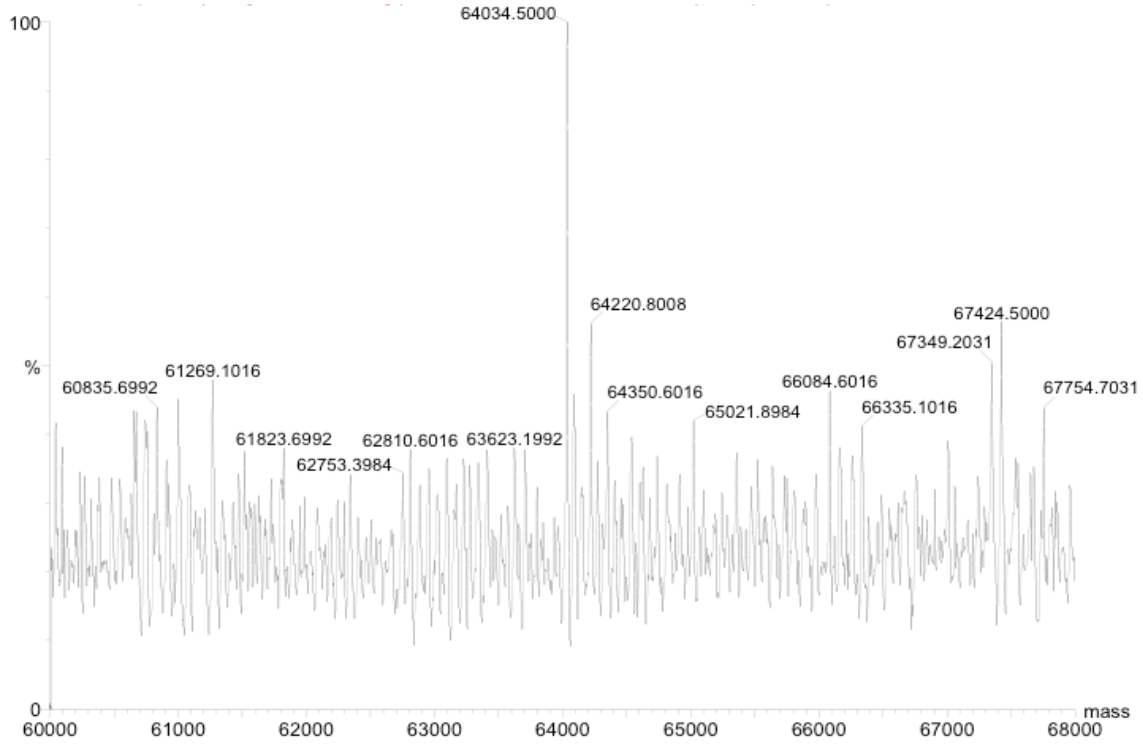
A


M
XPB1
XPB1 K192A
XPB2
XPB2 K96A
kDa
70
60
50
40
30

B


β-amylase (200 kDa)
Alcohol dehydrogenase (150 kDa)
Albumin (66 kDa)
XPB1 (59 kDa)
Carbonic anhydrase (29 kDa)
XPB2 (19.5 kDa)
Cytochrome C (12.4 kDa)
y = 4.373 - 1.3533x R= 0.99742
Log MWt
Ve/Vo
2.4
2.2
2
1.8
1.6
1.4
1.2
1
1.5
2
2.5

C


100
%
0
64034.5000
64220.8008
67424.5000
67349.2031
61269.1016
60835.6992
64350.6016
66084.6016
65021.8984
66335.1016
67754.7031
61823.6992
62810.6016
63623.1992
62753.3984
mass
60000
61000
62000
63000
64000
65000
66000
67000
68000

D

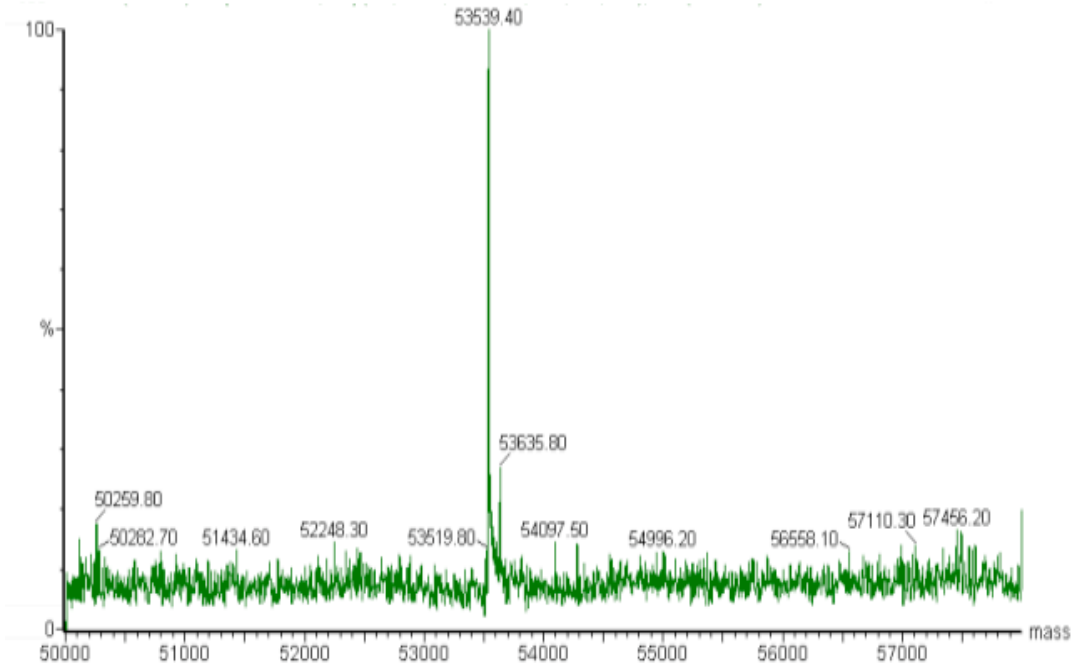


Figure 3.5. Size determination of XPB1 and XPB2
A) SDS-PAGE gel analysis of XPB1, XPB2 and their respective Walker A box mutants. M represents the size marker. B) Standards of known molecular weights were used to calibrate the gel filtration column. Their elution volumes divided by the void volume (retention time of blue dextran, the largest standard) was plotted against the log of their molecular weights. The resulting standard curve was used to estimate the sizes of XPB1 and XPB2, based on their elution volumes. The intact molecular weights were confirmed by mass spectrometry; C) XPB1, the major peak represented a protein with a molecular weight of 64,034.5 Da (data was processed to 0.1 Da) and D.) XPB2, the major peak represented a protein with a molecular weight of 53,539.4 Da (data was processed to 0.1 Da).

3.5 XPB1 and XPB2 preferentially bind to a ssDNA region

Electrophoretic mobility shift assays (EMSAs) were used to investigate the DNA binding abilities of both XPB1 and XPB2. The EMSAs provide some insight into DNA binding, although, especially if the binding interaction is weak, the migration of large DNA-protein complexes through a gel matrix can result in complex disassembly. These band shift assays only capture those molecules that are bound to the DNA with high affinity. Those that are in the process of translocating along the DNA undergo temporary dissociation and, hence, do not produce band shifting (McDougal and Guarino, 2001).

An apparent (estimate) Kd, therefore, is obtained form this type of experiment and in most cases the actual Kd value will be much lower.

XPB1 and XPB2 were incubated at varying concentrations with a variety of DNA substrates (see appendix 2 for oligonucleotide sequences); a 50mer single strand (J150B), 50mer dsDNA (J150B + J150B complement) and a 50mer duplex with either a 3'- or 5'-25 nt single stranded overhang region (J150B + complement with 3' or 5' 25 nt T tail). The protein was incubated with 10 nM DNA substrate at room temperature. The results of XPB1 and XPB2 binding to a 3'overhang DNA substrate are shown in figure 3.6, the gel pictures of the remaining substrates are shown later in figure 3.14 for a direct comparison with binding of XPB1 and XPB2 to the damaged DNA substrates.

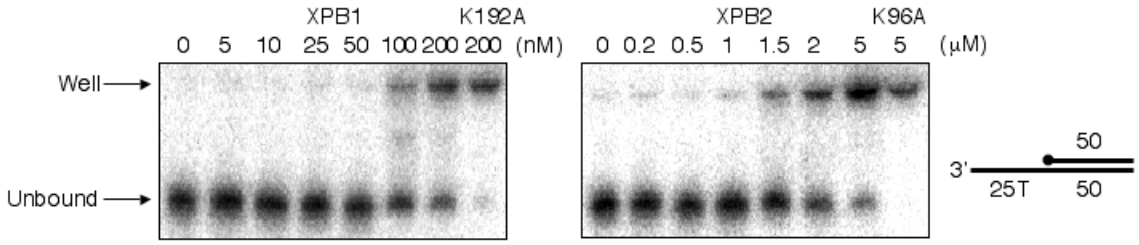


Figure 3.6. Electrophoretic Mobility Shift Assays of XPB1 and XPB2
The enzymes were incubated at the concentrations shown, with 10 nM 3'overhang DNA substrate (^{32}P-5'-radiolabelled, black circle) for 30 minutes at room temperature in a final volume of 15 μl. The samples were run on non denaturing 12 % polyacrylamide: TBE gels to separate unbound substrate from the protein bound DNA.

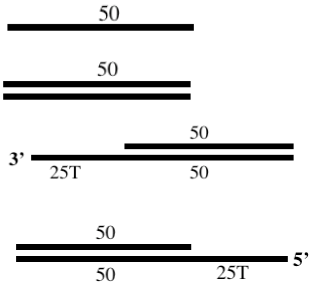
The Kd values were estimated from the band shift gel pictures and are presented in table 1. In each case, a proportion of the enzyme-DNA complex was held up in the wells or had migrated into the running buffer and, therefore, the Kd estimates were calculated based on the disappearance of unbound DNA.

Significantly tighter binding of both enzymes to ssDNA was seen compared to dsDNA, although the binding affinity of XPB1 (Kd~50 nM) was much higher than that of XPB2 (Kd~1 μM). Tighter binding of XPB1 to DNA compared to XPB2 was evident with each DNA substrate. The Kd estimates for XPB1 and XPB2 binding to an overhang substrate were higher than the corresponding ssDNA values, this was expected since the single stranded portion of the overhang substrate was shorter than the 50mer ssDNA.

Blunt ended dsDNA proved to be a very poor substrate for both XPB1 and XPB2, with Kd values of 1 μM and 10 μM respectively. The binding affinity of XPB1 was

similar to that seen for several other SF1 and SF2 helicases, for example PcrA and DinG, respectively (Dillingham *et al.*, 2001; Voloshin and Camerini-Otero, 2007), but XPB2 bound much more weakly.

Table 1. The apparent Kd values for XPB1 and XPB2
The DNA was used at a final concentration of 10 nM in the binding assays

	XPB1	XPB2
Substrate	**Kd (μM)**	**Kd (μM)**
	0.025 – 0.05 0.5 - 1 0.1 - 0.2 ~0.1	~ 1 ~10 ~ 2 1.5 - 2

On some of the gel pictures, slower migrating complexes were obvious (figure 3.6). This was explained by the presence of various DNA-protein complexes, likely the result of non-cooperative binding of multiple protein molecules to the DNA (McDougal and Guarino, 2001; Rudolf, 2006).

3.6 DNA dependent ATPase activity

All helicases have biochemical properties in common in that they posses the ability to bind nucleic acids and NTP molecules. They use the energy generated from NTP hydrolysis to translocate along ssDNA and catalyse nucleic acid duplex unwinding. Helicases are categorised into families based on conserved amino acid motifs (Hall and Matson, 1999; Singleton *et al.*, 2007). The XPB amino acid sequences contain the seven conserved motifs typical of the SF2 helicases (Gorbalenya and Koonin, 1993).

The ATPase activities of XPB1 and XPB2 were measured in the presence and absence of DNA. XPB1 and XPB2 were incubated at a concentration of 0.2 μM with 10 nM PhiX174 virion (ssDNA) or RFI DNA (covalently closed circular dsDNA) at 45 °C. Both enzymes exhibited similar levels of ATP hydrolysis in the presence of DNA, although minimal levels of phosphate were detected in the absence of DNA. The high concentrations (10 nM) of plasmid DNA in these assays probably compensated for the

observed weak binding affinities of XPB2. XPB1 and XPB2 exhibited ATP hydrolysis reaction rates of 62 ± 4.9 and 80 ± 3.3 pmoles phosphate.min^{-1}.pmole enzyme^{-1}, respectively, when stimulated by PhiX174 virion (ss) DNA. The rates of the reaction with PhiX174 RFI (ds) were 17 ± 1.1 and 15 ± 1.5 pmoles phosphate.min^{-1}.pmole enzyme^{-1} for XPB1 and XPB2, respectively, 4-6 fold lower than in the presence of PhiX174 virion (ss) plasmid (figure 3.7A). These observed reaction rates of XPB1 and XPB2 when incubated with ssDNA were approximately 10-fold lower than those evident with XPD. XPD ATPase activity, however, was not stimulated by dsDNA (Rudolf *et al.*, 2006).

A

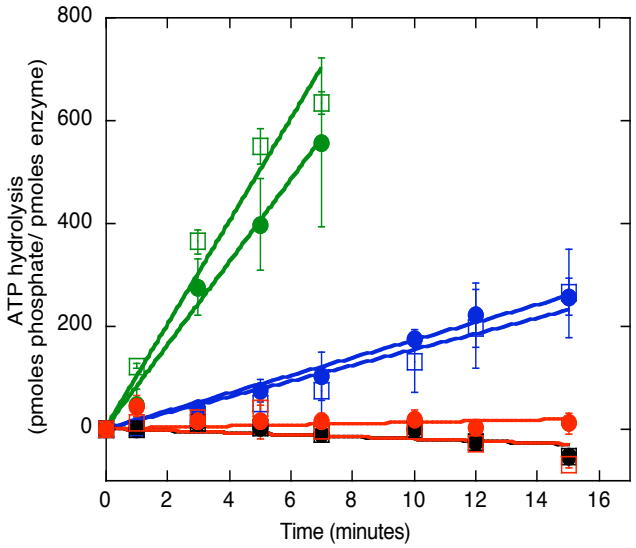


Table 2. The rates of ATP hydrolysis
ATPase assays were carried out with a final nucleotide concentration of 5.4 μM. The rate of ATP hydrolysis is shown as pmoles phosphate. min^{-1}.pmoles protein^{-1}.

Substrate	XPB1 Rate of ATP hydrolysis	XPB2 Rate of ATP Hydrolysis
50	110 ± 3.6	86 ± 6.6
50	94 ± 14	69 ± 8.2
3' 25T 50 50	180 ± 4.2	140 ± 9.2
50 50 25T 5'	60 ± 7.2	47 ± 4.0

Figure 3.7. ATPase activity of XPB1 and XPB2
A) The ATPase assays were carried out in triplicate; 10 nM PhiX174 DNA was incubated with 0.2 µM enzyme, in a final volume of 240 μl, for 1 minute. Samples were taken over a 15 minute time course at 45 °C and added to 0.3 M PCA to stop the reaction. Data was also collected for the ATPase inactive mutants K192A (XPB1) and K96A (XPB2). ssDNA is represented in green; dsDNA in blue; no DNA is shown in black; Walker A box inactive mutants with ssDNA are shown in red; XPB1 is represented by closed circles and XPB2 by open squares. Standard errors are shown.

The invariant lysine of the Walker A box motif binds to the β- and γ-phosphates of ATP; its positively charged side chain resides in the space that is occupied by the magnesium ion once the NTP-Mg^{2+} complex binds the enzyme. During catalysis, the lysine side chain may act to stabilise this transition state by contacting the β-phosphate of the bound NTP (Tuteja and Tuteja, 2004). The Walker A box lysine mutants XPB1 K192A and XPB2 K96A, as expected, were unable to hydrolyse ATP and were useful

negative controls. The ATPase activity, therefore, was not due to the presence of contaminating *E. coli* proteins and could be solely attributed to XPB1 and XPB2.

The ATPase assays were repeated with a 50mer ssDNA (J150B), 50mer dsDNA (J150B + complement) and 50mer duplex with a 25nt overhang (J150B + complement with a 3' or 5' 25 nt T tail) to identify an optimal substrate. The substrates were added to a final nucleotide concentration of 5.4 μM to ensure that the variation in the length of the substrates did not influence ATPase activity. The 3' overhang substrate stimulated the ATPase activity of both XPB1 and XPB2 to a higher level than either the dsDNA or the ssDNA substrate, the reaction rates are presented in table 2. In contrast to this, neither XPB1 nor XPB2 were able to efficiently hydrolyse ATP in the presence of a 5' overhang. Since both XPB1 and XPB2 were able to bind to 3' and 5' overhang with the same affinities, this suggested that both XPB1 and XPB2 were exhibiting 3'-5' polarity although it was by no means conclusive.

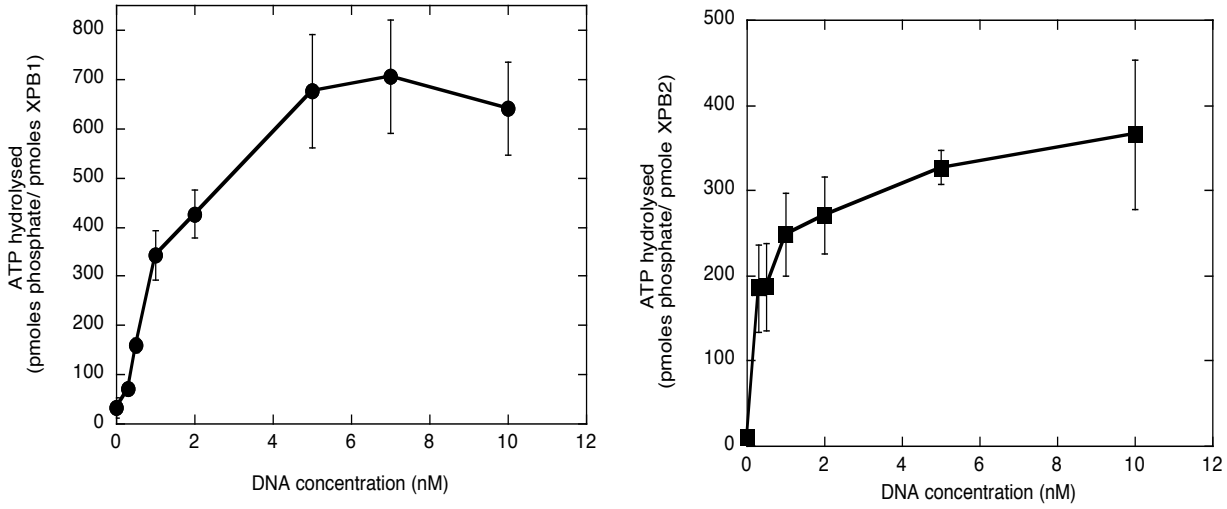


Figure 3.8. DNA dependent ATPase activity of XPB1 and XPB2
In each case 0.2 μM protein was assayed with PhiX174 virion DNA at the concentrations shown (0-10 nM). Samples of XPB1 were taken after 7 minutes, whereas XPB2 was incubated for 3 minutes at 45 °C. The reactions were prepared in a final volume of 50 μl. Each sample was added to 0.3 M PCA to stop the reaction and the levels of free phosphate were detected with malachite green. XPB1 is represented by circles and XPB2 by squares. The averages of triplicate measurements are plotted and the standard errors are shown.

The PhiX174 DNA substrates were used at a concentration of 10 nM (~54 μM nt) whereas these shorter substrates were assayed at a concentration of 5.4 μM nt. The effect of DNA concentration (PhiX174 virion) on the ATPase activities of XPB1 and XPB2 was

also investigated (figure 3.8). Essentially, the enzymes were saturated above 1 nM (~5 μM nt) DNA and, therefore, the data collected with 10 nM PhiX174 substrates was comparable to the reaction rates calculated for the shorter ss and dsDNA substrates.

The rates of ATP hydrolysis of XPB1 and XPB2, when stimulated by a 50mer ssDNA substrate, were in good agreement with the corresponding PhiX174 reaction rates. The short dsDNA substrates, however, stimulated the ATPase activity of both enzymes to a greater extent than expected compared to the PhiX174 RFI (ds). In addition, the difference between the ss and dsDNA-stimulated reactions was not so pronounced with the 50 mer substrates. The reason for this discrepancy was unclear, although it was possible that supercoiled, covalently closed plasmid could influence the enzyme in a different way to short linear substrates.

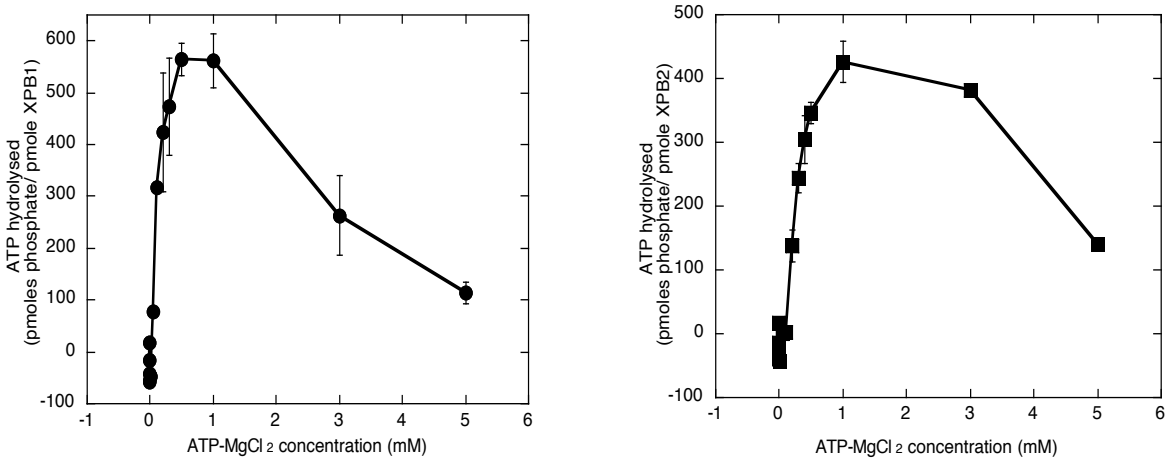


Figure 3.9. ATP dependent ATPase activity of XPB1 and XPB2
Graphs to show the effect of increasing ATP-$MgCl_2$ substrate concentration (0-5 mM) on the ability of XPB1 and XPB2 to hydrolyse ATP. In each case 0.2 μM protein was assayed with 10 nM PhiX174 virion DNA. The samples of XPB1 were taken after 7 minutes, whereas XPB2 was incubated for 3 minutes at 45 °C. The reactions were prepared in a final volume of 50 μl. Each sample was added to 0.3 M PCA to stop the reaction and the levels of free phosphate were detected with malachite green. XPB1 is represented by circles and XPB2 by squares. The averages of triplicate measurements are plotted and the standard errors are shown.

The effect of ATP-$MgCl_2$ concentration on the rate of ATP hydrolysis catalysed by XPB1 and XPB2 was also investigated in the presence of PhiX174 virion DNA (figure 3.9). Both XPB1 and XPB2 exhibited dramatically reduced activity above an ATP-$MgCl_2$ concentration of 1 mM, which was explained by substrate inhibition. One of the most common deviations from Michaelis-Menten enzyme kinetics is the inhibition of enzymes by excess substrate concentrations (Kuhl, 1994).

Since XPB1 and XPB2 are proteins from a hyperthermophilic organism, it was expected that they would be active at high temperatures (figure 3.10), their activities were investigated using PhiX174 virion DNA. This was the case for XPB1 for which the highest level of activity was detected between 50 and 65 °C but a significant level of free phosphate was still released at 80 °C. In contrast, the catalytic activity of XPB2 dropped rapidly above 50 °C and XPB2 precipitated above this temperature, which was consistent with the heat trial shown in figure 3.3.

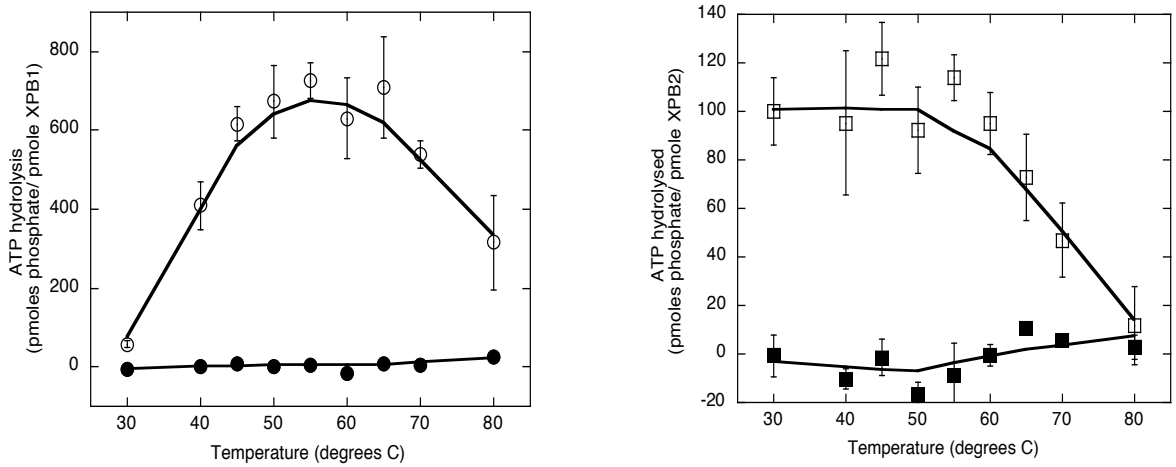


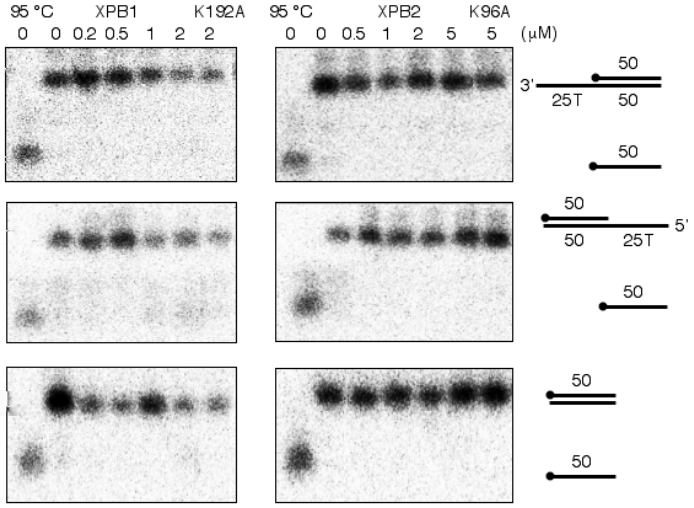
Figure 3.10. Temperature dependent ATPase activity of XPB1 and XPB2
0.2 μM protein was incubated with 10 nM PhiX174 virion DNA (final volume 50 μl) and samples of XPB1 were taken after 7 minutes, whereas XPB2 was incubated for 3 minutes. The reactions were incubated at the temperatures shown, 30-80 °C. Each sample was added to 0.3 M PCA to stop the reaction and the levels of free phosphate were detected with malachite green. XPB1 is represented by circles and XPB2 by squares; both were assayed in the presence (open shapes) and absence (closed shapes) of DNA. The averages of triplicate measurements are plotted and the standard errors are shown.

3.7 XPB1 and XPB2 do not unwind duplex DNA

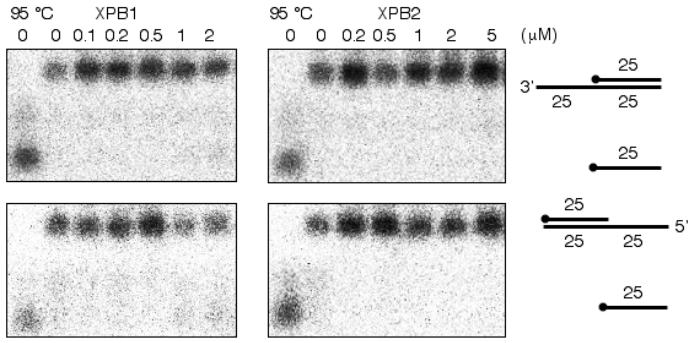
The majority of proteins that contain seven helicase motifs in their amino acid sequence do not actually possess the ability to separate the two strands of a duplex. Therefore identification of these consensus sequences is not sufficient to classify proteins as helicases, they must also be characterised biochemically for unwinding activity (Singleton and Wigley, 2002). Both XPB1 and XPB2 possess the characteristic helicase motifs and, therefore, their ability to catalyse strand separation was of particular interest. The helicase activity was investigated in the presence of blunt duplex (J150B + complement) and 3' or 5' overhang (J150B + complement with a 3' or 5' 25 nt T tail)

DNA substrates for 30 minutes at 45 °C. To stop the reaction the samples were added to STOP solution, which contained unlabelled DNA (same sequence as the labelled strand of the substrate) to prevent substrates reannealing.

A



B



C

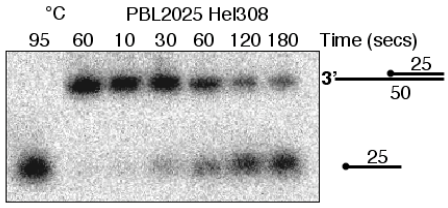
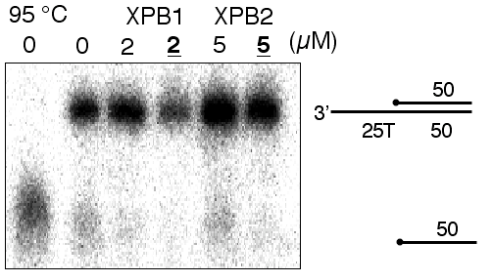


Figure 3.11. Helicase activity of XPB1 and XPB2
Helicase assays to determine the ability of XPB1 and XPB2 to unwind DNA substrates (^{32}P-5'-radiolabelled, black circle). XPB1 and XPB2 were incubated at the concentrations shown with 10 nM blunt ended duplex, 3' or 5' overhang substrates (total volume of 15 μl) at 45 °C for 30 minutes. The overhang substrates had either a A) 50 bp or B) 25 bp duplex portion. C) *S. solfataricus* PBL2025 Hel308 was used as a positive control. 0.5 μM protein was incubated with 10 nM 3'overhang substrate at 60 °C in a final volume of 60 μl. In each case the DNA substrate was incubated at the assay temperature and at 95 °C to show substrate stability at the temperature of the reaction and substrate melting, respectively. The samples were analysed on 12 % polyacrylamide : TBE gels.

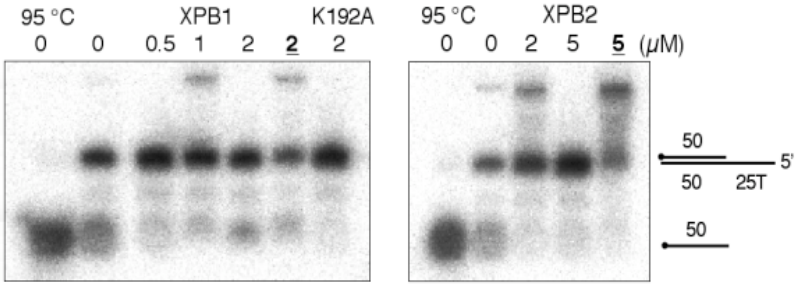
XPB1 was assayed up to a final concentration of 2 μM and XPB2 up to 5μM but no helicase activity was detected (figure 3.11 A). Overhang substrates with a 25 nt duplex region did not stimulate helicase activity either (figure 3.11B). *S. solfataricus* PBL2025 Hel308 was used as a positive control (figure 3.11 C) to ensure that the reaction conditions were suitable to promote helicase activity (Hel308 is discussed in detail in chapter 5). Hel308 was able to efficiently unwind the 3' overhang DNA substrate.

The assays were repeated at a temperature of 65 °C in an attempt to stimulate helicase activity; control reactions were also prepared without ATP-$MgCl_2$. The higher temperatures did not, however, promote duplex unwinding (figure 3.12). As shown before (figure 3.3), XPB2 was not stable at temperatures above 50 °C in contrast to XPB1, which would withstand temperature up to 80 °C. This provides an obvious explanation for the lack of XPB2 helicase activity although XPB1, it would seem, is not active as a helicase.

A



B



C

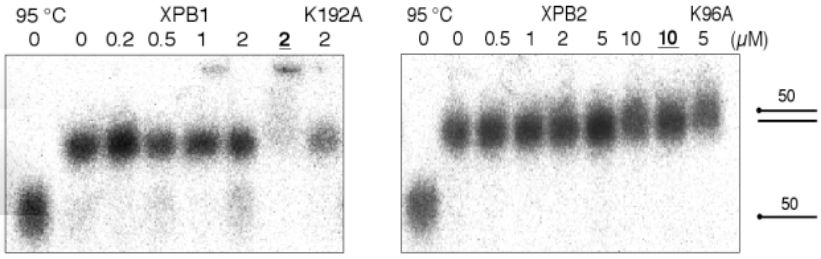


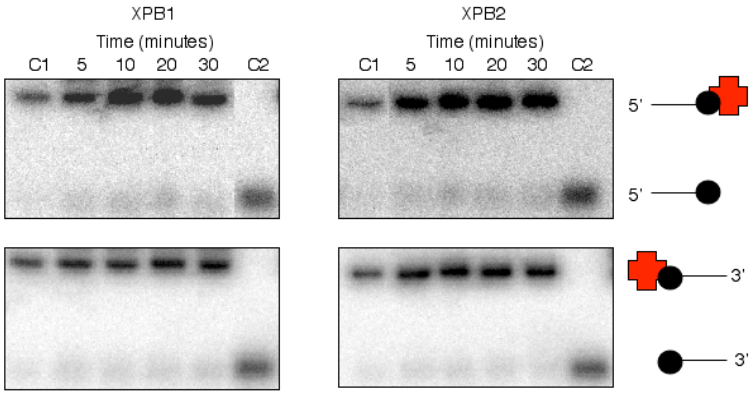
Figure 3.12. Helicase activity of XPB1 and XPB2 at elevated temperatures
The helicase assays were repeated at 65 °C to promote helicase activity of XPB1 and XPB2. The assays were performed with 5.4 μM DNA substrates (^{32}P-5'-radiolabelled, black circles) and protein at the concentrations indicated, to a final volume of 15 μl. After 30 minutes the samples were added to STOP solution to prevent reannealing of the substrate and to stop the reaction. The DNA substrates were incubated at the assay temperature to ensure their stability. Boiled substrate was run on the gel to show the size of the unwound product. A) 3' overhang, B) 5' overhang and C) blunt ended double stranded DNA. The reaction was also carried out in the absence of ATP-$MgCl_2$, this sample is shown in **bold and underlined.** The samples were analysed on 12 % polyacrylamide: TBE gels.

3.8 XPB1 and XPB2 are unable to displace streptavidin

Proteins that are unable to unwind DNA, but contain the helicase motifs, have been implicated in translocation and chromatin remodelling. That is, they possess the ability to move along ssDNA or dsDNA and displace protein from it rather than displacing the DNA strand. The molecular motors of proteins that possess the helicase motifs may not be sufficient to displace a single strand from duplex DNA without the function of additional domains (Singleton and Wigley, 2002).

A streptavidin displacement assay was used to determine whether XPB1 and XPB2 could translocate along ssDNA and function to dislodge streptavidin from DNA bound biotin, a property displayed by the Dda helicase (Morris and Raney, 1999; Morris *et al.*, 2001). Streptavidin was bound to a 3'- or 5'-biotin labelled DNA oligonucleotide (J150B), and incubated with either XPB1 or XPB2. Free biotin in the buffer was used to trap any displaced streptavidin (figure 3.13 A).

A



B

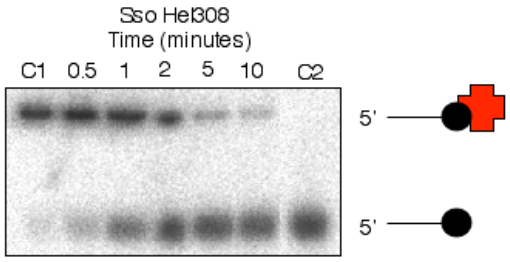


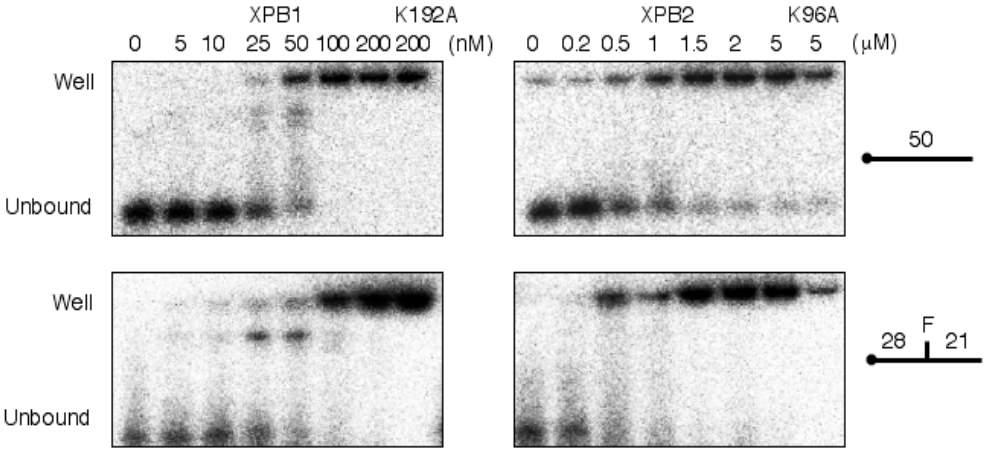
Figure 3.13. Streptavidin displacement assays with XPB1 and XPB2
A) The translocation activity of XPB1 and XPB2 was investigated by measuring the ability of these enzymes to displace streptavidin from a biotinylated probe (^{32}P-5'-radiolabelled). The assays were carried out at 45 °C with 1 mM ATP-$MgCl_2$, 300 nM Streptavidin and 6 µM free biotin in a total of 60 µl. The reaction was started by addition of protein to a final concentration of 0.5 µM. C1 and C2 are controls without XPB proteins; C1 indicates the biotin labelled oligonucleotide bound to streptavidin and C2 represents just the biotin labelled probe. The controls were incubated at 45 °C for 30 minutes to ensure streptavidin was not spontaneously disassociating. The samples were analysed on 12 % polyacrylamide: TBE gels. B) *S. solfataricus* PBL2025 Hel308 was used as a positive control. The reactions were prepared as described in A.

Neither XPB1 nor XPB2 were able to disrupt the streptavidin-biotin interaction. *S. solfataricus* PBL2025 Hel308 was used as a positive control (figure 3.13 B) and was able to efficiently displace the streptavidin molecule (this is discussed in more detail in chapter 5). Proteins, such as Dda, have been shown to work cooperatively to relieve a block. This is not thought to be the result of protein-protein interactions, but rather due to the presence of more than one monomer, which prevents slipping of the first monomer upon encountering a block (Byrd and Raney, 2004, 2006). With regard to the translocation ability of XPB1 and XPB2, the results were inconclusive. The enzymes may have translocated along the DNA strand and then stalled upon encountering a protein block.

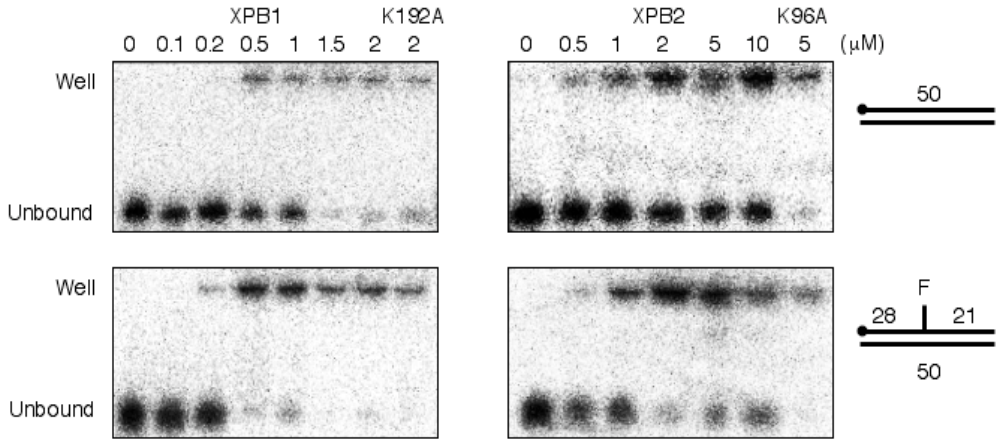
3.9 XPB1 and XPB2 are not implicated in damage recognition

Biochemical analysis of *A. fulgidus* XPB implied that it specifically bound to a damaged duplex DNA substrate. In addition, structural data identified a damage recognition domain (DRD) at the N-terminus. It was proposed that *A. fulgidus* XPB recognised distortions in the DNA induced by a lesion and that the DRD positioned XPB to allow duplex unwinding, therefore, implicating XPB in NER damage recognition (Fan *et al.*, 2006). To investigate this proposal the EMSA, ATPase and helicase assays were repeated with 'damaged' DNA substrates (the same oligonucleotides were used as previously described). The substrates were prepared with an internal bulky extrahelical fluorescein adduct attached to the oligonucleotide (J150B), which served as a good mimic for a NER-type lesion.

A



B



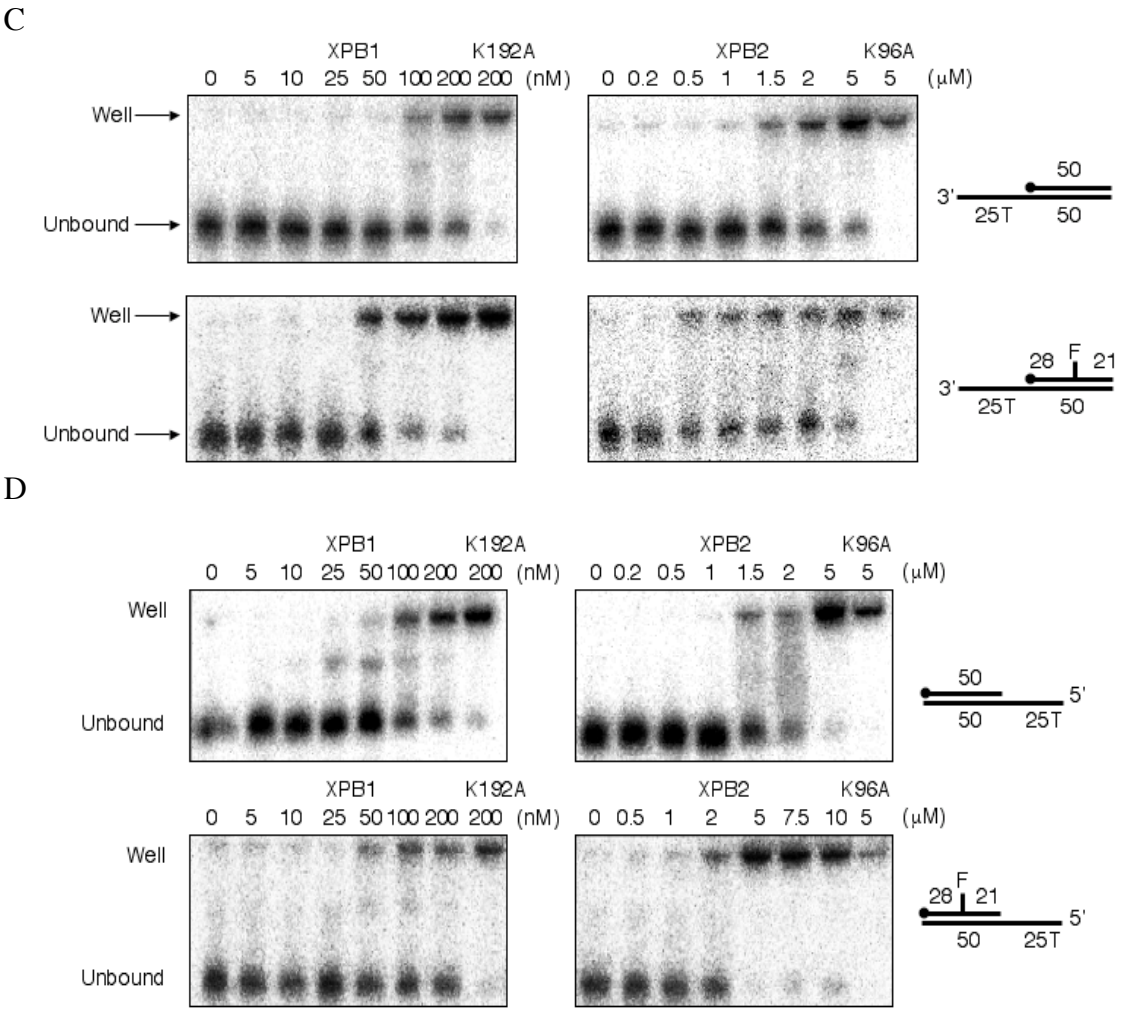
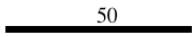
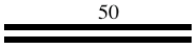
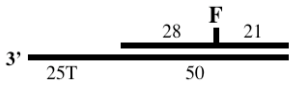
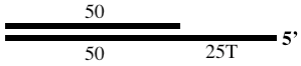
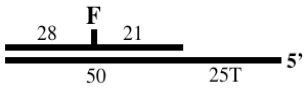


Figure 3.14. The binding affinity of XPB1 and XPB2 to damaged and undamaged DNA
Electrophoretic Mobility Shift Assay to show the binding affinity of XPB1 and XPB2 to damaged and undamaged DNA. The proteins were incubated, at the concentrations indicated, with 10 nM ^{32}P-5'-radiolabelled DNA shown as a black circle (final volume of 15 μl), at room temperature for 30 minutes. The samples were analysed on 12 % polyacrylamide: TBE gels. A) single stranded 50 mer, B) double stranded 50mer, C) 3' overhang and B) 5' overhang. K192A and K96A are the Walker A box mutants of XPB1 and XPB2, respectively. F represents the fluorescein label.

Both XPB1 and XPB2 showed equal, if not a slightly increased binding affinity for the DNA substrates with the fluorescein adduct when compared to the equivalent undamaged substrate (figure 3.14). In contrast, comparison of the ATPase reaction rates (table 3) indicated a possible trend towards lower DNA stimulated ATPase activity in the presence of the fluorescein adduct, although a higher rate of XPB1 activity with the damaged ssDNA substrate was identified.

Table 3. The rates of ATP hydrolysis exhibited by XPB1 and XPB2 with damaged and undamaged DNA substrates
The DNA was assayed at a final nucleotide concentration of 5.4 μM. The rates of ATP hydrolysis are calculated as pmoles phosphate.min^{-1}.pmole of protein^{-1} (2 sf).

	Rate of ATP hydrolysis	
Substrate	**XPB1**	**XPB2**
	110 ± 3.6	86 ± 6.6
	120 ± 3.8	26 ± 5.0
	94 ± 14	69 ± 8.2
	77 ± 14	36 ± 6.8
	180 ± 4.2	140 ± 9.2
	120 ± 8.8	32 ± 4.5
	60 ± 7.2	47 ± 4.0
	67 ± 6.3	45 ± 4.9

These observations suggest that the archaeal proteins may have a higher affinity for damaged DNA but this evidence to suggest a role in the DNA damage recognition event was inconclusive at this stage. The presence of a NER-type lesion in the DNA substrates still did not stimulate XPB1 or XPB2 to catalyse strand separation (figure 3.15).

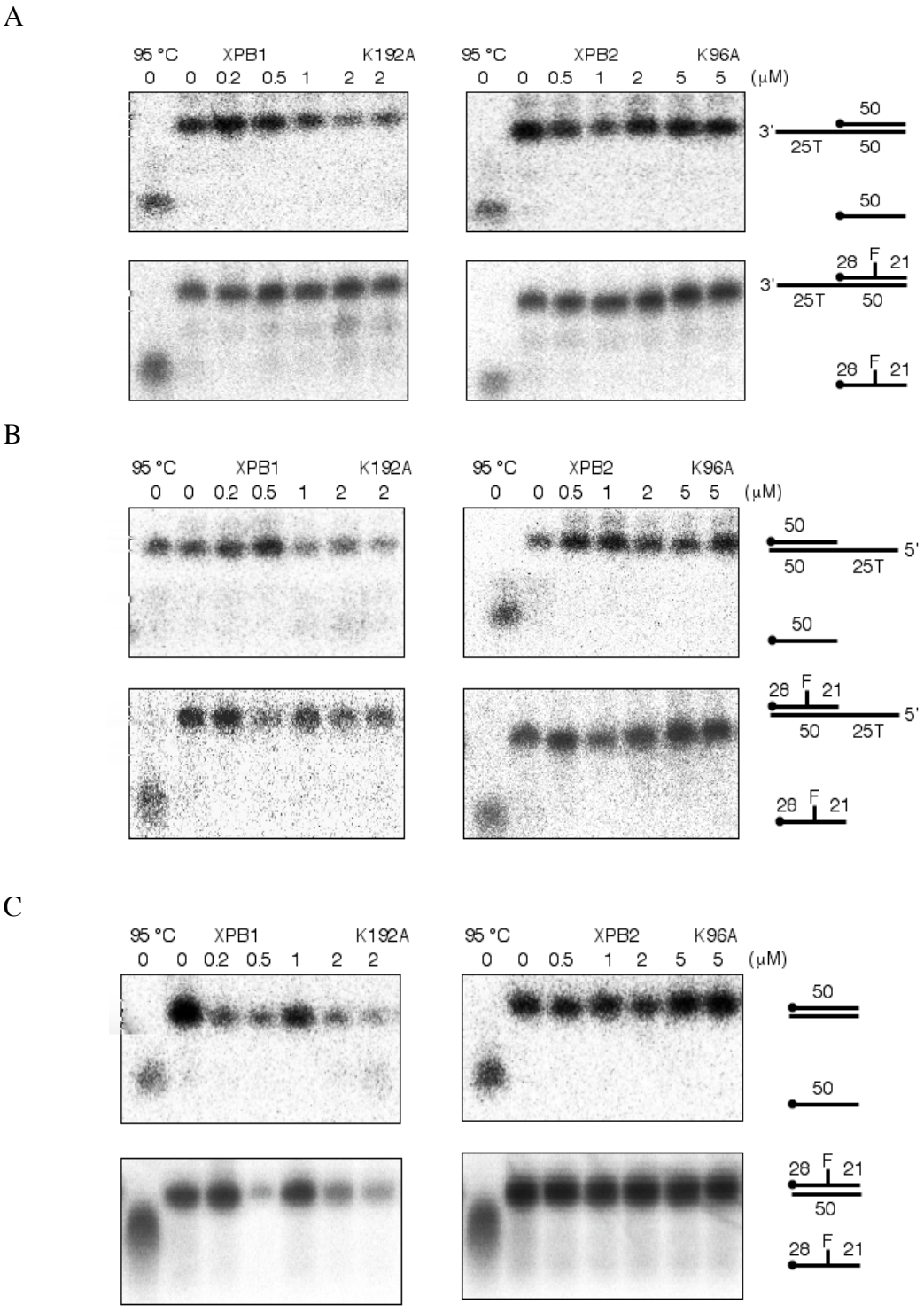


Figure 3.15. The helicase activity of XPB1 and XPB2 with damaged and undamaged DNA
Helicase assays were carried out with XPB1 and XPB2 in the presence of damaged and undamaged DNA substrates (^{32}P-5'-radiolabelled, black circles). The proteins were incubated at the concentrations shown with 10 nM DNA at 45 °C for 30 minutes. The samples were also analysed on 12 % polyacrylamide TBE gels. A) 3' overhang, B) 5'overhang and C) 50mer duplex DNA.

3.10 XPB1 and XPB2 are structured

XPB1 and XPB2 were apparently inactive helicases; XPB2, especially, was highly unstable and prone to precipitating. The secondary structures of XPB1 and XPB2 were studied by circular dichroism (CD) spectroscopy to confirm that the proteins were structured and that this instability was not the result of improper folding. CD spectroscopy is a method by which the difference in absorbance between the two circularly polarized components (L and R) of plane polarized light is measured (reported in terms of ellipticity (θ) in degrees). CD spectroscopy is discussed in more detail in chapter 6.

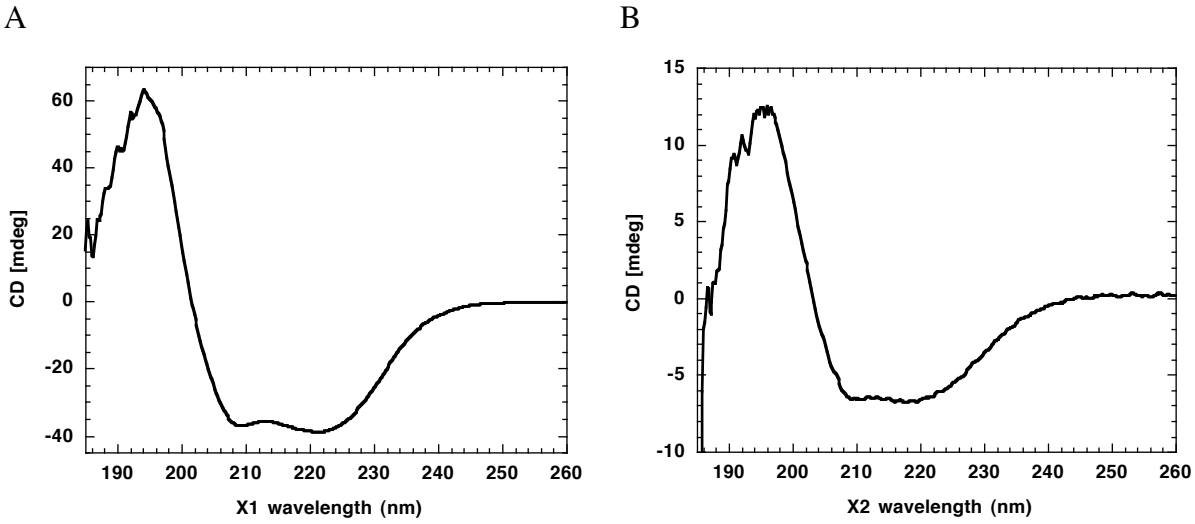


Figure 3.16. A CD absorbance spectrum of XPB1 and XPB2 in the far UV region
Protein was added to 300 μl of sodium phosphate buffer to a final concentration of 1 mg/ml. The absorbance of each was recorded between 260 and 180 nm. The spectra of A) XPB1 and B) XPB2 show the average of six readings automatically averaged by the spectrometer.

The absorbance spectra of XPB1 and XPB2 in the far UV region (<240 nm), corrected for buffer background absorbance, indicated that the secondary structure of both proteins were predominantly α-helical (figure 3.16). The lack of helicase and translocase activity cannot, therefore, be explained by a disordered structure of XPB1 or XPB2. The level of absorbance by XPB2 was very low compared to XPB1, consistent with the instability of XPB2.

3.11 Discussion

The expression, purification and characterisation of XPB1 and XPB2 from the crenarchaeon *S. solfataricus* are described in this chapter. Both XPB1 and XPB2 preferentially bound to ssDNA rather than to dsDNA. The Kd values estimated for XPB2 were much higher than the corresponding binding affinities of XPB1, which were more consistent with other SF2 helicases (Voloshin and Camerini-Otero, 2007). Both XPB proteins were able to hydrolyse ATP in the presence of DNA and, as expected, the Walker A box mutants XPB1 K192A and XPB2 K96A were inactive.

Despite the presence of the seven motifs, characteristic of SF1 and SF2 helicases, in the amino acid sequences of XPB1 and XPB2, neither protein was able to disrupt the interaction between streptavidin and a biotinylated oligonucleotide or separate the strands of duplex DNA. The inability to catalyse strand unwinding is consistent with recently published studies with human XPB (Coin *et al.*, 2007).

Eukaryotic XPB and XPD are both components of the transcription complex TFIIH involved in transcription and NER. Previous biochemical analysis of the roles of XPB (Rad25 in yeast) and XPD (yeast Rad3) in NER have shown that mutation to their respective ATP binding sites dramatically impaired DNA opening around the lesion (Sung *et al.*, 1988; Guzder *et al.*, 1994; Lin *et al.*, 2005; Coin *et al.*, 2007). This merely demonstrated that the hydrolysis of ATP was crucial for strand opening and provided no direct evidence concerning the requirement of the helicase activities of the two enzymes (Coin *et al.*, 2007).

XPB only exhibits very weak helicase activity compared to XPD, although mutations to the helicase motif III or IV of XPB had no effect on its activity in NER or promoter opening (Lin *et al.*, 2005; Coin *et al.*, 2007). Furthermore, the p8 subunit of TFIIH stimulates the ATPase activity of XPB through interactions with p52. This, however, is not accompanied by an increased helicase activity implying that the two are not entirely connected (Coin *et al.*, 2006). It was suggested that helicase activity was a requirement for promoter escape following opening and initiation resulting in the coupling of transcription to RNA processing (Lin *et al.*, 2005).

XPB is not thought to act as a conventional helicase (Kim *et al.*, 2000; Lin *et al.*, 2005). The results presented by Coin *et al.* (2007) favoured the model that describes

XPB as a 'wedge' where XPB wraps around the DNA promoting local unwinding of the duplex and, using ATP, is able to maintain strand separation. This enables correct positioning of the helicase XPD to unwind the DNA further (Kim *et al*., 2000; Coin *et al*., 2007). A second model describes ERCC3-XPB as an ATP-dependent 'ratchet wrench' that facilitates the promoter opening and escape (Kim *et al*., 2000; Lin *et al*., 2005).

It was concluded that, rather than a conventional helicase, XPB acts as an ATP-dependent conformational switch. The process of remodelling the NER preincision complex before the arrival of the endonucleases is supported by the ATPase activity of XPB and, therefore, not only is XPB a source of energy but it also provides a functional role (Coin *et al*., 2006). This role is mirrored at promoter opening, since XPB plays a role in the reorganization of the multiprotein preinitiation complex, which subsequently allows elongation and the preparation of the various RNA processing stages (Lin *et al*., 2005; Coin *et al*., 2006).

In contrast, *A. fulgidus* XPB was shown to be active as a helicase (Fan *et al*., 2006), albeit a weak one. A RED motif was identified and proposed to assist strand separation. The glutamic acid residue was suggested to be crucial in this role since mutation of this residue abolished helicase activity of XPB. *S. solfataricus* XPB1 contains this RED motif but the corresponding sequence of XPB2 is RSD. Initially, this provided the basis of an explanation for the inability of XPB2 (not XPB1) to unwind DNA. However, closer inspection of the XPB protein sequences from a number of eukaryotic and archaeal organisms indicated that this glutamic acid residue was not strictly conserved (figure 3.17).

```
Sso0473         LTATPERSDGRHKLY
Sac1657         LTATPERGDGKEVLY
Sto1613         LTATPERDDGKHELY
Pfu1902         LTAFPERADNLHELF
Afu0358         LTATFEREDGRHEIL
Sso0959         LSATPHREDNKHNEL
Sto1287         LSATPVREDGKHEEL
Sac1326         LSATPYRDDGKHEEL
                *:*   * *. .
```

Figure 3.17. A Clustal W alignment of archaeal XPB 'RED' motif
Clustal W was used to identify sequence conservation in the 'RED' motif in number of archaeal XPB sequences: *S. solfataricus* (Sso), *S. acidocaldarius* (Sac), *S. tokodaii* (Sto), *A. fulgidus* (Afu) and *P. furiosus* (Pfu).

S. solfataricus grows at 80 °C and the helicase assays were carried out at much lower temperatures to prevent spontaneous melting of the DNA substrates. The two *S. solfataricus* XPB proteins may exhibit a weak helicase activity that was not detectable at the temperatures used for the assays presented in this chapter. Although, since the positive control (Hel308) was also a protein from the hyperthermophile *S. solfataricus* (strain PBL2025) and it was able to efficiently unwind DNA at 45 °C and 60 °C (chapter 5) the results imply that neither XPB1 nor XPB2 are active as helicases.

A. fulgidus XPB was shown to preferentially bind to a DNA substrate containing a lesion. This, together with the identification of a DRD in the crystal structure, led to the proposal that XPB played a role in damage recognition (Fan *et al.*, 2006). The data presented in this chapter does highlight subtle differences in the DNA binding affinities and the ATPase activities of both XPB1 and XPB2 with damaged and undamaged DNA. There is not sufficient evidence, however, to support an active function of archaeal XPB in damage detection and the results to date remain inconclusive. The involvement of archaeal XPB in another pathway, for example transcription initiation, remains a possibility.

In eukaryotes, both XPB and XPD are components of a multiprotein complex and are known to interact with regulatory partners, p52 and p44 respectively (Coin *et al.*, 1998; Jawhari *et al.*, 2002; Costa *et al.*, 2003). Each partner protein is responsible for modulating the ATPase activities of XPB and XPD (Coin *et al.*, 2007). This is not uncommon as a number of the putative helicases require protein interactions or modifications, such as phosphorylation, in order to activate any catalytic ability (Singleton and Wigley, 2002). The phosphorylation of eukaryotic XPB, for example, inhibits NER; the unwinding event is not affected but rather the ERCC1-XPF catalysed 5' incision reaction. This is remote from the involvement of XPB in transcription (Coin *et al.*, 2004).

The evidence to support an active role in NER and/or transcription, as seen for human XPB, is so far lacking in *S. solfataricus*. The results presented in this chapter, however, provide some evidence to support the notion that *S. solfataricus* XPB1 and XPB2 act as DNA dependent ATPases rather than conventional helicases.

During the preparation of XPB2 it was evident that this was an unstable protein, especially compared to XPB1. Precipitation of XPB2 was seen at 4 °C and at temperatures of 50 °C and above; the solubility of XPB2 was also reduced in the absence of NaCl. The *S. solfataricus* XPB proteins, especially XPB2, are likely to be lacking vital interacting protein partners responsible for maintaining protein stability and controlling or stimulating catalytic activity. This is investigated further in chapter 4.

CHAPTER 4: A PROTEIN PARTNER FOR *SULFOLOBUS SOLFATARICUS* XPB2

4.1 Introduction

The DNA repair genes are well conserved across the three domains of life, which emphasises the importance of combating the constant assault on DNA. Maintaining the balance, however, between repair and the requirement for a certain level of genetic mutation, to fuel evolution, is absolutely critical (Martins-Pinheiro *et al*., 2004).

Homology is the term intended to describe entities that are derived from a common ancestral characteristic (Thornton and DeSalle, 2000) rather than those that are similar due to convergent or parallel evolution, termed analogues. Homology describes the relationship between two characters that have descended from a common ancestral character and homology is indivisible, that is something is either homologous or it is not. The character can be genic (molecular), structural or behavioural and usually descends with divergence (Fitch, 2000). Homologues may or may not serve the same purpose in different organisms and, therefore, functional similarities are better described as being analogous. Analogous characters are distinguished from those that are homologous as, although they are similar, the characters have descended convergently from unrelated ancestral characters (Fitch, 2000).

There are two major types of homology, the first describes those genes that have a common ancestor, yet as a result of a speciation event, reside in different genomes; these are referred to as orthologues. In contrast, those genes that share a common ancestor as a result of a gene duplication event and are found within the same genome are referred to as paralogues (Thornton and DeSalle, 2000; Hurles, 2004). The genes might descend and diverge while existing adjacent to each other in the same lineage (Fitch, 2000). Paralogues can be found either in clusters or dispersed throughout the genome; the latter correlates with functional diversity and is more common (Hurles, 2004).

Gene duplication can occur as the result of unequal recombination between two repeated elements flanking the gene or by retrotransposition. In eukarya, for example, retrotransposition is the integration of mature RNAs that have been reverse transcribed into the genome at random sites. The resulting gene, therefore, lacks introns and possesses a polyA tail (Hurles, 2004). Gene duplication has been particularly important in the evolution of eukaryotic genomes, which typically contain numerous multigene families. It is this creation of additional gene copies that has played a role in adaptive

evolution by allowing the evolution of new proteins and protein functions (Friedman and Hughes, 2001).

Duplicated repair genes have been identified in some organisms, for example *S. solfataricus* and related crenarchaea possess two homologues of the XPB protein, XPB1 (Sso0959) and XPB2 (Sso0473). The presence of multiple repair genes has been suggested to provide a form of backup that broadens the capacity of the repair pathway (Martins-Pinheiro *et al.*, 2004).

A further type of homology is evident as the result of an interspecies transfer of genetic material and is referred to as xenology (Fitch, 2000). This can also be described as horizontal gene transfer (HGT) or lateral gene transfer (LGT) and is a major force driving the evolution of prokaryotic genomes (Jain *et al.*, 1999). The complexity hypothesis suggests that the complexity of gene interactions, particularly of informational genes (those involved in transcription, translation and other related processes), is a limiting factor that restricts the success of LGT compared to the transfer rates of the operational/housekeeping genes (Jain *et al.*, 1999; Eisen, 2000; Garcia-Vallve *et al.*, 2000). The number of interactions a protein makes with other proteins, for example, is another factor affecting the probability of a successful transfer. Complexity of the interactions, however, may not be the key factor (Eisen, 2000). The likelihood of gene transfer may also be influenced by the sequence conservation between species. Additionally, the information genes are more likely to be required for survival and, therefore, are probably much less likely to be lost from a species (Eisen, 2000).

LGT is a major contributor to prokaryotic evolution as it is crucial for speciation and the molecular evolution of microorganisms; a view that is supported by the genetic relationships between the archaea and bacteria (Garcia-Vallve *et al.*, 2000). Usually only small genetic changes result from mutations, whereas much larger changes are usual with genetic transfer. LGT allows the adaptation of the microorganism to a changing environment and is possibly more important than mutagenesis for the evolution of new functions (Jain *et al.*, 1999; Eisen, 2000; Garcia-Vallve *et al.*, 2000).

There are three main mechanisms by which genetic information is transferred between prokaryotic species: transformation, transduction and conjugation. Transformation describes the uptake of naked DNA from the environment; transduction

involves a bacteriophage that has replicated and packaged random DNA fragments within a donor microoorganism; and conjugation requires the physical interaction between the donor and the recipient cells to mediate the transfer of genetic material (Ochman *et al.*, 2000).

A small number of organisms outside of the bacteria, including *M. thermoautotrophicus*, possess the complete *uvr* operon including the *uvrA*, *B* and *C* genes (Costa *et al.*, 2003). This provides a prime example of LGT between bacteria and archaea.

4.1.1 XPB1 and XPB2 contain a conserved core

The *S. solfataricus* XPB protein paralogues share 28 % and 24 % sequence identity with human XPB and are termed XPB1 and XPB2 accordingly. This level of similarity suggests that the archaeal and eukaryotic XPB proteins share a common ancestor.

Human XPB is divided into three modules: the N-terminal domain, the catalytic core domain and the C-terminal region. This is a similar arrangement to that described for *A. fulgidus* XPB (see section 3.1) (Fan *et al.*, 2006). The helicase motifs of human XPB are all clustered within the catalytic core, the putative DNA binding domain resides in the N-terminal portion and has been shown to correspond to the region that makes contacts with p52 (Jawhari *et al.*, 2002). Human XPB is much longer than either *S. solfataricus* XPB1 or XPB2 (782 versus 551 and 444 amino acid residues, respectively) due to the N-terminal extension.

Comparison of the XPB1 and XPB2 protein sequences emphasised the presence of an additional N-terminal region of approximately 100 residues in XPB1. Upon closer inspection the N-terminal sequence is conserved in other crenarchaeal XPB1 proteins, but not in the eukaryotic XPB sequences. Therefore, it would seem that it is only the catalytic core of the XPB proteins that is conserved throughout the three domains (figure 4.1).

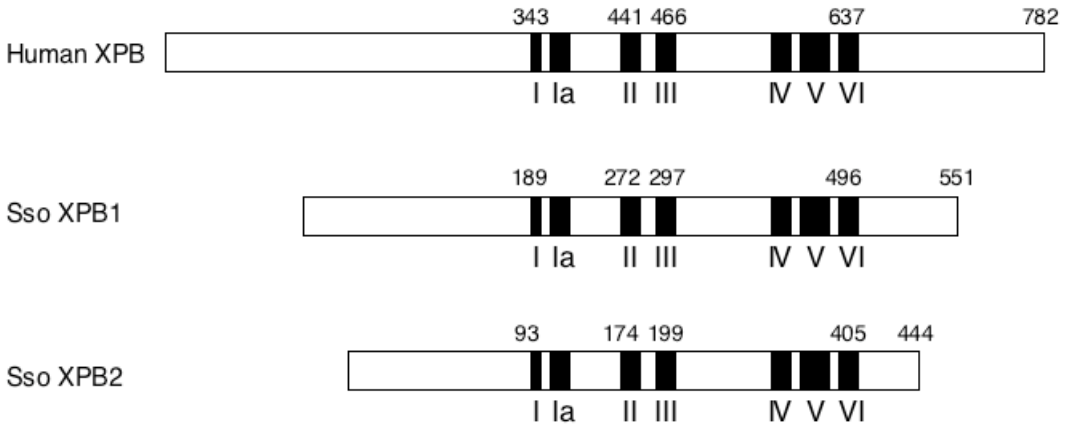


Figure 4.1. Comparison of human XPB with SsoXPB1 and SsoXPB2
The Human XPB sequence comprises an N-terminal extension. An N-terminal extension was identified on XPB1 when compared to XPB2. The sequence of this region, however, is conserved in other crenarchaeal XPB1 proteins, but not in the eukaryotic sequences. Only the catalytic core seems to be conserved.

This chapter focuses on XPB1 and XPB2 from an evolutionary perspective to determine whether these proteins play a similar role in the cell, or whether they have evolved to function separately. The relative expression levels of the XPB1 and XPB2 proteins were studied since differences in the levels of protein expression have functional implications. The gene neighbourhoods around *xpb1* and *xpb2* genes were searched with the intention of finding a protein partner for XPB2.

4.2 XPB1 is more abundant than XPB2 in *S. solfataricus* cells

4.2.1 RT PCR quantification of *ssoxpb1* and *ssoxpb2* mRNA

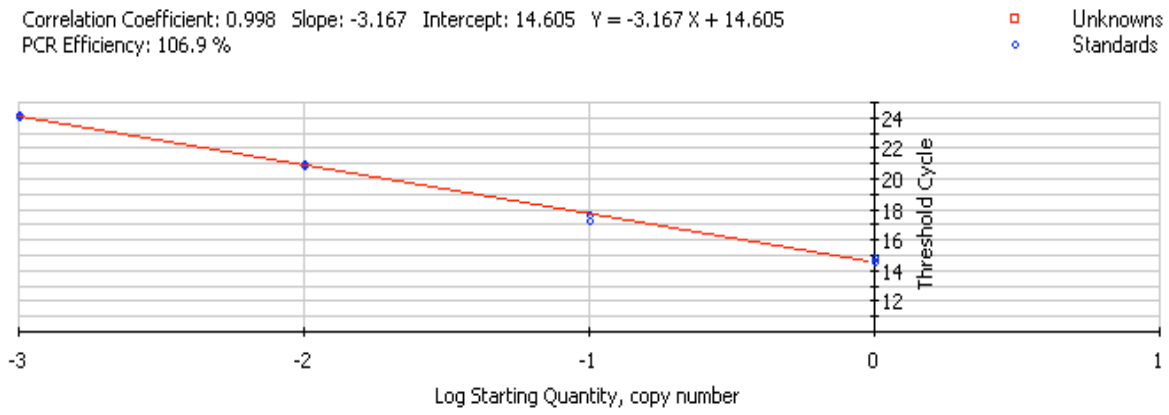
The relative levels of *S. solfataricus xpb1* and *xpb2* mRNA were detected and compared using one-step reverse transcriptase real time (RT) PCR. RT PCR quantification is achieved by measuring a fluorescent signal during the exponential phase of amplification, this is a reliable and sensitive method as the measurements are recorded before limiting factors, for example polymerase inactivation, can affect the reaction (Stratagene, 2005). There are a variety of fluorescent reporter molecules that fall into one of two categories; the first includes the sequence specific DNA probes with a fluorescent dye on one end and a quencher on the other end. The second group includes non-specific DNA binding dyes such as SYBR® Green I. SYBR® Green I displays relatively low fluorescence in solution, upon dsDNA binding its fluorescence increases

1000-fold (Stratagene, 2005). The level of fluorescence is in direct proportion to the DNA concentration and, therefore, can be used to measure the concentration of PCR product in the sample.

mRNA was extracted from a *S. solfataricus* culture in mid log growth phase and used at a final concentration of 1.5 μg/μl in each RT PCR reaction. Gene specific primers were designed to amplify a 100-200 bp region of the *xpb1* and *xpb2* genes. The RT PCR mixture contained reverse transcriptase in the reaction solution in order for both reverse transcription and iTaq polymerase amplification to proceed during the same protocol (see materials and methods for PCR conditions). Negative controls were prepared without RNA to rule out the possibility of contamination.

A standard curve is produced based on the amplification of a standard dilution series of mRNA, the iCycler automatically plots the Ct value of each reaction against the log of the initial template concentration (Stratagene, 2005). The plot should be a straight line if the efficiencies of each reaction are equal. PCR reactions with the *xpb1* and *xpb2* primers were prepared with 130 ng, 12 ng (1/10 dilution), 1.4 ng (1/100 dilution) and 0.3 ng (1/1000 dilution) of template RNA. The primer efficiencies were 1.069 and 1.027 (106.9 % and 102.7 %) respectively (figure 4.2). The values were slightly greater than 1 (100%), which was explained by slight experimental errors, for example subtle pipetting inaccuracies.

A



B

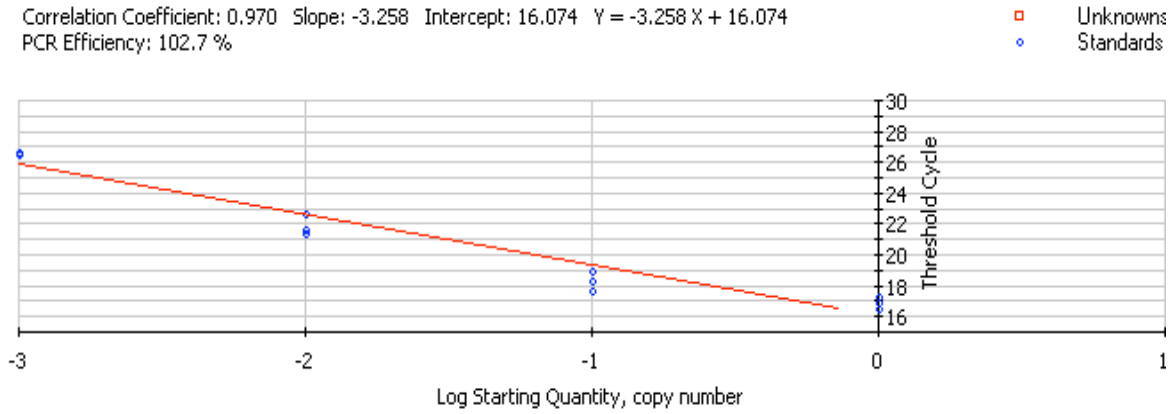


Figure 4.2. Determining the efficiency of *xpb1* and *xpb2* RT PCR primers
PCR reactions with the A) *xpb1* and B) *xpb2* primers were prepared with a standard dilution series of RNA. The Ct values of each reaction were plotted against the log of the initial template concentration and the iCycler automatically calculated the primer efficiencies. The plot is a straight line if the efficiencies of each reaction are equal.

The Ct value is the number of PCR cycles it takes for the amplification level to cross a pre-defined concentration threshold, a limit set by the iCycler using a specific algorithm (Stratagene, 2005). The level of amplification is related to the mRNA concentration in the original sample and, therefore, the Ct values directly relate to the starting mRNA concentration. More specifically, the Ct value is inversely proportional to the mRNA concentration in the samples, since the higher the Ct value the lower the level of PCR product, which corresponds to a lower level of mRNA. The number of PCR copies doubles with each Ct value assuming the primer efficiency is 1 (Schefe *et al.*, 2006).

The Ct values of *xpb1* and *xpb2* were 21.9 ± 0.1 and 24.7 ± 0.03, respectively, and were obtained from triplicate experiments (figure 4.3). The difference between the Ct values corresponds to approximately 2.8 cycles. Since each cycle represented a doubling of the amount of PCR product, this is a $2^{2.8}$-fold difference between the two products. These results, therefore, suggest that the levels of *xpb1* mRNA are 7-fold higher than *xpb2* mRNA. An increase in fluorescence intensity was recorded for the negative control containing the *xpb1* primers, the Ct value for which was ~29. There should not be any amplification in the negative control samples, however, the data is assumed to be reliable if the control Ct value is at least 5, preferably 10, cycles after the last data set. The Ct value of ~29 was seven cycles later than the *xpb1* samples implying that the data was dependable.

SYBR® Green I will bind to any dsDNA in the samples and so to ensure that the fluorescence signal was the result of the amplification of the gene of interest only, a melt curve analysis was performed. Following the PCR reaction the products were heated to 95 °C. As the temperature increases, the dsDNA melts and the fluorescence signal is lost. The temperature at which the product melts is partly dependent upon the length of the duplex and, therefore, a pure sample would result in a single peak. Contaminants of different lengths melt at different temperatures and would be detected by the presence of multiple peaks. The *xpb1* melt curve was consistent with a pure sample, the *xpb2* melt curve looked to show slight contamination, the levels of which were decreased by optimising the initial mRNA concentration (figure 4.3).

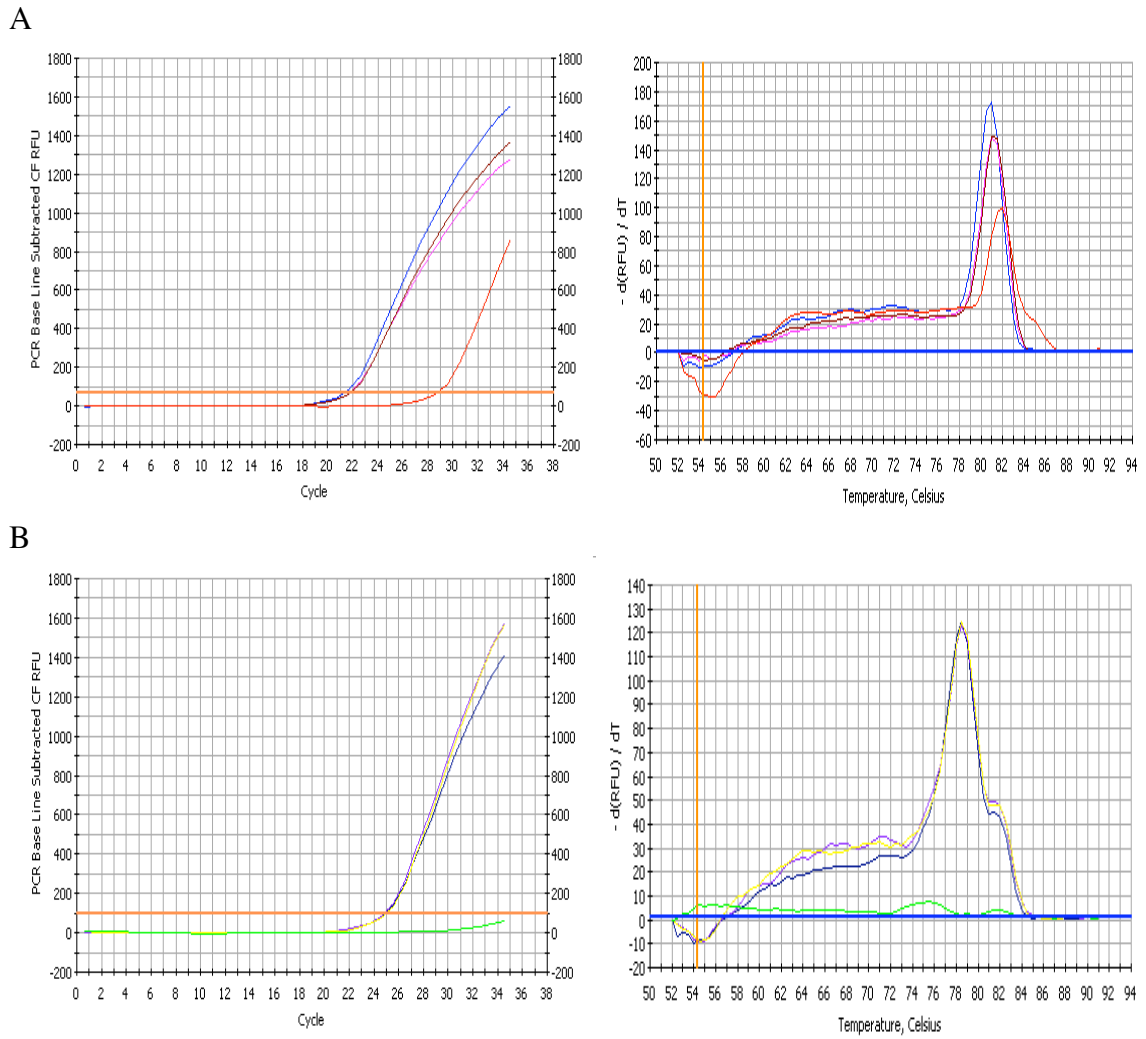


Figure 4.3. Analysis of the relative levels of the *S. solfataricus xpb1* and *xpb2* mRNA
The relative levels of A) *xpb1* and B) *xpb2* mRNA was analysed by one step RT PCR. mRNA was extracted from a growing *S. solfataricus* culture and used in each reaction at a concentration of 1.5 μg/μl. The measurements were carried out in triplicate. Negative controls, minus mRNA, are shown in red for XPB1 and green for XPB2. To identify any contaminating duplex structures the PCR products were heated to 95 °C and the temperatures at which they melted were recorded. The real time amplification (left) and the melt curves (right) are both shown.

4.2.2 Western Blot detection of XPB1 in *S. solfataricus* lysate

Soluble cell lysate prepared from *S. solfataricus* culture, also in mid log phase of growth, was used to detect the levels of the XPB1 and XPB2 proteins. Pure recombinant XPB1 and XPB2 protein samples from 10 ng – 100 ng were prepared and run on an SDS-PAGE gel along with a 222 μg sample of the clear cell lysate. Levels of XPB1 equivalent to 10.2 μg of protein from 1 ml of soluble cell lysate were detected (figure

4.4) with XPB1 specific antibodies. XPB2 expression levels, however, were too low to be detected using this method. The antibodies against XPB2 could not detect 10 ng of recombinant protein on the blot, but 20 ng of XPB2 (and the higher concentrations) showed obvious bands comparable to those seen with recombinant XPB1. This indicated that both antibodies exhibited a similar sensitivity for their respective antigens. Up to 551 μg of clear lysate was analysed, yet still no XPB2 was detected.

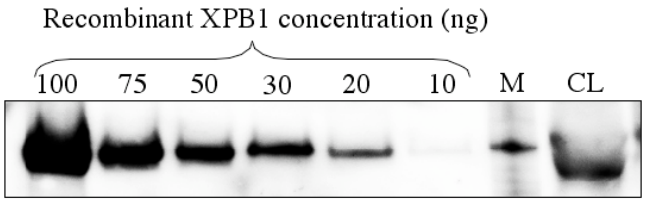


Figure 4.4. Western blot analysis of XPB1 protein in *S. solfataricus* cell lysate
The relative protein level of XPB1 in soluble *S. solfataricus* lysate was analysed by western blot. A dilution series of recombinant XPB1 was prepared from pure recombinant protein to quantify the levels detected in the cell. M represents a 60 kDa protein size marker, and 222 μg of clear lysate was loaded in the lane marked as CL

These results implied that the expression levels of XPB1 were higher than XPB2, consistent with the RT PCR results. These two experiments cannot, however, be directly compared, since analysing the mRNA levels does not account for translation levels of each protein.

4.3 Identification of a protein partner for SsoXPB2

The TFIIH subunit p52 interacts directly with the N-terminal region of XPB and stimulates its ATPase activity (Coin *et al.*, 2007). Therefore, p52 is considered as a regulatory subunit controlling the ATPase activity of XPB in the TFIIH complex (Coin *et al.*, 2007). It was proposed that the C-terminal domain of p52 can adopt a three-dimensional structure that would trap and stabilise XPB inside TFIIH (Jawhari *et al.*, 2002). The tenth subunit of TFIIH, p8/ TTDA, is also involved in the regulation of XPB, albeit indirectly. p8/ TTDA interacts with p52 and is likely responsible for stabilising the interaction between XPB and p52 (Coin *et al.*, 2006; Coin *et al.*, 2007). According to the proposed conformational change of XPB upon ATP hydrolysis (Fan *et al.*, 2006), it is

possible that both p52 and p8/ TTDA regulate this movement through stimulating the catalytic activity of XPB (Coin *et al.*, 2007).

XPB2 was very unstable and prone to precipitating during purification and biochemical assay protocols (chapter 3). Under the assumption that *S. solfataricus* XPB2 required a protein partner for stabilisation and stimulation, similar to eukaryotic p52 (Jawhari *et al.*, 2002), the gene neighbourhood around the two *xpb* genes was examined.

The three *Sulfolobus* species sequenced contain two *xpb* genes in their genomes, a feature that seemed to be confined to the crenarchaea. Closer inspection of a number of archaeal genomes revealed a gene residing adjacent to *xpb2*, named *sso0475* in the *S. solfataricus* genome (figure 4.5). A few bacterial species possess an *xpb* gene in close association with a 'partner'. The distribution is very limited and is, therefore, likely the result of LGT, similar to the previous observation of bacterial NER UvrABC genes in some halophilic and methanogenic archaea (Aravind *et al.*, 1999).

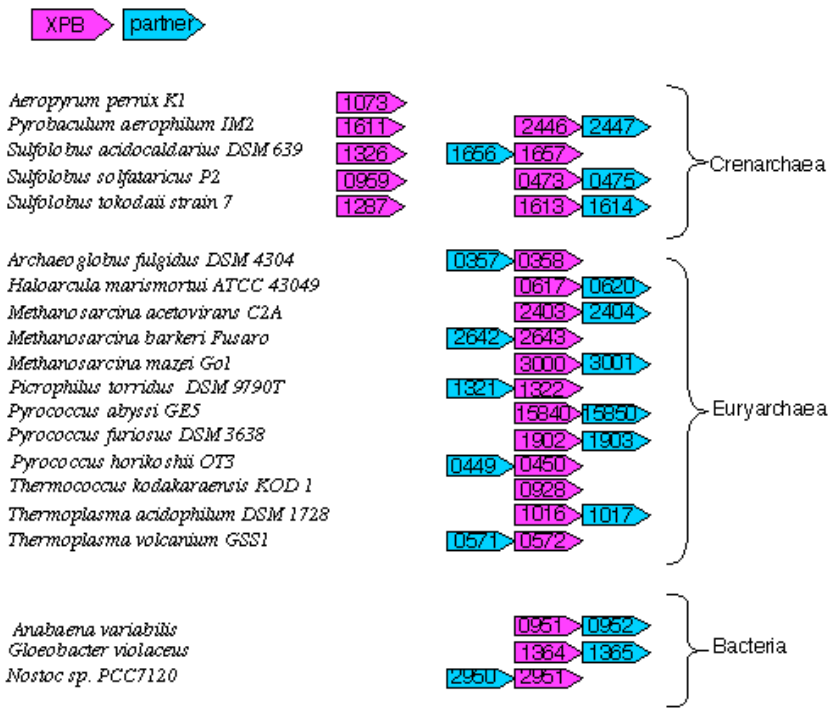


Figure 4.5. A cartoon showing the conservation of *xpb* genes and a possible partner
Close inspection of the archaeal genomes identified a gene residing adjacent to *xpb2* that encodes a possible protein partner for XPB2. A few of the archaeal species contain two copies of the *xpb* gene, although it looks to be mainly confined to the crenarchaea. The few bacterial examples identified are most likely the result of lateral gene transfer.

There is no known function of the Sso0475 protein, or its homologues. Bioinformatic analysis, such as PSI-BLAST, failed to identify significant protein matches with known structures or functions. Such a close association with *xpb2*, however, suggested that these proteins may interact, functionally as well as physically. The presence of the partner gene next to only one of the archaeal *xpb* genes, in those species that possess two, further implicated the two XPB proteins in different cellular functions.

4.4 Expression and purification of Sac1656

sso*0475* was amplified by PCR and cloned into the pET28c and pETDuet plasmids, no protein expression was seen. *sso0475* was also cloned into pETDuet with *xpb2*, although attempts to co-express the proteins were also unsuccessful.

The 'partner' protein from the closely related *S. acidocaldarius* (Chen *et al.*, 2005a), Sac1656, shares 47 % identity and 71 % similarity with Sso0475 protein sequence (figure 4.6).

```
Sso0475         MLTSDLARFKIENQRIIPLFATDSDIDVAKEVIDMFKIGAKVGDILEDLKYLSKIYDYKL
Sac1656         MLPWELARFKIQGSRIIPLFIDNS--EIAQEVLDLFQKGERLGNVIEQAKYLEKIYDYKL
                **. :******:..******  :*  ::*:**:*:*: * ::*:::*: ***.*******

Sso0475         VKGLGKIYLRYCIVESATKIDYIELRRQLFSRGPVLEENDKERVLKEVGDLFHVDPIKAM
Sac1656         VRGLIRLIIRVTSLEEDSPIDPSILRRELFSRGPVFDKKVKEDVIEEIKKKYNIEPIRYM
                *:** :: :*   :*. : **   ***:********:::: ** *::*: . ::::**: *

Sso0475         YEDLDVEKKIVELPKFSPEDLLKIYNLSLLQTIIFNAYKVTVSVSDGWKEIARRIKMLGL
Sac1656         FSDLDEEKRIIEKPNISADELIKWYNLSLLQTLLFKGYRLSIRVSSNWKEIIWRAKLLGL
                :.*** **:*:* *::*.::*:* ********::*:.*:::: **..****  * *:***

Sso0475         MYLAYENPLRIEIFGPLSLVKMTEKYGRNLAALVPFLVSKNKWTIIADIVLGKNKRRTYR
Sac1656         MYYAYKEPLELEFIGPATLVKMNEKYGRNFAVLLPYVVSSPDWSIRAEIVLGKLKKRIYK
                ** **::**.:*::** :****.******:*.*:*::**. .*:* *:***** *:* *:

Sso0475         LELSSGYSKLFKYINDEEMEKRFDSSIEEKFYEEFRRVIRDWNIVREPEPIVIEKRLYFP
Sac1656         LEASN----LKNLREIKIDEKKFDSSIEEKFYNDFTSILKDWKITREPEPLVVKDKIFLP
                ** *.    * :  : :  **:**********::*  :::**:*.*****:*::.::::*

Sso0475         DFVLSKGNIKVYVEIMGFWTKEYVNSKIEKLRNFKYPILVLLNEELSYENYIPDTLNVIK
Sac1656         DFVAEKGKLKVYIEIVGFWTKKYLQNKLEKVSDVENPLLLLVNEELSIDGKIDLPN-IIL
                *** .**::***:**:*****:*::.*:**: :.: *:*:*:***** :. *  .  :*

Sso0475         FKKKIDIGKLYSALRGFQANVNEDID---LSDVNDDIILIKELSAKYNVNEKIVRSKLMQ
Sac1656         YKKKIDVMKVYKWLRDYERQFLQSLPNNIDYTIEGDVVSVSEIANRLSLPLEHLRKNLKN
                :*****: *:*. **.:: :. :.:       ::.*:: :.*:: : .:  : :*.:* :

Sso0475         RPDYIVLKNYAIKKTFIEELKKEDFSNTQLSSLVNKYGNYIVDIIDYLGYKIVWKNISDA
Sac1656         FPGYILLKNYYVKEDLINKIKSYQLEGRKLSDLKKEYGDYIVDVIDYLGYKLMWRGLTDA
                 *.**:**** :*: :*:::*. ::.. :**.* ::**:****:*******::*:.::**

Sso0475         IVEKIKEV
Sac1656         VVVKK---
                :* *
```

Figure 4.6. A Clustal W sequence alignment of Sso0475 and Sac1656
Sac1656 from the closely related species *S. acidocaldarius* shares 47 % identity and 71 % similarity with the XPB protein partner (Sso0475) from *S. solfataricus*.

The *sac1656* gene was cloned into pDEST14 (Gateway® Cloning system) and donated to the project by Dr Huanting Liu (Biomolecular Sciences, St Andrews University).

Sac1656 was successfully expressed in *E. coli* with an N-terminal polyhistidine tag and purified by a combination of nickel column and gel filtration chromatography (figure 4.7). Some obvious contaminants were visible on the SDS-PAGE gel after size exclusion chromatography. Sac1656 was subjected to heat treatment and further purification on a heparin column, although neither of these attempts successfully removed the contaminants. The purification was repeated with chilled lysate to prevent protease degradation and also with a heat step, prior to applying the lysate to the column, to denature any bacterial proteases. This, however, only resulted in a decreased yield of Sac1656. The contaminants were identified as degradation products by mass spectrometry. The interaction of Sac1656 with XPB2 but not the catalytic activity of Sac1656 itself was the main focus of study. For this purpose, further investigation was carried out with the Sac1656 sample that eluted from the gel filtration column (figure 4.7B). The polyhistidine tag was not removed from Sac1656, since this procedure previously proved to be inefficient (Chapter 3 section 3.3, XPB2).

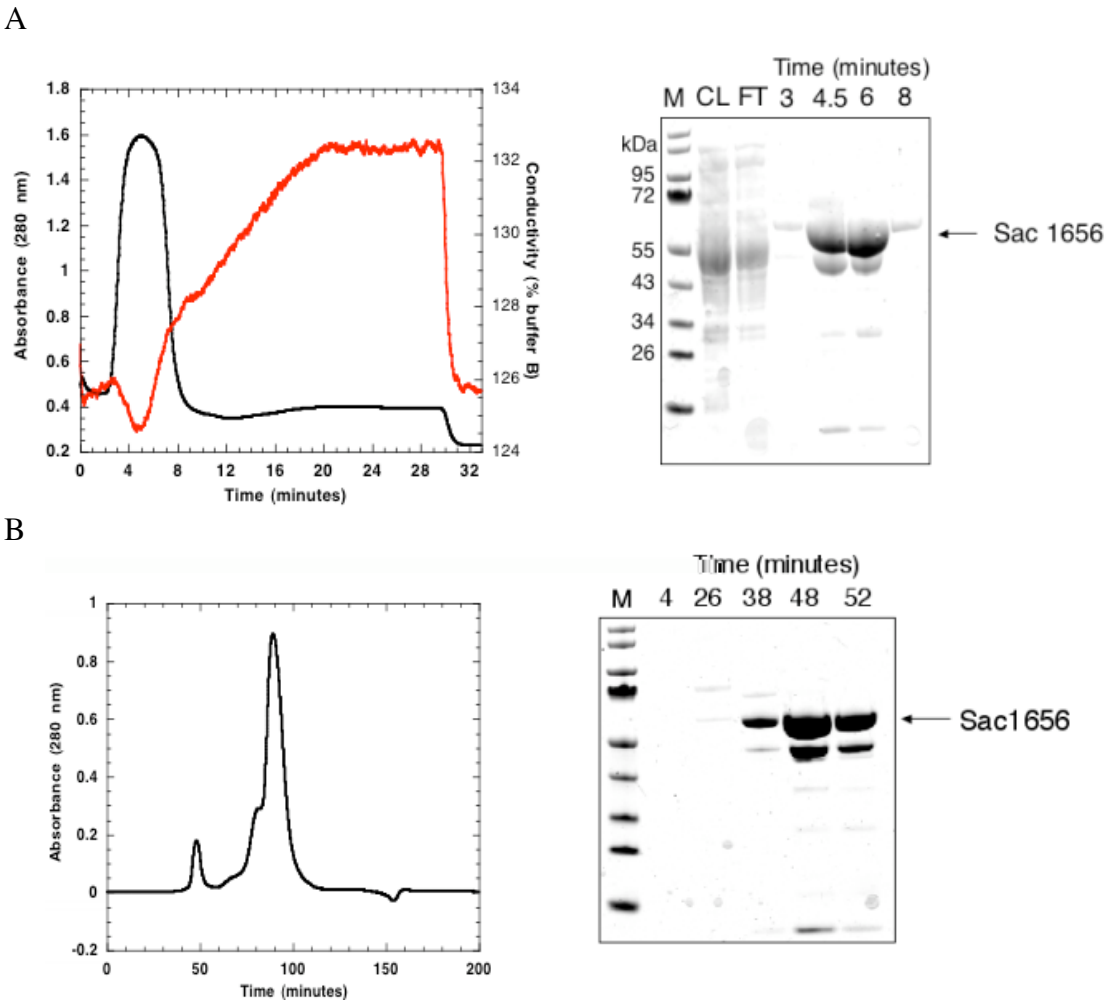


Figure 4.7. Purification of polyhistidine tagged Sac1656
Sac1656 was purified by a combination of A) nickel column and B) gel filtration column chromatography. Soluble *E. coli* expression culture cell lysate was loaded onto an equilibrated nickel column. Sac1656 was eluted from the column across a 30 – 500 mM imidazol gradient. The fractions containing Sac1656 were identified on an SDS-PAGE gel, pooled and purified further by gel filtration. The absorbance spectra and the SDS-PAGE gel analysis of the peak fractions of each column are shown. M represents the marker, CL is the clear lysate sample and FT represents the flow through.

4.5 Sac1656 and XPB2 interact

An Affi-gel column 'pull down' experiment was used to identify a physical interaction between Sac1656 and XPB2. One column, the negative control, was prepared without bait protein and a second, experimental column, was prepared with Sac1656. 3-4 mg of Sac1656 was immobilised on the Affi-gel by covalent interactions with the gel

matrix and the remaining binding sites of the Affi-gel were blocked with 1 M ethanolamine. Both columns were incubated with 3-4 mg of the 'target' protein XPB2.

Each column was washed with 0.1 M NaCl wash buffer to remove any unbound protein and then subjected to 0.25, 0.5 and 1M NaCl buffer elutions. Any eluted proteins were precipitated according to the TCA-DOC protocol, described in materials and methods, and the protein pellets were separated on an SDS-PAGE gel. A sample of the Affi-gel from each column was also run on the SDS-PAGE gel and the samples were analysed by western blot.

A small amount of XPB2 eluted from the column in both the 0.25 M and 0.5 M NaCl washes but no XPB2 was detected in 1 M NaCl elution sample. The control column was used to ensure that XPB2 was not interacting with the Affi-gel itself and since no XPB2 was detected in the control samples, this was a valid assumption. A convincing band representing XPB2 in the gel sample was evident in the experimental column and not the control column, indicating that XPB2 was interacting with Sac1656 (figure 4.8).

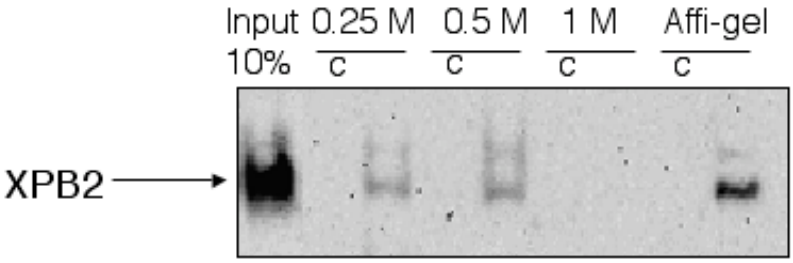


Figure 4.8. XPB2 and Sac1656 physically interact
Sac1656 was covalently immobilised on an Affi-gel column, the remaining binding sites were blocked with 1 M ethanolamine. The column was incubated with pure recombinant XPB2 and washed with 0.1 M NaCl buffer. The column was subjected to 0.25 M, a 0.5 M and 1 M NaCl washes. The samples were precipitated according to the TCA-DOC protocol (materials and methods) and run on an SDS-PAGE gel. A sample of the gel matrix was also run on the gel and all the samples were analysed by western blot. A control column, shown as c, was prepared without Sac1656 to ensure XPB2 was not binding to the Affi-gel.

XPB2 and Sac1656 were passed down a calibrated gel filtration column, both separately and in combination, to examine the interaction in solution (figure 4.9). The elution volumes of known molecular weight standards were plotted against the log of their molecular weights to produce a standard curve that was subsequently used to

calculate the molecular weights of the proteins of interest (the standard curve is presented in appendix 4). XPB2 eluted from the column with a retention volume that corresponded to a molecular weight of 12 kDa compared to an expected molecular weight of 50,817.2 Da. As explained in chapter 3, XPB2 consistently eluted much later than expected and corresponded to a monomeric species of XPB2. The calculated molecular weight of Sac1656 eluted from the gel filtration column was 53 kDa, which was in good agreement with 56,555.3 Da, the expected molecular weight of a monomer. There was an obvious absorbance peak shift when the two proteins were passed through the column in combination. The elution volume correlated with a molecular weight of 98 kDa, which correspond approximately to the sum of the two expected values. The fractions from the centre of each peak were analysed by SDS-PAGE to confirm that the absorbance shift was due to protein interactions and not experimental differences (figure 4.9).

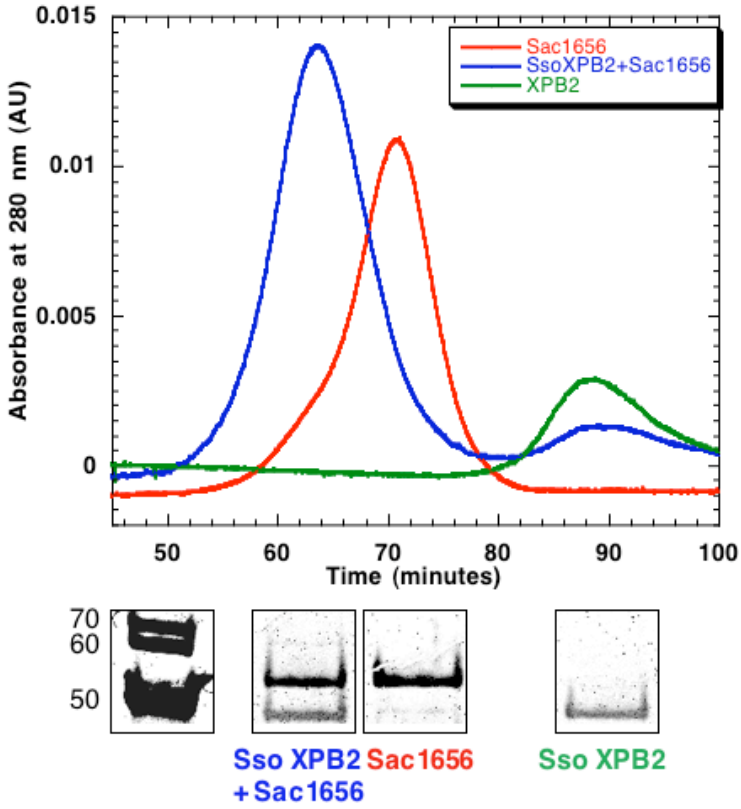


Figure 4.9. Analysis of the Sac1656-XPB2 interaction in solution
A gel filtration column was calibrated with known molecular weight standards. Sac1656 and XPB2 were passed through the calibrated gel filtration column separately and in combination to investigate their interaction in solution. Fractions from the centre of each peak were analysed on an SDS-PAGE gel.

Both the calculated molecular weight values and the SDS-PAGE gel confirmed a stable interaction between Sac1656 and XPB2, consistent with the interaction observed from the pull-down assays. As a consequence, this partner protein Sac1656 will be referred to as Bax1 (**B**inding **A**rchaeal **X**PB). No interaction was detected between XPB1 and Sac1656 by pull-down or gel filtration column analysis. XPB1 was, therefore, used as a negative control in the subsequent assays.

4.6 *S. acidocaldarius* Bax1 does not stimulate *S. solfataricus* XPB2

It was proposed that Bax1 might stimulate XPB2 in a manner analogous to the human XPB-p52 partnership. Consequently, the DNA binding, ATPase and helicase assays were repeated by keeping XPB2 at a constant concentration and increasing the concentration of Bax1. Bax1, however, had no effect on the DNA binding, the ATPase or helicase activities of XPB2; the helicase assay gel is shown in figure 4.10 (otherwise, data not shown).

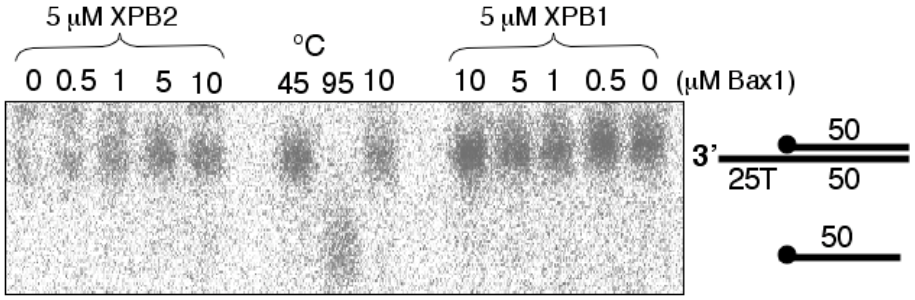


Figure 4.10. The effect of Bax1 on the helicase activity of XPB2
The helicase assays were carried out with 10 nM 3' overhang DNA substrate (^{32}P-5'-radiolabel represented by the black circle) at 45 °C; 5 μM XPB2 was incubated with Bax1 at the concentrations shown, in a total volume of 15 μl. The DNA substrate was incubated at the assay temperature and at 95 °C to show substrate stability and substrate melting, respectively. A control reaction was carried out with Bax1 alone to confirm that Bax1 was unable to unwind DNA. The samples were analysed on a 12 % polyacrylamide: TBE gel.

4.7 Discussion

It is now widely recognised that similarities exist between archaeal and eukaryotic information processing pathways (Bell and Jackson, 1998; Kelman and White, 2005). The archaeal repair pathways, for example, are similar to the eukaryotic system but also possess some unique archaeal-specific features. Most archaeal genomes possess

homologues of repair genes, such as XPB (Grogan, 2000; She *et al.*, 2001). Two *xpb* genes were identified in the *S. solfataricus* genome (She *et al.*, 2001), a feature that seems to be confined to the crenarchaea. This was most likely the consequence of a duplication event, however, interest lay with the function of these paralogues and whether they had evolved to play separate roles within the cell. This provoked investigations to identify differences between *S. solfataricus* XPB1 and XPB2.

Primarily, RT PCR and western blot analysis identified more *xpb1* mRNA and protein, respectively, compared to that of XPB2 in the *S. solfataricus* cell. This suggested a possible functional difference between the two proteins. Secondly, a gene in close association with the *xpb2* gene was present in a number of archaeal genomes examined and also a couple of bacterial species as a result of LGT. Conservation of this arrangement implied that the two gene products may physically and functionally interact. The results presented in this chapter support this proposal since *S. solfataricus* XPB2 and *S. acidocaldarius* Bax1 were shown to form a stable interaction. This conserved partnership was not evident with *xpb1*, suggesting that XPB1 and XPB2 have evolved separately.

Human XPB interacts directly with the p52 subunit of TFIIH, which subsequently stimulates its ATPase activity (Coin *et al.*, 2007). XPB hydrolyses ATP in order to promote local unwinding of the DNA around a DNA lesion, to allow XPD to further unwind the duplex (Coin *et al.*, 2007), and to catalyse promoter opening and escape to prevent premature arrest of the RNA polymerase II complex (Dvir *et al.*, 1997). The regulation achieved by p52 is, therefore, crucial for correct timing of the ATPase activity of XPB. No sequence homologues with a known function have been identified for Bax1. This, however, does not rule out the possibility of a distinct relationship between Bax1 and p52.

In contrast to p52, *S. acidocaldarius* Bax1 was unable to stimulate the activity of *S. solfataricus* XPB2. This may have been due to the fact that the two proteins were from heterologous origins and may present a functional barrier, which the interaction may not be sufficient to overcome. The tenth subunit of TFIIH, p8/TTDA, associates with the TFIIH complex in the presence of NER-specific lesions (Giglia-Mari *et al.*, 2004; Coin *et al.*, 2006; Coin *et al.*, 2007). p8/TTDA stimulates the ATPase activity of XPB by

interacting directly with p52 and stabilising the XPB-p52 interaction (Coin *et al*., 2006; Coin *et al*., 2007). Bax1 could be involved in a role consistent with human p8/TTDA, despite the observation that p8/TTDA does not interact with XPB (Coin *et al*., 2006; Coin *et al*., 2007); Bax1 may function to recruit XPB2 in addition to a role in modulating the catalytic activity of XPB2 in archaeal transcription and/or repair. More information is required to understand the importance of the XPB2-Bax1 interaction.

CHAPTER 5: NEGATIVE REGULATION ACHIEVED BY AUTOINHIBITION OF *SULFOLOBUS SOLFATARICUS* HEL308

5.1 Introduction

Many DNA processing pathways rely on the activity of helicases, which perform a crucial role in all organisms to ensure stable DNA replication and subsequently to maintain genome stability (Sharma *et al.*, 2006).

The helicases of the RecQ family play key roles in maintaining the genomic integrity by stabilising stalled replication forks and removing intermediates of DNA recombination (Sharma *et al.*, 2006). Hypersensitivity to DNA damaging agents, as well as replication defects of a number of *recQ* mutant cell lines implicates them in both DNA repair and replication processes (Sharma *et al.*, 2006). RecQ proteins have been shown to target specialised DNA structures, implying that the different RecQ helicases most likely have specialized unwinding functions that are modified for specific DNA intermediates, especially branched substrates that mimic replication forks and Holliday junctions (Bennett and Keck, 2004; Sharma *et al.*, 2006). One proposed function of this family of proteins, for example, is to prevent deleterious recombination events when the replication machinery is perturbed by DNA damage or secondary DNA structures. RecQ helicase, however, possesses an almost contradictory role, since it functions to initiate homologous recombination via the RecF pathway as well as to suppress "illegitimate recombination" (Bennett and Keck, 2004; Sharma *et al.*, 2006).

The importance of the RecQ helicases is emphasized in humans in particular. Mutations in the RecQ family helicases result in the onset of genome instability syndromes associated with premature aging and/or cancer predisposition (Sharma *et al.*, 2006). Mutations, for example, in three human homologues of RecQ, RECQ4, BLM, WRN are associated with the rare autosomal disorders Rothmund Thomson Syndrome (RTS), Bloom's Syndrome (BS), and Werner's Syndrome (WS), respectively (Bennett and Keck, 2004; Opresko *et al.*, 2004).

Similarly, the Hel308 helicases are SF2 members and have been implicated in DNA repair, recombination and genome stability. The first helicase to be identified in eukaryotes was Mus308 from *D. melanogaster*, a protein involved in the initiation stage of repair of DNA interstrand crosslinks (Harris *et al.*, 1996), which are repaired by excision-recombination mechanisms in prokaryotes (Sladek *et al.*, 1989; Marini and Wood, 2002). Mus308 consists of a polymerase and a helicase domain and was the first

example where homologous regions of two motifs (helicase and polymerase) were apparent on the same polypeptide (Harris *et al*., 1996). The C-terminus of Mus308 and the polymerase domains of the bacterial DNA polymerase I like enzymes share more than 55 % sequence similarity, whereas the N-terminal portion of Mus308 contains the seven motifs characteristic of DNA and RNA helicases. This domain shows some similarity to *E. coli* RecG, particularly in motif V (Harris *et al*., 1996).

A number of sequences characteristically similar to Mus308 were identified, but only one with a biological function; *Saccharomyces cerevisiae* SKI2 antiviral protein (Harris *et al*., 1996). Motif I of Mus308 helicase domain is most closely related to the SKI2 family (Seki *et al*., 2003). SKI2 (super killing) proteins are SF2 RNA helicases involved in the inhibition of translation of viral RNA and also poly (A) and/or uncapped cellular RNA (Dangel *et al*., 1995).

The SF2 Mus308 subfamily, however, is not obviously more closely related to any one of the helicase families since these proteins seem to contain a combination of motifs from a number of the helicase families. Classification of these proteins, therefore, is not so simple (Harris *et al*., 1996). Motif II of the Mus308 proteins seems to fit with the DEXH helicases that have highly diverse functions in repair and recombination and the sequence of motif IV would classify the proteins in the DEAD family of RNA helicases. The sequence of motif V is found in the SNF2/RAD54 family of helicase-like ATPases and, therefore, implicates Mus308 in transcriptional regulation and repair (Harris *et al*., 1996).

DNA polymerases rarely catalyse strand displacement from a DNA substrate with a nick or a small gap without a helicase or other accessory factors. However, both helicase and polymerase catalytic activities reside on the Mus308 polypeptide, which couples the two events and allows the coordination of strand separation and polymerization (Marini and Wood, 2002; Sharma *et al*., 2006).

D. melanogaster Mus308 family homologues have been identified in *Caenorhabditis elegans*, *Arabidposis thaliana* as well as in humans and mice; they have not, however, been found in *S. cerevisiae* or other yeast or bacteria (Marini and Wood, 2002). The domain arrangement of *D. melanogaster* Mus308 was evident in the human orthologue PolQ (figure 5.1), although the region connecting the two motifs is

significantly longer in PolQ (Seki *et al.*, 2003). Human PolQ is an A family polymerase that has unusual functions. It can, for instance, bypass DNA adducts and is also a DNA dependent ATPase (Seki *et al.*, 2003). In addition to the helicase-polymerase fusion protein PolQ, human Hel308 was identified as another orthologue of Mus308 based on its similarity to the helicase portion. Human Hel308 catalyses ssDNA dependent ATPase activity and is able to unwind DNA in a 3'-5' direction with limited processivity (Marini and Wood, 2002; McCaffrey *et al.*, 2006).

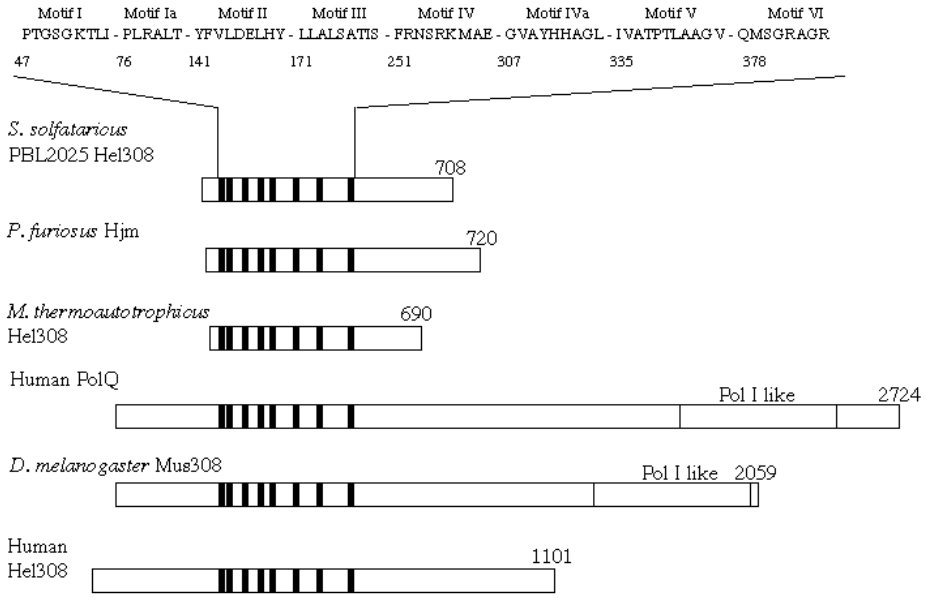


Figure 5.1. A cartoon of archaeal Hel308 (Hjm) and eukaryotic Mus308, PolQ and Hel308 proteins
The N-terminal portions of *Sulfolobus solfataricus* PBL2025 Hel308, *Pyrococcus furiosus* Hjm, *Methanothermobacter thermoautotrophicus* Hel308, Human PolQ, *Drosophila melanogaster* Mus308 and Human Hel308 are shown to contain the helicase motifs I-VI. The sequence of each motif for *S. solfataricus* has been shown at the top of the figure. *D. melanogaster* Mus308 and the human orthologues PolQ both comprise predicted helicase and polymerase domains and they both possess a family A DNA polymerase (bacterial pol I)-like sequence in their C-terminal regions. The Hel308 proteins (including Hjm) were identified by homology to the putative helicase domain of Mus308. This figure was adapted from Fujikane *et al.* (2005) and Guy *et al.* (2005).

Archaea possess clear Hel308 homologues also (figure 5.1); biochemical analysis of Hel308 from both *Methanothermobacter thermoautotrophicus* and *Pyrococcus furiosus* (called Hjm) has implicated Hel308 in replication fork repair and replication restart. Since most archaeal species lack homologues of RecQ, Hel308 may perform a crucial role in genome maintenance and stability (Guy and Bolt, 2005). Futhermore, the functional interaction of *P. furiosus* Hjm with PCNA via a C-terminal PIP box suggests

that Hjm could be recruited to a stalled replication fork by PCNA in order to prepare the stalled fork for restoration (Fujikane *et al.*, 2006). *M. thermoautotrophicus* Hel308 and *P. furiosus* Hjm both function and exhibit substrate specificity like RecQ in a genetic screen for synthetic lethality in *E. coli* (Guy and Bolt, 2005; Fujikane *et al.*, 2006), which reinforces the RecQ like properties of these two proteins.

A homologue of Hel308 was identified in *S. solfataricus*. The focus of this chapter is the structural and biochemical analysis of *S. solfataricus* strain PBL2025 Hel308, referred to as Hel308. The main focus is on domain 5 of Hel308, since removal of this resulted in an increase of helicase activity. This led to the proposal that domain 5 acts as an autoinhibitory domain that may be relieved upon interaction with DNA or other proteins.

5.2 Crystal structure of Hel308

Recently, the crystal structures of Hel308 from *Archaeoglobus fulgidus* (Buttner *et al.*, 2007) and *Sulfolobus solfataricus* PBL2025 (Richards *et al.*, 2007 in press) have been solved and are in good agreement with each other.

The crystal structure of Afu Hel308 (figure 5.2), in complex with a 15 mer duplex with a 10 nt single stranded region, captures initial strand separation (Buttner *et al.*, 2007). The overall structure of Sso PBL2025 Hel308 (figure 5.3A) is very similar to Afu Hel308, although the proteins only share 39 % identity (Richards *et al.*, 2007 in press).

Hel308 is organized into 5 domains, domains 1 and 2 contain the composite ATP binding sites (RecA-fold) at their interface and domain 3, which is tightly packed against domain 1, contains a winged helix fold. Domain 4 consists of a seven-helix bundle, which is relatively rarely identified in proteins and, therefore, no function has yet been assigned to this domain (Buttner *et al.*, 2007; Richards *et al.*, 2007). Together, domains 1-4 form a central pore through which the DNA passes. The ring, however, is not covalently closed, the gap between domains 2 and 4 provides flexibility for Hel308 to load onto DNA substrates lacking a 3' DNA end, such as stalled replication forks (Richards *et al.*, 2007). There is a long loop connecting domains 2 and 3 (416-426), which could act as a hinge for this movement (figure 5.3A), however, this is not conclusive.

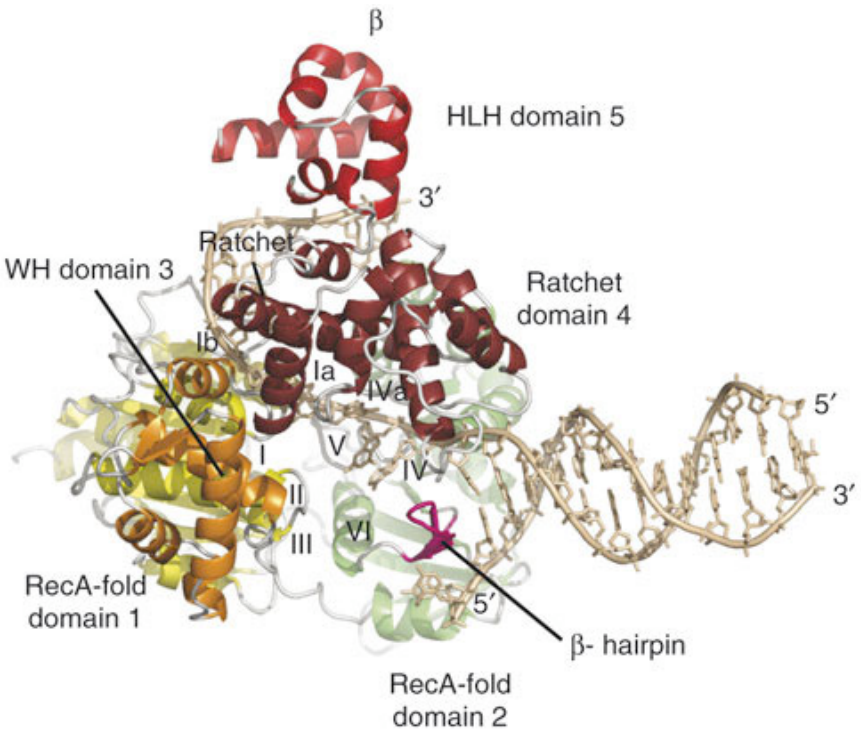


Figure 5.2. A crystal structure of *Archaeoglobus fulgidus* Hel308
Ribbon structure of Afu Hel308 in complex with 15 mer duplex DNA with a 10 nt single stranded region (shown as beige sticks). Domain 1 (1-199) is shown in yellow, domain 2 (200-398) is represented in mint green, domain 3 (399-485) is coloured in orange, domain 4 (486-625) in maroon and domain 5 (626-691) is shown in red. Roman numerals represent the SF2 helicase motifs and the β-hairpin is coloured pink (Buttner *et al.*, 2007).

Domain 5 consists of two helix-hairpin-helix motifs, commonly associated with DNA binding and it is positioned behind the central pore similar to a cap (Richards *et al.*, 2007). The RAR motif in this domain was identified on an α-helix and is conserved in all the Hel308 family proteins.

The Sso PBL2025 Hel308 apo structure superimposes well onto the DNA bound structure of Afu Hel308 suggesting that there is very limited movement of Hel308 upon DNA binding (Buttner *et al.*, 2007). Also, four sulphates present in the crystal structure of Sso PBL2025 Hel308 indicated possible binding sites of the phosphate residues in the DNA backbone. The DNA, therefore, could be modelled into the Sso PBL2025 Hel308 structure (figure 5.3B, C). The DNA duplex is split by a β-hairpin loop that sits almost at the entrance of the channel. This β-hairpin loop was identified in both structures. The DNA is threaded through the channel forming interactions with domains 3 and 4, engaging domain 5 upon exiting the pore (Buttner *et al.*, 2007).

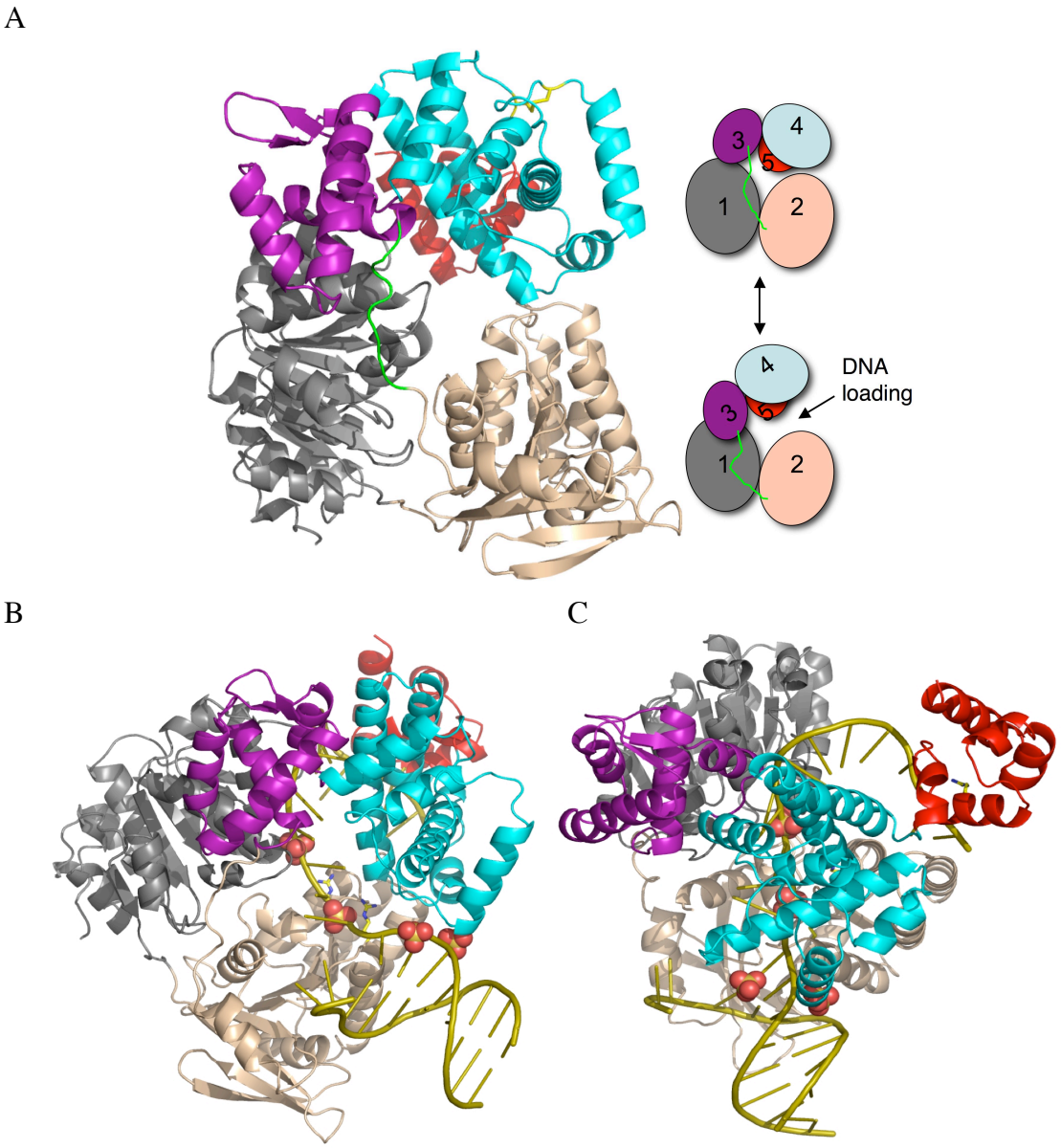
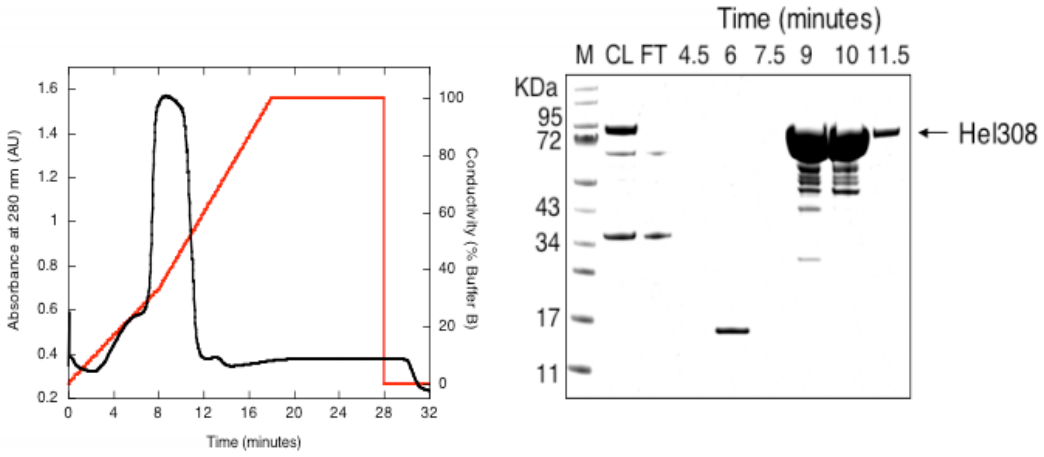


Figure 5.3. A crystal structure of *Sulfolobus solfataricus* PBL2025 Hel308
A) Ribbon structure of *S. solfataricus* PBL2025 Hel308. Domain 1 (1-199) is shown in grey, domain 2 (200-416) is coloured wheat, domain 3 (426-501) is purple, domain 4 (501 – 646) is shown in cyan and domain 5 (648-705) is red. The green line represents the extended linker (417-425) connecting domains 2 and 3. Yellow indicates the disulphide bond linking the only two cysteine residues in the sequence. The cartoon shows the arrangement of the Hel308 domains in the closed conformation and the proposed opening, which exposes the interface between domains 2 and 4. B) Ribbon representation of Hel308 modelled with DNA double helix. The domains are coloured as described in A. The space filling spheres (oxygen is shown in red and sulphur in yellow) represent the experimentally located sulphate ions, the mutated residues R320, R255 and R662 are shown as sticks. C) The same figure as B, Hel308 is rotated 90 ° about a horizontal axis in the plane of the paper. The figures were taken from Richards, *et al*. (2007, in press).

5.3 Expression and purification of PBL2025 Hel308

The *hel308* gene was amplified by PCR from *S. solfataricus* PBL2025 genomic DNA and cloned (Gateway® cloning system) into the expression vector pDEST14, with an N-terminal polyhistidine tag. This construct was provided by Dr Huanting Liu (Biomolecular Sciences, St Andrews University) and the gene sequence was verified by DNA sequencing (The Sequencing Service, School of Life Sciences, Dundee). Purification of the *hel308* gene product, Hel308, was achieved by a combination of heat treatment, immobilized metal affinity chromatography and finally size exclusion on a gel filtration column (figure 5.4). The protein was confirmed as *S. solfataricus* PBL2025 Hel308 by mass spectrometry.

A



B

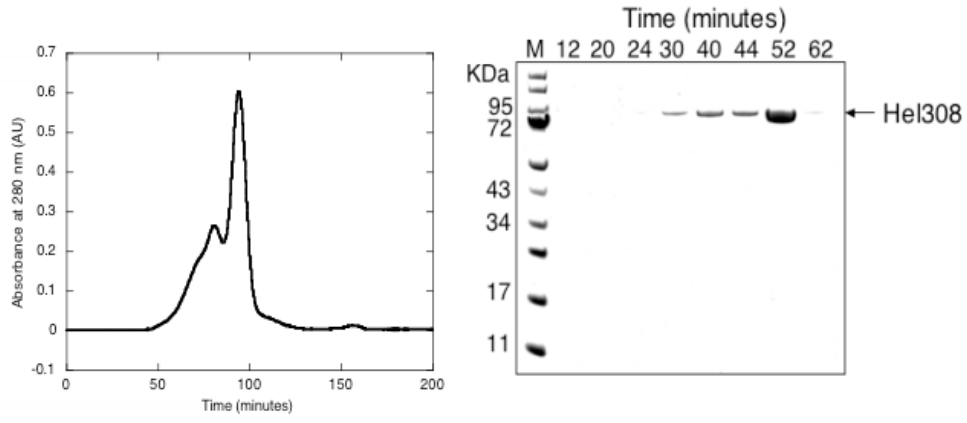
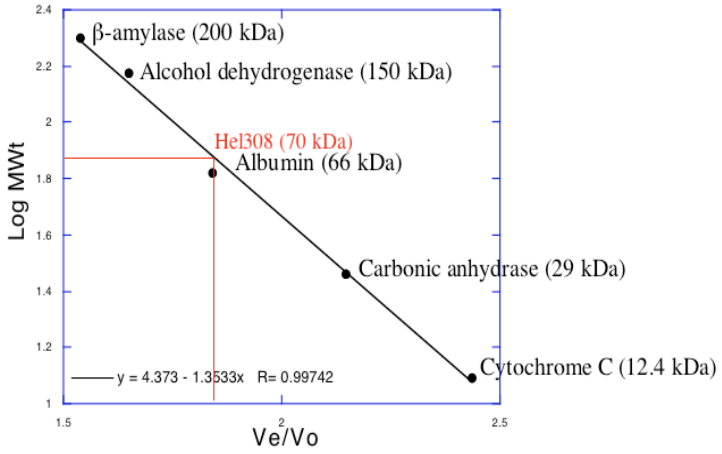


Figure 5.4. Purification of polyhistidine tagged Sso PBL2025 Hel308
Hel308 was immobilised on a metal affinity column and eluted from it across a 30 mM-500 mM imidazol gradient. A) The absorbance trace and SDS-PAGE gel analysis of the peak fractions from the nickel column. CL represents the clear lysate (soluble) and FT is the column flow through. B) The absorbance spectrum and SDS-PAGE gel analysis of the peaks eluted from the gel filtration column. M represents the size marker.

The oligomerisation state of a helicase provides information regarding its mechanism of action (Lohman and Bjornson, 1996; Marini and Wood, 2002). Hel308 was passed through a calibrated gel filtration column and eluted with an approximate molecular weight of 70 kDa, which corresponded to a monomeric species (figure 5.5A). This was in good agreement with the actual molecular weight of the monomer 84,456.0 Da, as determined by mass spectrometry (figure 5.5B), and the expected molecular weight of 81,476.4 Da.

A



B

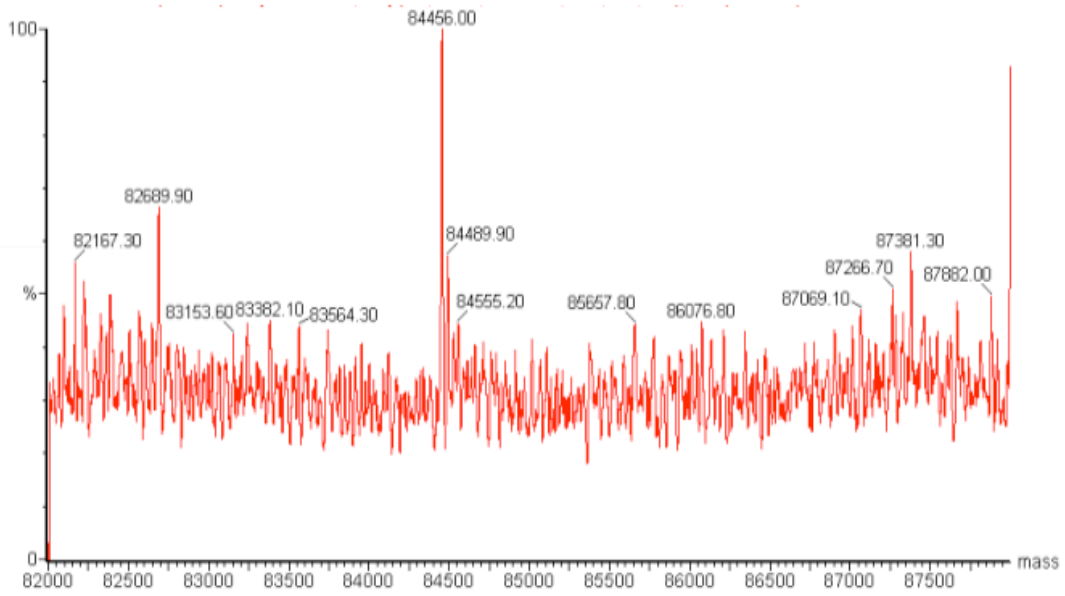


Figure 5.5. Size determination of Hel308

A.) Standards of known molecular weight were used to calibrate the gel filtration column. Their elution volumes divided by the void volume (retention time of blue dextran, the biggest standard) were plotted against the log of their molecular weights. The resulting standard curve was used to estimate the size of Hel308, according to its elution volume. B.) The intact molecular weight of Hel308 was confirmed by mass spectrometry; the major peak represented a protein with a molecular weight of 84,456.0 Da (data was processed to 0.1 Da) compared to an expected molecular weight of 81,476.4 Da.

5.4 Site-directed mutagenesis of Sso Hel308

The crystal structure of Hel308 enabled the identification of key amino acids involved in DNA binding, these amino acids are indicated in table 4. The ring formation adopted by domains 1-4 of Hel308 provides a passage for ssDNA through the centre of the molecule. Three conserved arginine residues R155, R255 and R320 protrude into the centre of this channel, which suggests an involvement in DNA binding. To investigate this proposal each of these arginine residues was mutated to an alanine, by site directed mutagenesis, to give the mutants R155A, R255A and R320A.

Domain 5, which does not contribute to the ring structure, consists of two helix-turn-helix motifs. Such motifs are known to be involved in DNA binding (Luscombe *et al.*, 2000), thereby implicating domain 5 in a DNA binding role. Domain 5 contains an arginine-alanine-arginine (RAR) motif that is conserved in the related helicases PolQ, Mus308 and human Hel308, suggesting that this motif is important for functional interactions relevant for the eukaryotic enzymes (Buttner *et al.*, 2007). The first arginine of the conserved RAR motif was one of two residues targeted to domain 5; it was also mutated to an alanine, R662A. The second was K646, which was mutated to a STOP codon, this resulted in the removal of the entire domain 5 (K646STOP). Domain 5 alone was also constructed using The Companion™ pET Directional TOPO® Expression Kit (materials and methods). The Walker A box mutant K52A was prepared as a negative control for ATPase and helicase activity. All of the primers used for site directed mutagenesis are listed in appendix 1.

Table 4. A description of Hel308 mutagenesis
Mutagenesis was achieved using the WT plasmid construct as the template.

Hel308 mutagenesis		
Residue	**Description**	**PCR conditions**
K52A	Canonical Walker A box residue	Pfu Ultra II Fusion Polymerase: 95 °C 2' (X1); 95 °C 20", 65 °C 20", 72 °C 2'30" (X30); 72 °C 3' (X1)
R155A	Conserved arginine in the protein channel, possible involvement in DNA binding	Pfu Ultra II Fusion Polymerase: 95 °C 2' (X1); 95 °C 20", 58 °C 20", 72 °C 2'30" (X30); 72 °C 3' (X1)
R255A	Conserved arginine in the protein channel, possible involvement in DNA binding	Pfu Polymerase: 95 °C 2' (X1); 95 °C 30", 59 °C 30", 72 °C 18' (X30); 72 °C 10' (X1)
R320A	Conserved arginine in the protein channel, possible involvement in DNA binding	Pfu Turbo Polymerase: 95 °C 2' (X1); 95 °C 30", 59 °C 30", 72 °C 9' (X30); 72 °C 10' (X1)
K646STOP	Reisdue at the end of domain 4	Pfu Ultra II Fusion Polymerase: 95 °C 2' (X1); 95 °C 20", 58 °C 20", 72 °C 2'30" (X30); 72 °C 3' (X1)
R662A	One of three conserved arginine residues in domain 5 thought to be involved in DNA binding to promote ssDNA dependent ATP hydrolysis	Pfu Polymerase: 95 °C 2' (X1); 95 °C 30", 57 °C 30", 72 °C 18' (X30); 72 °C 10' (X1)
Domain 5	Domain 5 alone	Platinum® PCR Supermix High Fidelity Polymerase: 95 °C 2' (X1); 94 °C 30", 50 °C 30", 68 °C 30" (X30); 68 °C 5' (X1)

The WT and mutant genes were sequenced (The Sequencing Service, School of Life Sciences, Dundee) using the T7 primers and two internal primers (appendix 1) in order to cover the whole gene. The mutants were expressed and purified as described for WT Hel308 (figure 5.6) and each mutation was confirmed by mass spectrometry. No expression of Hel308 R155A mutant was seen, so further experiments were carried out on Hel308 WT and the remaining mutants.

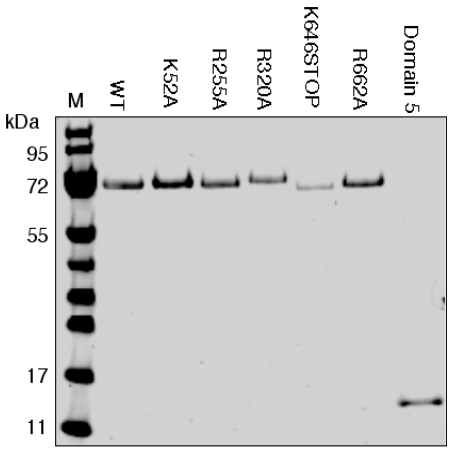


Figure 5.6. Hel308 WT and mutant protein
SDS PAGE gel of pure recombinant WT Hel308 and mutants: K52A, R255A, R320A, R662A, truncation mutant K646STOP and domain 5. M represents the size marker.

5.5 Mutation of conserved arginine residues reduces DNA binding

The DNA binding properties of WT Hel308 and the mutants were investigated using fluorescence anisotropy as an indicator of complex formation. Anisotropy provides an accurate measurement of binding interactions in solution. Anisotropy is more precise than methods involving gel separation of bound and unbound product, since these disturb the reaction equilibrium and result in inaccurate quantification (Lundblad *et al.*, 1996). The measurements are based on the rotation of fluorescein labelled macromolecules in solution. Fluorescent molecules absorb light in a particular plane, termed the absorption dipole. When this absorption dipole is parallel to the plane of the light, maximum absorption is achieved. The emission dipole, however, differs from the absorption dipole and therefore the emitted light is depolarized. Any additional depolarization of the light is due to rotational movement of the labelled molecule and the difference between the two provides the basis for quantification of complex formation (Lundblad *et al.*, 1996). The difference between the emitted vertical and horizontal fluorescence intensity divided by the total intensity defines the anisotropy (Hey *et al.*, 2001; Reid *et al.*, 2001). The G (grating) factor is automatically calculated for each measurement and the fluorimeter routinely uses this value to correct for any difference in the response to vertical and horizontal polarized light (Reid *et al.*, 2001). All fluorescence intensity measurements were taken using 'magic angle' conditions where the excitation polariser is in the vertical

orientation and the emission polariser is positioned at 54.7° from vertical in order to enhance the horizontal fluorescence intensity ($I_{\perp}$) two-fold over the vertical fluoresecence intensity ($I_{\parallel}$). This results in the correct sum for the total fluorescence intensity (I_T) (Lakowicz, 1983; Lundblad *et al*., 1996):

$$I_T = I_{\parallel} + 2\, I_{\perp}$$

Rotational motion is greatest in smaller molecules due to an increase in mobility. The observed anisotropy is reduced by rotation of the fluorescein labelled molecule between excitation and emission, according to the Perrin equation shown below:

$$A = \frac{A_0}{1 + (\tau/\phi)}$$

Where A is the observed anisotropy, A_0 is the limiting anisotropy, τ is the fluorescence lifetime (the time between excitation and emission) and ϕ is the rotational correlation time (Lundblad *et al*., 1996).

The anisotropy value of a molecule that rotates rapidly, with respect to the fluorescence lifetime, approaches zero. When the molecule rotates slowly, with respect to the fluorescence lifetime, the anisotropy reading approaches the limiting anisotropy. The measured anisotropy is the sum of free and bound molecules (Lundblad *et al*., 1996).

Binding of Hel308, WT and mutant proteins, to a single stranded 15 mer (the sequence is shown in appendix 2) with a 5'-fluorecein dye was investigated by direct titration (figure 5.7A), the data was fitted to the equation described in the materials and methods. Hel308 WT bound ssDNA (15 nts) with a Kd of 0.14 μM whereas binding at levels almost 40-fold weaker were evident with a dsDNA substrate (15mer). This result was consistent with previous biochemical analysis of *M. thermoautotrophicus* Hel308, which was unable to form stable complexes with duplex DNA (Guy and Bolt, 2005). The same oligonucleotide was used to prepare the ssDNA and dsDNA (plus complement) substrates to avoid sequence bias.

As expected, K52A bound DNA much like wild type protein, whereas both R255A and R320A presented reduced DNA binding. Interestingly, however, the Kd value of R320A was 0.88 μM whereas R255A showed a huge decrease in affinity for ssDNA with a Kd of 3.7 μM. This provided strong evidence to confirm the ssDNA binding role proposed for R320 and particularly R255, both of which reside in the central channel of

Hel308 through which the ssDNA is threaded. The K646STOP truncated protein and the R662A mutant exhibited DNA binding comparable to WT, a result that was surprising since domain 5 possesses typical DNA binding motifs. Domain 5 alone bound to ssDNA although the affinity was much lower than WT Hel308. This was expected since removal of domain 5 did not impair DNA binding suggesting the remaining domains account for most for the binding observed. The Kd values are presented in table 5.

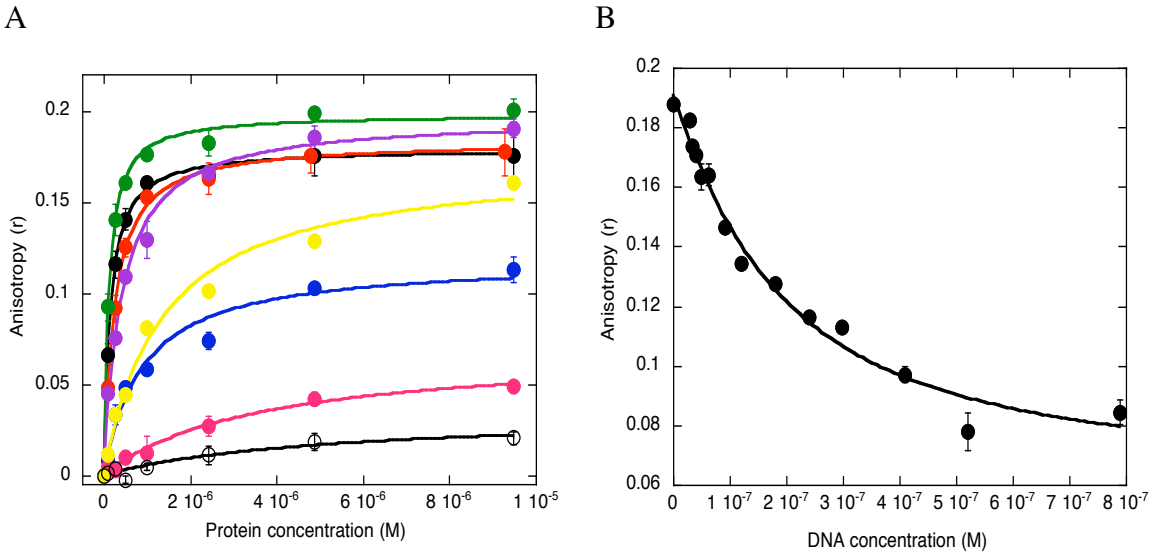


Figure 5.7. Binding of Hel308 WT and mutants to ssDNA, measured by anisotropy
The binding reactions were carried out at 20 °C; A) the protein was directly titrated into the assay mix containing 20 nM DNA substrate. The starting volume was 153 μl and the concentrations were corrected according to changes in the volume upon titration. The mean of triplicate measurements was plotted and the standard errors are shown. WT is shown in black; K646STOP in red; K52A in green; R320A in blue; R255A in pink; R662A in purple and domain 5 in yellow. Closed circles represent ssDNA and open circles represent dsDNA (WT).
B) Competition assay: 1 μM WT Hel308 was incubated with 100 nM fluorescently labelled ssDNA. Unlabelled ssDNA competitor was titrated into the mixture and the anisotropy measurements recorded. The results plotted were the average of three data sets and the standard errors are shown. See materials and methods for curve fit.

Table 5. The accurate Kd values for Hel308, WT and mutants
The DNA was incubated at a concentration of 20 nM and the assays were carried out at 20 °C

Anisotropy	ssDNA 15 nts	dsDNA 15mer
	Kd (μM)	Kd (μM)
Hel308 WT	0.14 ± 0.01	5.3 ± 3.1
Hel308 K646STOP	0.23 ± 0.01	3.7 ± 1.5
Hel308 K52A	0.10 ± 0.01	8.5 ± 2.6
Hel308 R320A	0.88 ± 0.2	8.8 ± 5.5
Hel308 R255A	3.7 ± 0.7	-
Hel308 R662A	0.40 ± 0.04	-
Hel308 Domain 5	1.3 ± 0.2	-

A decrease in the fluorescence intensity of up to 30 % was observed. Changes to the local environment of the fluorophore resulting, for example, from interactions between the protein and the fluorescein dye can affect the fluorescence intensity. This in turn could influence complex formation (Hey *et al.*, 2001). To verify that the molecular interaction was not influenced by the fluorescein moiety, competition assays were performed with WT Hel308. Hel308 was incubated with 100 nM fluorescently labelled ssDNA at a final concentration of 1 μM. An unlabelled competitor DNA strand was titrated into the mixture to induce binding free from the influence of the fluorescein (Reid *et al.*, 2001). Both the labelled and unlabelled strands were the same sequence to avoid sequence bias.

The decrease in anisotropy observed (figure 5.7B) confirmed a DNA specific interaction; the Kd of 0.13 μM was in good agreement with that calculated from the direct titration data. The protein-DNA interaction was not influenced by the fluorescein moiety and local changes around the fluorophore, such as a change in the pH of the immediate environment rather than a direct influence from the protein molecule, could account for the change in fluorescence intensity (Lundblad *et al.*, 1996).

5.6 Hel308 is a DNA dependent ATPase

Helicases require nucleotide binding and hydrolysis for strand separation (Sharma *et al.*, 2006). The ATPase activities of WT Hel308, K646STOP, K52A and R320A were

investigated in the presence of ssDNA since the binding affinity of Hel308 for dsDNA was poor. The proteins were incubated, at a concentration of 0.5 μM, with 10 nM PhiX174 virion (ss) DNA and samples were taken across a time course (figure 5.8A).

The ability of WT Hel308 to hydrolyse ATP was a DNA dependent reaction, since the levels of free phosphate in the absence of the DNA substrate were minimal.

A B

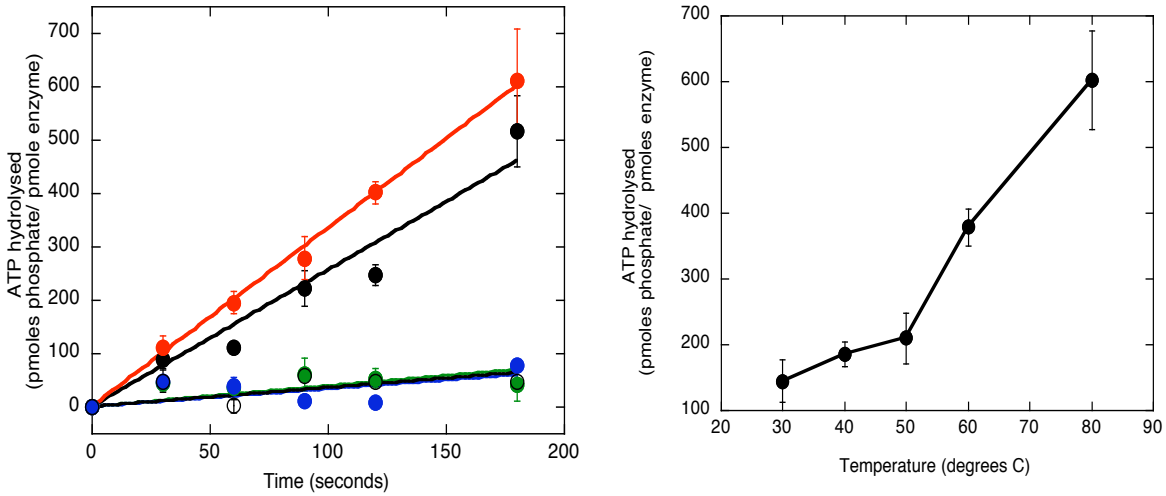


Figure 5.8. ATPase activity of Hel308 WT and mutants
A) The ATPase activity of Hel308 was measured by incubating 0.5 μM protein with 10 nM PhiX174 virion (ssDNA) in a final volume of 240 μl. The reactions were carried out at 60 °C and the reaction was started by the addition of 0.5 mM ATP-$MgCl_2$. The samples were taken at the time points indicated and added to 0.3 M PCA to stop the reaction. The level of free phosphate was measured on a plate reader after a 12 minute incubation with 20 μl malachite green. WT Hel308 is plotted in black; the truncated protein K646STOP is shown in red; the Walker A box mutant K52A in green; the DNA binding mutant R320A in blue. WT Hel308 minus DNA is represented by open black circles. B) The effect of temperature on the ATPase activity of WT Hel308 was investigated by incubating 0.5 μM WT Hel308 with 10 nM ssDNA (PhiX 174 virion). The reactions were started by the addition of 1 mM ATP-$MgCl_2$, samples were taken after 1 minute and added to 0.3 M PCA, as described. The average of triplicate measurements was plotted and the standard errors are shown in each case.

As expected the Walker A box mutant K52A was unable to hydrolyse ATP and was a useful negative control. R320A was also unable to catalyse ATP hydrolysis. Since the ATPase activity of WT Hel308 was DNA dependent, this result was consistent with the reduced ability of R320A to bind DNA (figure 5.7A). K646STOP and WT Hel308 both bound to ssDNA, similarly the ATPase activity exhibited by K646STOP was comparable to, if not faster, than WT Hel308. The reaction rate for WT Hel308 was 2.6 ± 0.15

pmoles phosphate.sec-^{1}.pmole enzyme^{-1} compared to a rate of 3.3 ± 0.15 pmoles phosphate.sec-^{1}.pmole enzyme^{-1} for K646STOP.

As expected for proteins from hyperthermophilic organisms, the ATPase activity of WT Hel308 increased as the temperature was raised (figure 5.8B). Consistent with the data for the Walker A box mutant the catalytic activity could, therefore, all be attributed to Hel308.

5.7 Hel308 displaces streptavidin from a biotinylated oligonucleotide

Translocation of Hel308 along ssDNA was investigated by observing its ability to displace streptavidin bound to a biotinylated oligonucleotide. Wild type Hel308 and the K646STOP, K52A and R320A mutants were incubated with 1 mM ATP-$MgCl_2$ and 10 nM 3' or 5' biotinylated 50 nt ssDNA molecule (J150B) bound to streptavidin at 45 °C. Samples were taken across a time course and added to a STOP solution. The STOP solution was optimized, compared to that described by Morris and Raney (1999), to eliminate band shifting which masked the streptavidin displacement (materials and methods). The results are presented in figure 5.9.

Hel308 was able to displace the streptavidin bound to the 5' or the 3' biotin (figure 5.9). WT Hel308 was able to displace more than 80 % of the streptavidin bound to a biotin probe in 10 minutes (figure 5.9 B). As expected the K52A Walker A box mutant was unable to displace the streptavidin, which is an ATP dependent reaction. Less than 40 % streptavidin displacement was seen with the R320A mutant during the time course, significantly lower than that seen for WT Hel308. This result was consistent with the weak DNA binding and ATPase activity observed with the R320A mutant. The K646STOP truncation mutant, lacking domain 5, retained the ability to displace streptavidin from the 5' biotinylated DNA, exhibiting slightly faster activity than that observed with WT Hel308 (figure 5.9 A, B). Translocation activity was slightly reduced when K646STOP was incubated with the streptavidin bound to the 3' biotinylated oligonucleotide (figure 5.9 B, C).

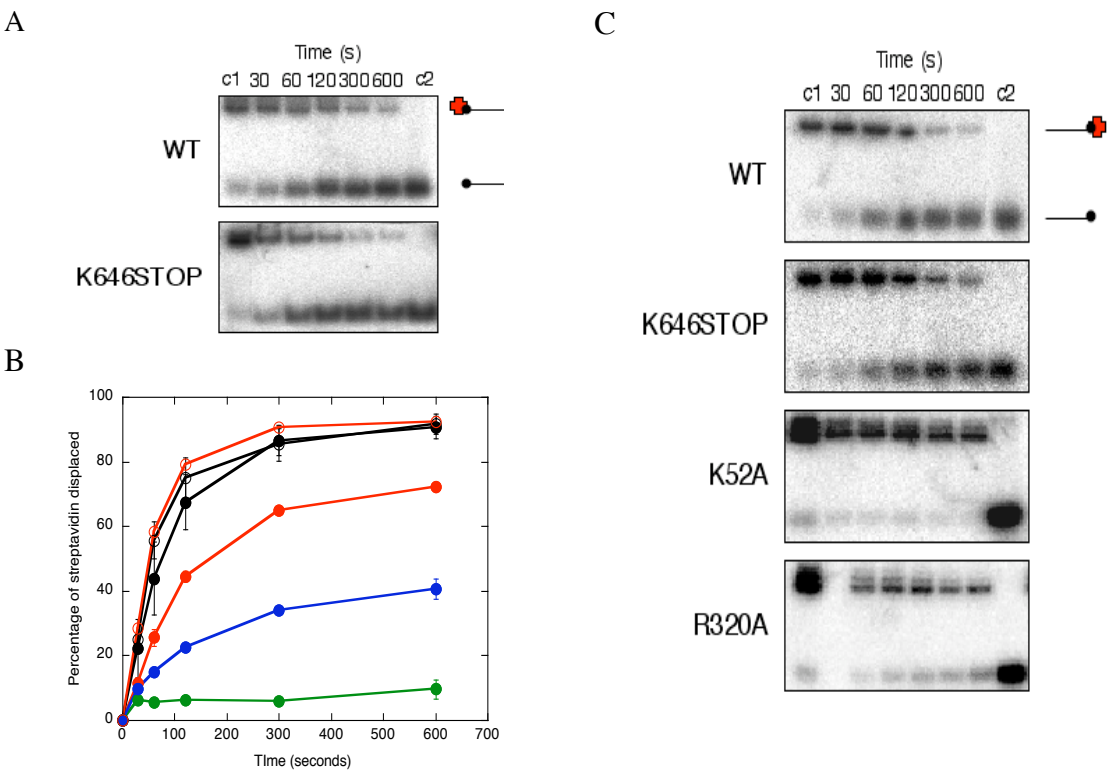


Figure 5.9. Translocation activity of WT Hel308 and mutants
10 nM biotin labelled oligonucleotide (^{32}P-5'-radiolabelled), bound to streptavidin, was incubated with 1 mM ATP-$MgCl_2$ at 45 °C and the reaction was started by addition of 0.5 μM mutant or WT Hel308 (final volume 60 μl). Samples were taken at the time points indicated and added to a STOP solution. The samples were analysed on a 12 % acrylmide: TBE gel. A) Gel pictures of WT Hel308 and K646STOP incubated with a 5' biotinylated probe and B) quantification of translocation activity exhibited by WT and mutant Hel308: the gel bands were quantified in Image Guage and the average of 3 data sets was plotted, the standard errors are shown. WT is plotted in black; K646STOP in red; K52A in green; R320A in blue; closed circles represent the 3' biotinylated DNA and open circles show the 5' biotinylated DNA. C) Gel picture of WT Hel308 and mutants incubated with a 3' biotinylated probe. c1 and c2 represent the controls without protein.

Hel308 was able to translocate along ssDNA and disrupt the tight interaction between streptavidin and biotin. The ability of helicases to remove proteins from DNA may be a common feature. This has important biological implications since most DNA sequences are bound, to some degree, by nucleic acid binding proteins and they must be removed for any of the DNA processing pathways, including replication and repair (Morris and Raney, 1999; Byrd and Raney, 2004). Streptavidin displacement may mimic the removal of protein, for example, from the DNA by chromatin remodelling protein *in vivo* and, therefore, these results implicate Hel308 in this role (Byrd and Raney, 2004).

5.8 Helicase activity of WT Hel308 and mutants

Human Hel308 is the orthologue of the *D. melanogaster* Mus308 protein, its identification was based on the similarities between the helicase domains of each (Harris *et al.*, 1996; Marini and Wood, 2002; McCaffrey *et al.*, 2006). The Mus308 helicases possess the seven motifs typical of SF2 DNA and RNA helicases, although they possess a combination of motifs from the various families.

5.8.1 Hel308 K646STOP unwinds DNA faster than wild type

The helicase activity of Hel308 was initially assayed using a minimal substrate; a 3' overhang consisting of a 25 bp duplex and a 25 nt ssDNA region (3' overhang - J150X and R26-50 see appendix 2 for nucleotide sequence). The DNA unwinding abililty of Hel308 was investigated at 60 ˚C over a 3 minute time course (figure 5.10A). The overhang substrate was incubated at 60 ˚C for 4 minutes to ensure the substrate was stable and that any single stranded product was produced due to catalytic activity and not substrate melting. Boiled substrate was run on the gel to indicate the size of the unwound product.

The WT Hel308 protein was able to displace the DNA strand efficiently, as observed for Hel308 homologues (Marini and Wood, 2002; Guy and Bolt, 2005; Fujikane *et al.*, 2006); WT Hel308 unwound approximately 85% of the DNA substrate (figure 5.10A, B). The Walker A box mutant, K52A, was unable to function as a helicase and the R255A and R320A mutants showed a dramatically reduced helicase activity compared to WT; only 20-25% of the DNA substrate was unwound by either mutant. This observation was consistent with significantly impaired DNA binding capabilities identified by anisotropy.

Surprisingly, both the truncation mutant K646STOP (lacking domain 5) and the R662A mutant, lacking the first arginine of the RAR motif in domain 5, showed significantly faster rates of unwinding than WT Hel308. K646STOP, especially, unwound 100% of the DNA substrate within 30 seconds at 60 °C. This data implicates domain 5 of Hel308 in an autoinhibitory role, since its presence seems to impose a significant level of control (Pufall and Graves, 2002).

Consistent with the streptavidin displacement data and in contrast to previous helicase data published on eukaryotic and archaeal Hel308 (Marini and Wood, 2002; Guy

and Bolt, 2005; Fujikane *et al*., 2006), WT Hel308 was able to efficiently unwind a 5' overhang DNA substrate (J150X and B1-25). This reaction, however, was not quite comparable to the level of unwinding evident in the presence of the 3' overhang (~ 55 % compared to 85 % in 3 minutes).

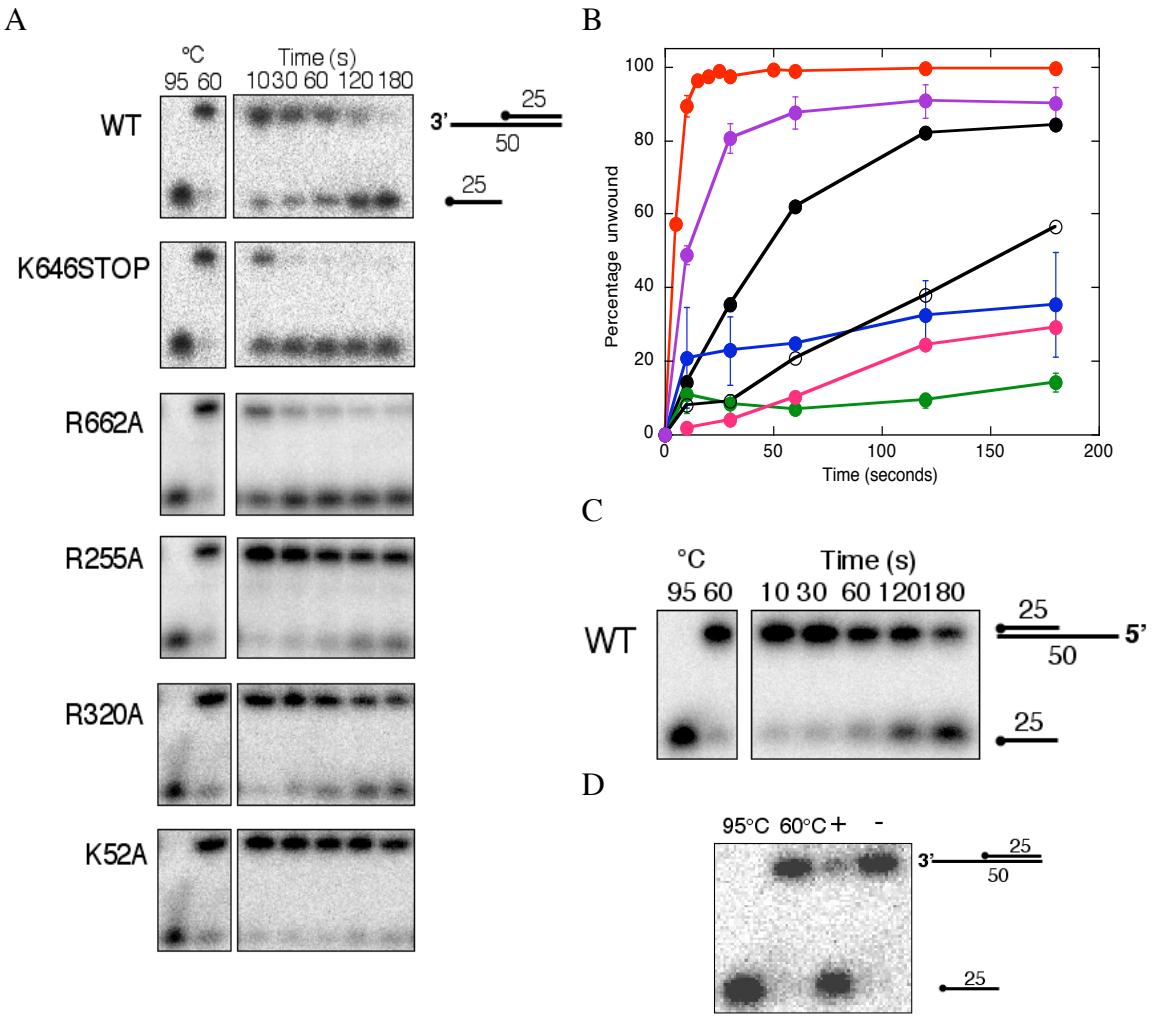


Figure 5.10. Helicase activity of WT Hel308 and mutants on a minimal substrate
A) 10 nM ^{32}P labelled (black circles) 3'overhang substrate was incubated with 0.5 μM protein at 60 °C for 1 minute in a final volume of 60 μl. The samples were analysed on 12 % acrylmide: TBE gels. B) the gel bands were quantified in Image Guage and the average of three data sets was plotted, the standard errors are shown. WT is plotted in black; K646STOP in red; R320A in blue; K52A in green; R255A in pink; R662A in purple. Closed circles represent the 3' overhang substrate and open circles respresent WT incubated with a 5' overhang substrate. C) WT Hel308 was also incubated with a 5' overhang substrate as described in A. D) The helicase activity of 2 μM WT Hel308 was investigated in the presence and absence of 1mM ATP-$MgCl_2$, the reactions were incubated for 3 minutes.
In each experiment the substrate was boiled as an indicator of the size of the unwound product and the substrate was incubated at 60 °C to ensure that the substrate was not melting. These are shown on the gels by 95 °C and 60 °C respectively.

Without testing the helicase activity of Sso Hel308 on a blunt ended duplex, the results presented here do not confirm a bidirectional helicase activity. Other possible explanations of the activity exhibited by Hel308 on the 5' overhang include 3'-5' helicase activity from a blunt DNA end and or from single stranded DNA exposed as a result of thermal fraying at 60 °C. The helicase activity of Hel308 was ATP dependent (figure 5.10C) since the unwinding ability was abolished in the absence of ATP-$MgCl_2$.

5.8.2 Hel308 cannot unwind long duplex regions efficiently

Human Hel308 shows a limited ability to displace longer duplex strands (Marini and Wood, 2002). Therefore, to follow up on this observation and to slow the rate of strand separation as catalysed by K646STOP, the helicase activity of both WT Hel308 and K646STOP was studied on a 3'overhang substrate with a longer duplex region of 50 bp (J150B with and J150B complement with a 25 T 3' tail) (figure 5.11). The control lane (60 ˚C) shows that the substrate was stable at the assay temperature therefore any unwound product was the result of the helicase activity of the enzyme. The boiled substrate (95 ˚C) indicates the size of the unwound product.

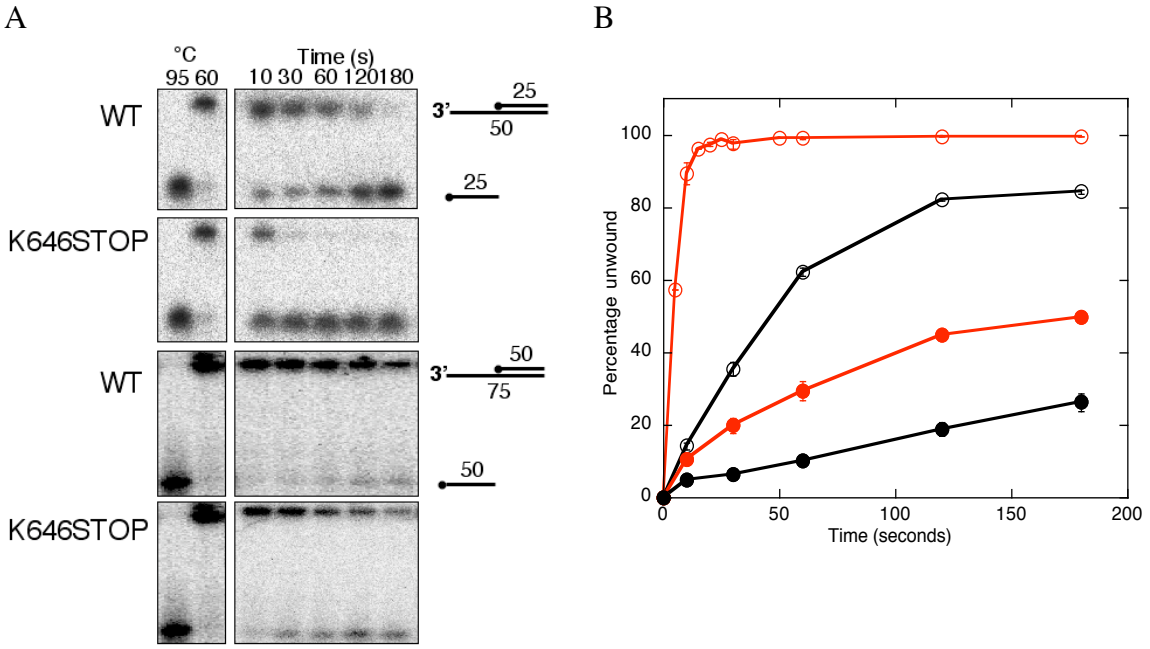


Figure 5.11. Helicase activity of WT hel308 and mutants on longer duplex strands
10 nM ^{32}P labelled (black circles) 3'overhang substrate was incubated with 0.5 μM protein at 60 °C for 1 minute in a final volume of 60μl. A) The samples were analysed on 12 % acrylmide: TBE gels. In each case the substrate was boiled to indicate the size of the unwound product (95 °C) and incubated at 60 °C to ensure the substrate was stable at the assay temperature (60 °C). B) the gel bands were quantified in Image Guage, the average of triplicate measurements was plotted and the standard errors are shown. WT Hel308 is represented in black; K646STOP in red; closed circles represent the long 3'overhang; and the open circles represent the shorter 3' overhang.

Consistent with human Hel308, both WT Hel308 and the truncation mutant K646STOP (lacking domain 5) exhibited a significantly weaker helicase activity when assayed with the longer substrate; almost 50% reduction in activity was evident. However, in both cases, significantly better unwinding was seen with the K646STOP protein.

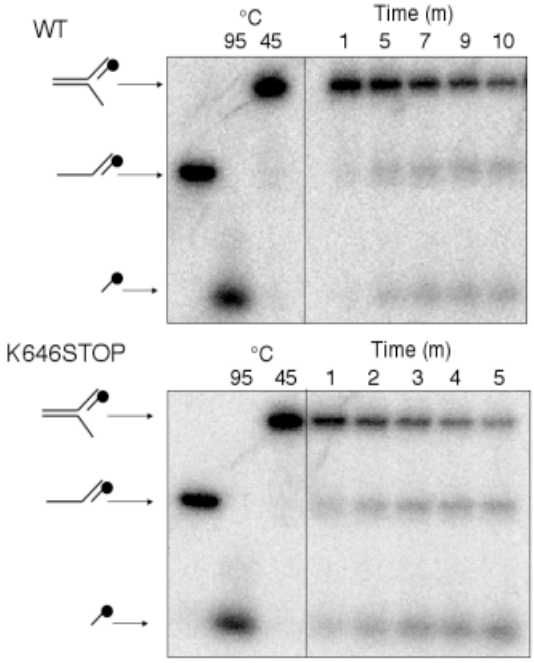
5.8.3 Hel308 requires a ssDNA portion to initiate efficient helicase activity

Previous biochemical studies on human Hel308 and the homologues from *P. furiosus* and *M. thermoautotrophicus* have shown that Hel308 is able to displace a DNA strand from a duplex structure that contains either a nick or a ssDNA region with 3'-5' polarity (Marini and Wood, 2002; Guy and Bolt, 2005; Fujikane *et al.*, 2006). Fork structures resembling stalled DNA replication forks were identified as target substrates; the strand corresponding to a nascent lagging strand was most efficiently removed by Hel308 homologues from *M. thermoautotrophicus* and *P. furiosus* (Guy and Bolt, 2005; Fujikane *et al.*, 2006).

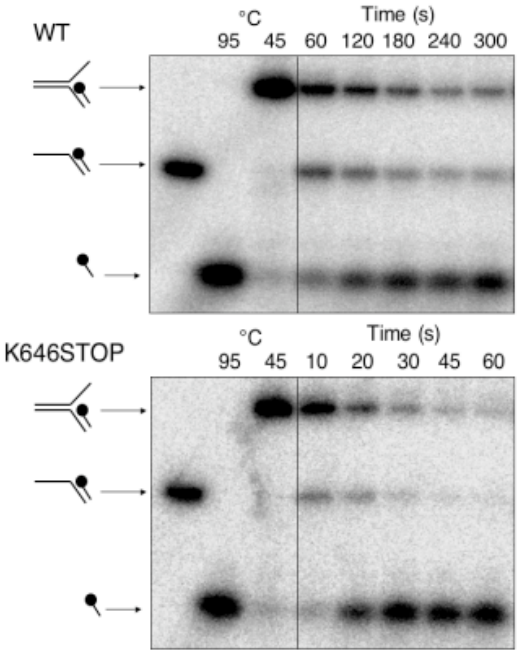
The substrate specificities of WT Hel308 and K646STOP were investigated using a number of forked substrates (figure 5.12), including a nicked 4-way junction (J150B, J150X, J150H, R1-25 and R26-50); a nicked 3-way junction (J150B, J150X, H1-25 and R26-50); a stalled fork with a leading strand (J150B, J150X and H1-25) and a stalled fork with a lagging strand (J150B, J150X and R26-50). The oligonucleotide sequences are presented in appendix 2. In each case K646STOP exhibited significantly faster helicase activity than WT Hel308. As described previously, the substrate was incubated at 45 ˚C for the same period of time as the assay. In each case the substrate was stable at this temperature. The boiled substrate was run on the gel to indicate the size of the unwound product. Substrate instability was obvious at 60 ˚C and in the helicase assay conditions described previously (materials and methods). Addition of 300 mM NaCl to the reaction mixture prevented the DNA substrates essentially falling apart. This, however, also inhibited the helicase activity of Hel308. Since the substrates were stable during the reaction time but were dissociating during the incubation period in STOP solution, NaCl was added to the STOP solution to a final concentration of 300 mM (added to the STOP solution last to prevent precipitation of SDS).

A range of possible intermediate substrates were also run on the gel and used as markers to identify intermediate products.

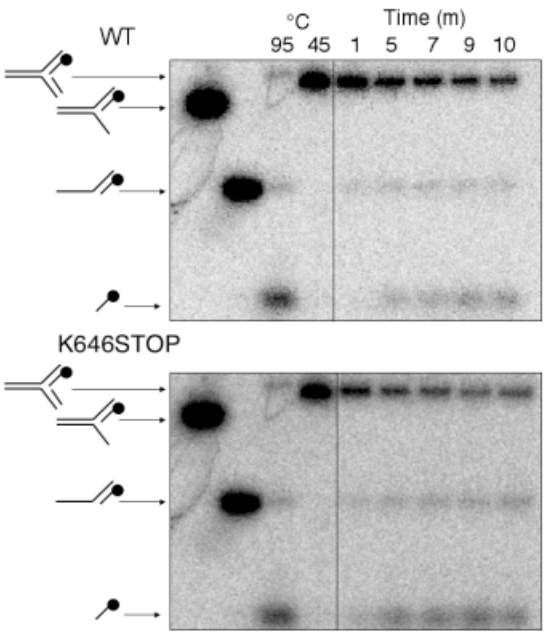
A



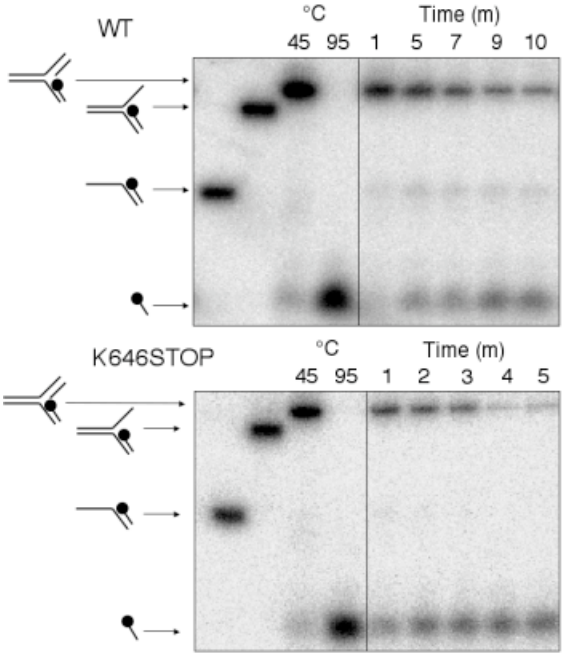
B



C



D



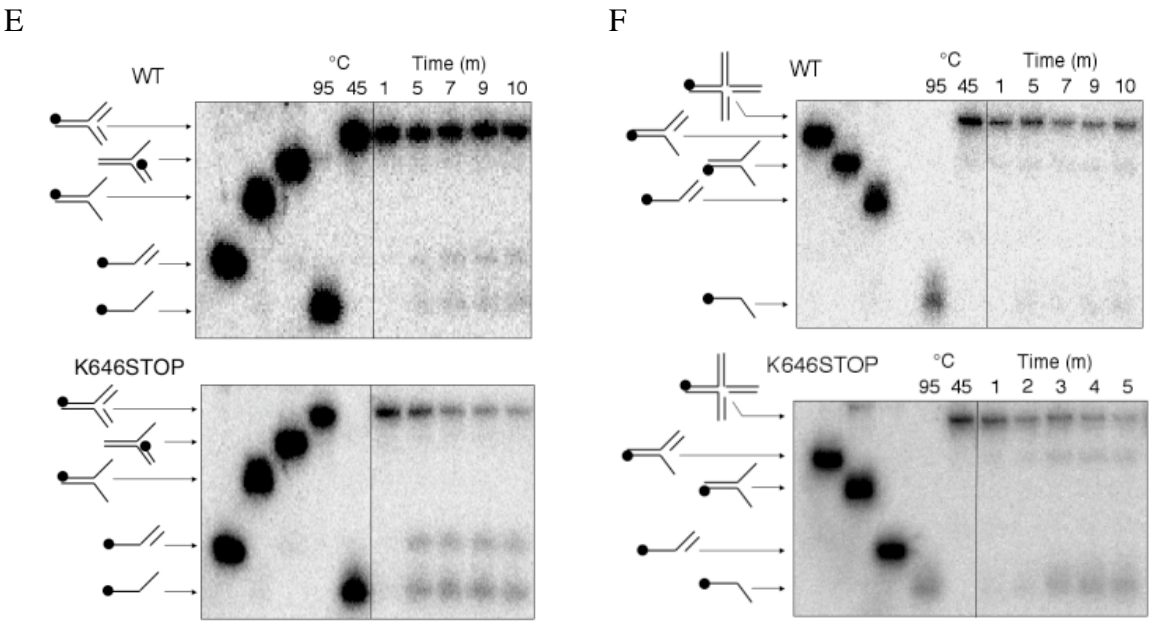


Figure 5.12. Helicase activity of WT Hel308 and K646STOP on fork substrates
The substrate specificity of Hel308 was investigated by incubating WT protein and K646STOP with a range of DNA fork and junction substrates. 10 nM ^{32}P 5' labelled fork substrate was incubated with 0.5 μM protein at 45 °C (final volume 60 μl) for 1 minute and the reaction was started by addition of 1 mM ATP-$MgCl_2$. Samples were taken at the time points indicated and added to STOP solution to inactivate the enzyme and prevent substrate reannealing. The samples were analysed on 12 % acrylmide: TBE gels. A) stalled replication fork with a leading strand; B) stalled replication fork with a lagging strand; C-E) nicked 3-way junctions, each labelled on a different strand as shown; and F) nicked 4-way junction. The black circle represents the ^{32}P label. The substrates were incubated at the assay temperature to show that any unwinding was due to catalysis and not substrate melting. The substrate was boiled and run on the gel with various possible structural intermediates as size markers.

The stalled replication fork structure with the nascent lagging strand (figure 5.12B) proved to be the target substrate for Hel308, this is consistent with previous studies (Guy and Bolt, 2005; Fujikane *et al.*, 2006). However, SsoHel308 looked to unwind the parental duplex before the lagging strand. Once again the truncation mutant K646STOP proceeded to unwind the lagging strand at a much faster rate compared to WT Hel308; K646STOP unwound ~ 95% of the lagging strand in 1 minute compared to ~ 80% of the substrate unwound by WT Hel308 in 5 minutes.

The equivalent stalled replication fork with a nascent leading strand (figure 5.12A) did not stimulate efficient unwinding, only ~ 60 % of the leading strand was removed by K646STOP in 5 minutes.

The helicase assays were repeated with a nicked 3-way junction (figure 5.12 C, D, E). The same junction was prepared with the ^{32}P label on the 5'-3' strand of the parental

duplex, the leading strand or the lagging strand. Neither WT Hel308 nor K646STOP were able to efficiently unwind the lagging strand from this nicked 3-way junction, compared to the fork with the lagging strand, consistent with the limited ability of Hel308 to bind dsDNA. There was, therefore, a requirement for a ssDNA portion for efficient helicase activity. This is not in keeping with *M. thermoautotrophicus* Hel308, which possesses the ability to remove the lagging strand from a substrate with a ssDNA region or a nick in the DNA backbone (GUY, 2005). Although the level of helicase activity was very low, the gel pictures in figures 5.12 C, D and E indicated that Hel308 unwound the parental duplex before targeting the resulting overhang substrate. In the case of the junction with the labelled lagging strand, the resulting labelled substrate is a 3' overhang, a known substrate of Hel308, which could account for the higher level of unwinding seen in figure 5.12 D compared to 5.12 C and E.

No helicase activity of WT Hel308 was detected with a nicked 4-way junction in 10 minutes. K646STOP, however, exhibited limited unwinding across a 5 minute time course. The initial unwinding event resulted in the formation of a fork substrate with a leading or lagging strand. The latter was more likely since this was then further unwound to ssDNA (figure 5.12F).

With each of the substrates assayed, Hel308 unwound the parental duplex before the strand corresponding to the lagging strand. This was in contrast to the activity exhibited by *M. thermoautotrophicus* Hel308, which exclusively unwound the lagging strand of this structure (Guy and Bolt, 2005) A nicked duplex and a lagging strand gapped 3-way junction substrate would be good controls to investigate this substrate-intermediate-product unwinding event further.

5.8.4 Hel308 does not require the parental duplex for activity

Hel308 unwound both the overhang and the stalled replication fork substrates. In order to compare the two, the gel pictures were quantified using Image Guage and the percentage of unwound substrate was plotted (figure 5.13A).

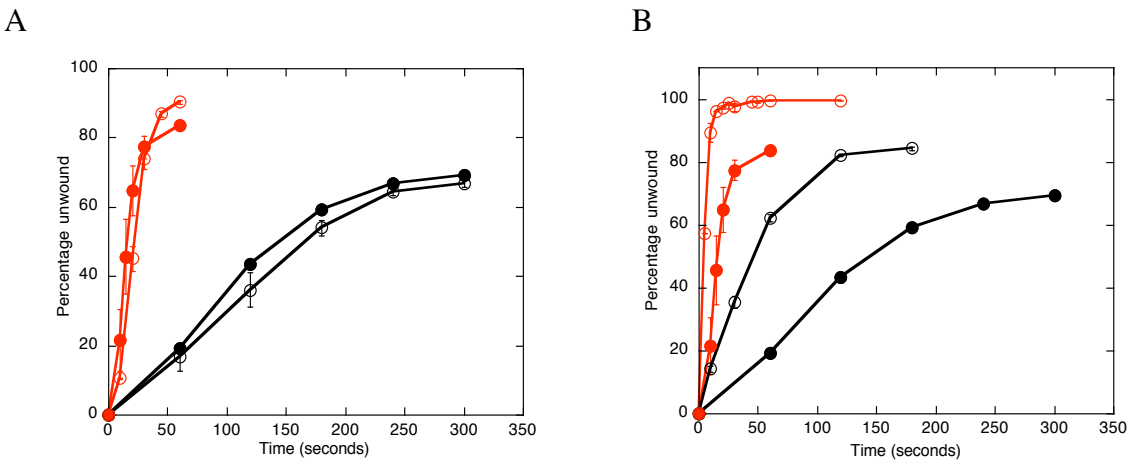


Figure 5.13. Quantified helicase assay data comparisons
A) The comparison of helicase activity at 45 ˚C on a 3' overhang structure (closed shapes) and a stalled replication fork with a lagging strand (open shapes); B) The comparison of helicase activity on a 3' overhang at 45 ˚C (closed circles) and 60 ˚C (open circles). WT is shown in black and K646STOP in red. Assays were carried out as described in the experimental procedures, data was quantified using Image Guage and plotted as an average of triplicate experiments and the standard errors are shown.

WT Hel308 activity did not seem to change regardless of the substrate, whereas the helicase activity of Hel308 K646STOP was slightly faster when stimulated by the stalled fork. In contrast *M. thermoautotrophicus* Hel308 unwound a nicked duplex substrate less efficiently than a nicked 3-way junction, exhibiting a requirement for the parental duplex (Guy and Bolt, 2005). Therefore, although Hel308 unwinds the parental duplex of the DNA substrate, there is no obvious requirement for it. The data, therefore, does not confirm that Sso Hel308 is a DNA fork specific helicase.

The helicase activity was investigated at 45 °C and at 60 °C (figure 5.13B) and, as expected, the activities of both WT Hel308 and K646STOP at the higher temperature were much faster. Figure 5.13B emphasises the effect of domain 5 as even the helicase activity of WT Hel308 at 60 °C was slower than the activity of K646STOP at 45 °C.

5.9 Discussion

5.9.1 Mechanism of Hel308 helicase activity

The crystal structure of *S. solfataricus* PBL2025 Hel308 was fully consistent with the structure of *A. fulgidus* Hel308 in the presence and absence of a DNA substrate (Buttner *et al.*, 2007). Comparison of the two structures provides a great deal of

information concerning the mechanism by which Hel308 unwinds duplex DNA. Hel308 consists of 5 domains; the first four adopt a ring structure large enough to accommodate a single DNA strand but not duplex DNA (Buttner *et al.*, 2007). This ring is not covalently closed promoting a degree of flexibility that would allow opening of the gap between domains 2 and 4 and loading of Hel308 onto the DNA lacking a 3' end, such as a stalled replication fork (Richards *et al.*, 2007). Domain 5 sits perpendicular to the plane of the ring formed by domains 1-4 and has been proposed to confer substrate specificity (Buttner *et al.*, 2007).

A β-hairpin loop extruding from domain 2 sits at the ss/dsDNA junction and is involved in melting 2 bp of the duplex by disrupting the base stacking in both strands without the requirement for ATP. Structural data of Hel308 revealed that residues of this loop wedge between the last paired bases and the first unpaired bases of the DNA strands (Buttner *et al.*, 2007). One of the single DNA strands is then drawn across domains 1 and 2, powered by ATP binding and hydrolysis induced conformational changes in the motor domains. The presence of two binding sites suggests that Hel308 engages in unwinding according to the inchworm mechanism (Lohman and Bjornson, 1996; Buttner *et al.*, 2007). The second DNA strand is passed through the ring, as a result of interactions with conserved arginine residues that line the central pore, and is bound by domain 5 upon exiting the channel. The substrate duplex and the single strand product adopt an almost collinear conformation. This is in contrast to the mechanism employed by SF1 helicases PcrA and UvrD where the ssDNA product is bound at right angles to the duplex, the strands are effectively peeled apart by concerted ssDNA translocation and duplex rotation (Velankar *et al.*, 1999; Lee and Yang, 2006; Buttner *et al.*, 2007).

Büttner *et al.* (2007) proposed the following mechanism of unwinding (figure 5.14). Motif IV on domain 2 is responsible for pushing the DNA towards domain 1, upon ATP binding. This movement would promote contact with the ratchet helix of domain 4, which stacks with the DNA bases in the 3' tail. This pushing action of domain 2 and the subsequent movement of domain 4 would loosen the association of DNA with domain 1 to promote backbone slipping of the 3' tail across motif Ia. Domain 2 is able to rebind the upstream DNA upon ATP hydrolysis and the β-hairpin loop is positioned to melt the next base pair (Buttner *et al.*, 2007).

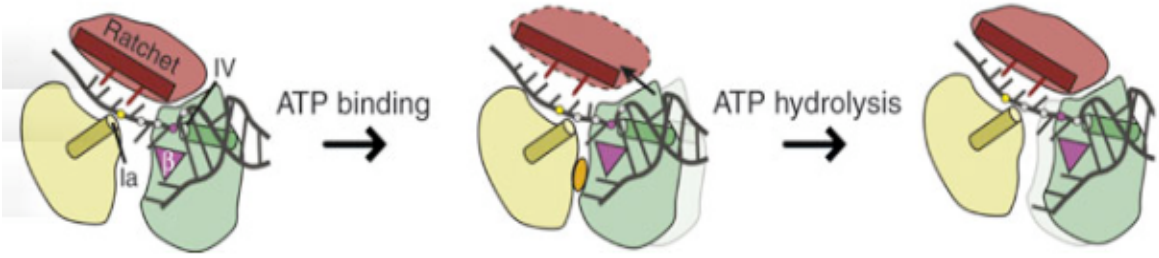


Figure 5.14. A cartoon of the two-step model for processive Hel308 DNA unwinding
Upon ATP binding, motif IV on domain 2 pushes the DNA towards domain 1 and via direct contacts with the ratchet helicase, it also pushes on domain 4 that stacks with the bases on the 3' tail. This combined movement of domains 2 and 4 loosens the grip of domain 1 on the DNA, slipping the backbone of the 3' tail across motif Ia. ATP hydrolysis promotes domain 2 rebinding to upstream DNA, this positions the β-hairpin loop in order to melt the next base pair of the duplex. Domains 1, 2 and 4 are coloured yellow, green and red respectively. Motif Ia is represented by the yellow helix, motif IV by the green helix, the ratchet helix is shown in red and the magenta triangle represents the β-hairpin loop. Domains 3 and 5 have been omitted for clarity (Buttner *et al*., 2007).

5.9.2 Autoinhibitory control of Hel308 helicase activity

Hel308 is a ssDNA binding protein that hydrolyses ATP in a DNA dependent manner, displaces streptavidin from a biotinylated oligonucleotide and is active as a helicase. The reduced catalytic activities of the R255A and R320A mutants were consistent with an obvious reduction in DNA binding ability, as predicted from the crystal structure. The Walker A box mutant K52A bound ssDNA, but was unable to hydrolyse ATP and therefore any translocation or helicase activity was obliterated, as expected.

Büttner *et al*. (2007), however, suggested that domain 5 acted as a specificity domain targeting Hel308 to branched DNA substrates. In contrast to this, the data in this chapter clearly shows that removal of domain 5 (K646STOP), or mutation of the conserved RAR motif (R662A), did not impair DNA binding or catalytic activity of Hel308. Moreover, it resulted in an increase in the activity of the helicase. K646STOP (truncation mutant), in particular, retained a faster helicase activity than the wild type protein and was able to displace the single strand from a longer duplex region more effectively. This same effect was evident with all of the substrates used (overhangs, forks and junctions), regardless of Hel308 substrate preference.

Fujikane *et al*. (2006) presented evidence showing that PCNA stimulated the activity of *P. furiosus* Hel308 by direct interactions via a PIP (PCNA Interacting Protein) box in the C-terminal sequence of Hel308 (figure 5.15).

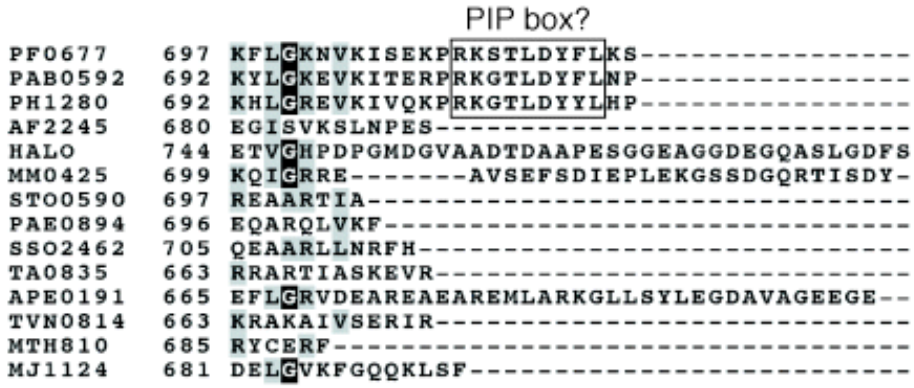
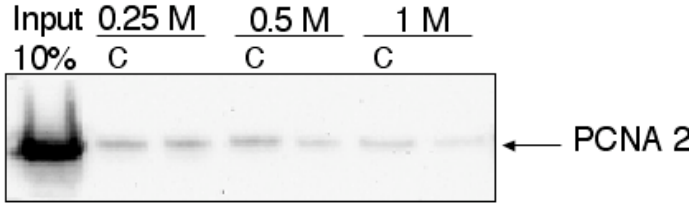


Figure 5.15. A multiple sequence alignment of the C-terminal region of Hel308
The alignment compares the C-terminal portion of the Hel308 homologues from *Pyrococcus furiosus* (PF), *Pyrococcus abyssi* (PAB), *Pyrococcus horikoshii* (PH), *Archaeoglobus fulgidus* (AF), *Halobacterium* Sp. NRC-1 (HALO), *Methanosarcina mazei* (MM), *Sulfolobus tokodaii* (STO), *Pyrobaculum aerophilum* (PAE), *Sulfolobus solfataricus* (SSO), *Thermoplasma acdiophilum* (TA), *Aeropyrum pernix* (APE), *Thermoplasma volcanium* (TVN), *Methanothermobacter thermoautotrophicus* (MTH), and *Methanocaldococcus jannaschii* (MJ). The potential PIP box is shown inside the rectangle (Fujikane *et al*., 2006).

Hel308 did not bind PCNA (figure 5.16A) and, accordingly, PCNA had a slight negative affect on the helicase activity of Hel308 (figure 5.16B). This was likely due to PCNA bound to the DNA substrate and hindering progression of Hel308 rather than the result of a functional interaction, consistent with the lack of the C-terminal PIP box identified in a group of archaea, including *P. furiosus* (figure 5.15).

A



B

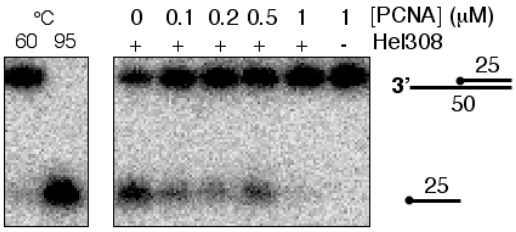


Figure 5.16. Hel308 and PCNA do not interact
A) Hel308 was covalently immobilised on an Affi-gel column and incubated with heterotrimeric PCNA. The column was washed with 0.25 M, 0.5 M and 1 M NaCl buffer, the elution samples were precipitated according to the TCA-DOC protocol and analysed by western blot. The blots were probed with a primary antibody against PCNA2. A control column, shown as c, was prepared without Hel308 to ensure PCNA was not binding to the Affi-gel.
B) The helicase assay was carried out with 10 nM 3' overhang substrate (^{32}P-5'-radiolabel represented by a black circle) at 60 °C. 0.2 μM Hel308 was incubated with increasing concentrations of heterotrimeric PCNA, as indicated, for 1 minute in a total volume of 15 μl. PCNA was incubated with the DNA substrate in the absence of Hel308 to ensure it did not unwind DNA itself. The substrate was incubated at assay temperature and at 95 °C to ensure substrate stability and to show substrate melting.

The evidence suggests that domain 5 controls the helicase activity of Hel308 by acting as a molecular brake. Autoinhibitory domains negatively regulate the activity of a second domain, acting as a kinetic block. This is an effective method of achieving tight regulation and a high level of specificity (Wang *et al.*, 2006). Autoinhibition has been observed in a number of proteins, for example, domain 4 of bacterial DNA repair helicase UvrB and subdomain 2B of *E. coli* replication fork restart protein Rep have both been identified as autoinhibitory domains (Brendza *et al.*, 2005; Wang *et al.*, 2006). Inhibition of Rep activity by subdomain 2B is relieved by dimerisation, therefore the relative position of 2B controls activation of Rep protein (Brendza *et al.*, 2005). Similarly domain 4 inhibition of UvrB is relieved by interactions with UvrA, this allows regulation of the UvrABC repair activity to prevent unwanted incision events (Wang *et al.*, 2006). Bacterial transcription coupling factor Mfd is another example of a protein that is controlled by an autoinhibition mechanism. C-terminal domain 7 of Mfd has been shown to prevent oligonucleotide displacement *in vitro*, a block that is relieved by the binding of RNA polymerase (Smith *et al.*, 2007). This interaction is thought to result in a conformational change in domain 7 which promotes the interaction of Mfd with UvrA repair protein additional to alleviating the repression (Smith *et al.*, 2007).

In each of these cases, enzyme inhibition is the default state and only by formation of protein-protein interactions is the autoinhibition lifted. Further research is required to determine whether the autoinhibitory domain of Hel308 has a similar role, or whether it functions as a constitutive brake on the helicase activity.

5.9.3 Model for the action of Hel308 *in vivo*

Hel308 has been shown here to exhibit 3'-5' and 5'-3' translocase and helicase activity. It is possible, therefore, that *in vivo* this could be a property that is governed by the strand or the orientation by which Hel308 is loaded onto the DNA (figure 5.17). Positioning of Hel308 on the lagging strand template would be appropriate for unwinding the nascent lagging strand (3'-5) or unwinding the parental duplex (5'-3'). Alternatively, if Hel308 loaded onto the leading strand template the nascent leading strand (5'-3') would be displaced or, once again, the parental duplex (3'-5').

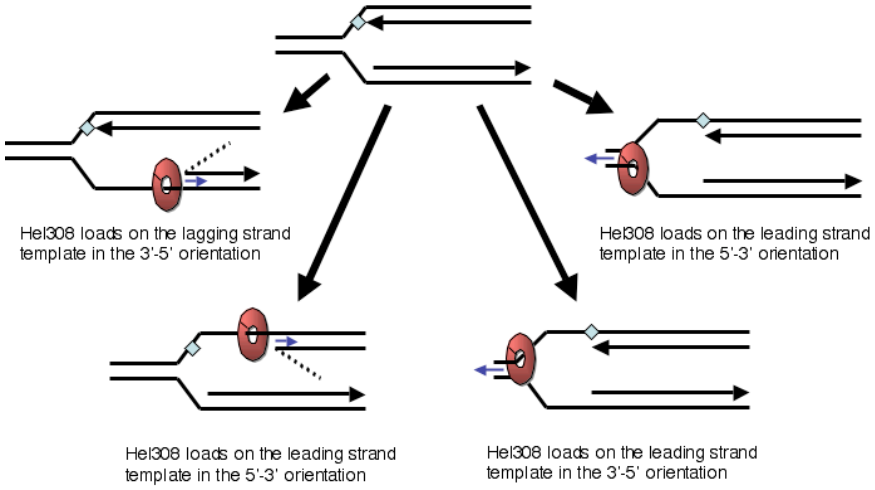


Figure 5.17. A cartoon of the potential orientations of Hel308 on the DNA
Hel308 can catalyse 3'-5' and 5'-3' translocation and duplex unwinding. Therefore, the strand and orientation by which Hel308 is loaded onto the DNA may be vital for the appropriate function. Hel308 could load onto the lagging strand (3'-5') or leading strand (5'-3') template and displace the nascent lagging strand or leading strands, respectively. Alternatively, Hel308 could load onto the strands in the opposite orientations and, in both cases, destabilise the parental duplex.

Since the truncation mutant K646STOP was able to unwind all of the substrates to a certain extent, domain 5 could have a vital role in the correct positioning and orientation of Hel308 to ensure the correct strand is unwound. There may, also, be a requirement for interacting proteins to ensure correct loading of Hel308 and to alleviate the block imposed by domain 5. Alternatively, there may be an obligation for Hel308 to

partially unwind the parental duplex before removing the lagging strand, although this requires further investigation in order to fully understand the nature of this pattern of unwinding.

Hel308 exhibited the most efficient helicase activity, *in vitro*, when stimulated by a stalled replication fork structure with a corresponding lagging strand or a 3' overhang structure. The schematic in figure 5.18 is a model proposed based on the results presented in this chapter and those from previous biochemical analysis (Guy and Bolt, 2005; Fujikane *et al*., 2006).

A lesion on the leading strand template causes the replication fork to stall, resulting in a gap since the leading strand cannot be primed downstream of the start site (Gregg *et al*., 2002). Similar to bacterial RecBCD and PriA at a replicative helicase block, it is thought that Hel308 is targeted to the stalled fork. The DNA displacement model describes the process by which Hel308 displaces the nascent lagging strand clearing the template for ssDNA binding proteins, for example RadA.

The *E. coli* RecFOR complex is known to remove ssDNA binding protein SSB from the nascent strand template and promote binding of RecA, which forms a nucleofilament on the DNA capable of initiating strand exchange (Umezu *et al*., 1993; Morimatsu and Kowalczykowski, 2003). As described here in the protein displacement mode (figure 5.18) an additional function of Hel308 may be to remove proteins from the ssDNA, since Hel308 was able to displace streptavidin from biotinylated ssDNA. Analogous to the proposed RecFOR pathway where SSB is displaced to allow RecA to bind, Hel308 could dislodge archaeal SSB (RPA in eukaryotes) from the DNA clearing the strand for binding of RadA (Rad51 in eukaryotes). This could form a nucleofilament, an active strand exchange species. The collinear relationship of the duplex substrate and the ssDNA product lends itself to many functions, including stripping proteins from DNA, since this mechanism is very versatile (Buttner *et al*., 2007). The combined action of RecQ helicases with nucleases is a common feature of this protein family, for example in the RecFOR pathway RecQ works in concert with RecJ (Morimatsu and Kowalczykowski, 2003; Bennett and Keck, 2004). It may be interesting to identify whether Hel308 works in combination with nuclease in this way to process DNA ends.

Each of these models results in a RadA nucleofilament that can initiate recombination or promote reannealing of the parental duplex for NER (figure 5.18). An alternative pathway involves the 'chicken foot' intermediate. An additional proposal describes Hel308 catalysing the unwinding of the nascent leading and lagging strands, which would then anneal together. This structure could then be resolved causing the fork to collapse and replication could restart via a D-loop intermediate, analogous to the action of RecG at a blocked fork (McGlynn *et al.*, 2001; McGlynn and Lloyd, 2002; Krejci *et al.*, 2003).

Bacterial UvrD, another 3'-5' helicase, has been shown to disrupt RecA nucleoprotein filaments, limiting recombination (Veaute *et al.*, 2005). Rather than promoting strand exchange, Hel308 may function in this way to prevent it. RecQ proteins are also known to both promote and inhibit recombination (Bennett and Keck, 2004).

P. furiosus Hel308 was shown to interact directly with RadA, the archaeal RecA (Fujikane *et al.*, 2006). This is a result that compliments either of these suggestions for Hel308.

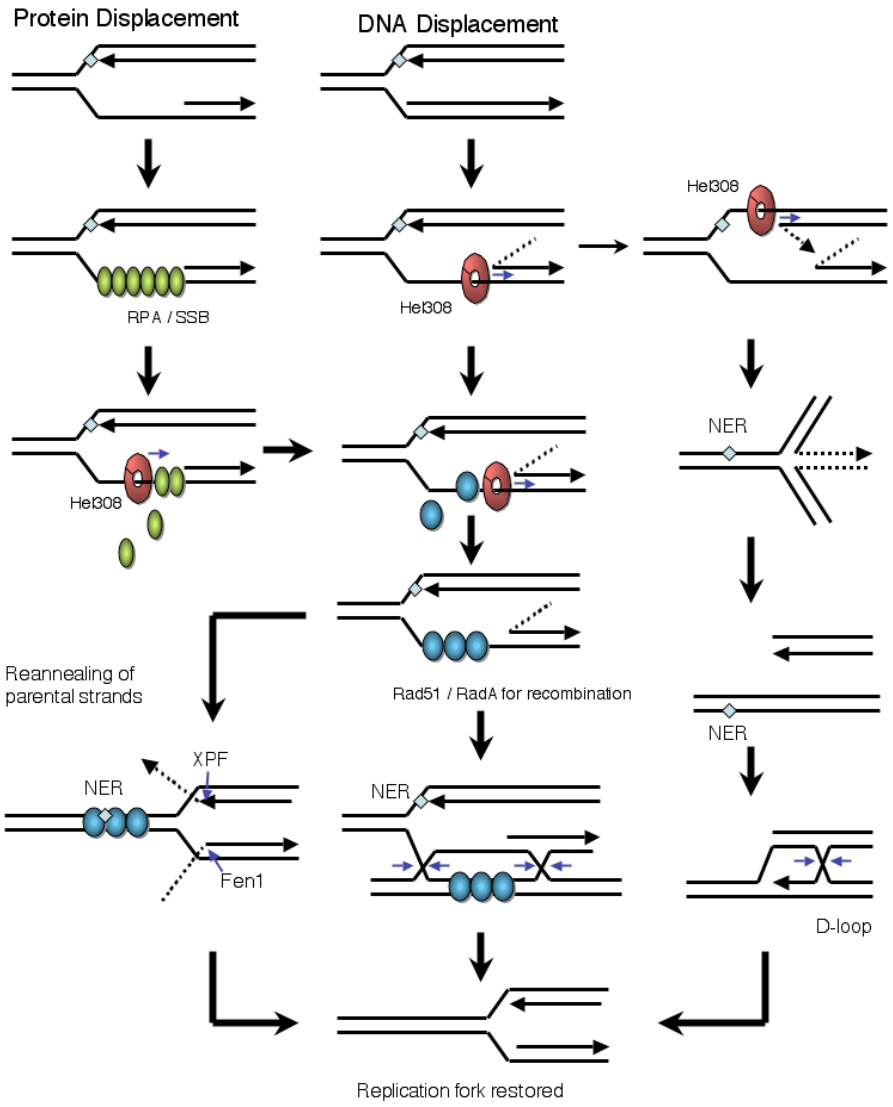


Figure 5.18. A model for the action of Hel308 at a stalled replication fork
A lesion on the leading strand template causes the replication fork to stall, which results in a gap since the leading strand cannot be primed downstream of the start site. It is thought that Hel308 (red ring) is targeted to the stalled fork, similar to bacterial RecBCD and PriA. The DNA displacement model describes the process by which Hel308 displaces the nascent lagging strand clearing the template for ssDNA binding proteins, for example RadA (euk Rad51, *E. coli* RecA). The protein displacement model suggests that Hel308 acts in a manner analogous to bacterial RecFOR and displaces the single stranded binding protein SSB (euk RPA), shown as a green sphere, clearing the strand for fork regression or RadA (blue sphere) loading. Both of these models lead to the production of a RadA nucleofilament active in strand exchange. Alternatively, RadA could promote reannealing of the parental duplex for NER. The replication fork is restored by either Holliday junction resolution or end processing of the nascent strands, respectively.
A further model implicates Hel308 in unwinding both the nascent leading and lagging strand to promote the formation of the 'chicken' foot intermediate. This is then resolved causing fork collapse, leading to homologous recombination. The replication fork is reset via a D-loop intermediate.

CHAPTER 6: A HYPOTHETICAL PROTEIN FROM *SULFOLOBUS SOLFATARICUS*

6.1 Introduction

Members of the AAA ATPase family are associated with diverse cellular activities in archaea, bacteria and eukaryotes (Karata *et al.*, 1999). They have functional roles in a large variety of processes; protein folding, membrane trafficking, intercellular motility and DNA replication are just a few examples (Vale, 2000; Ogura and Wilkinson, 2001). The AAA ATPases form a subfamily of the Walker-type ATPases, defined on the basis of an amino acid sequence referred to as the second region homology motif (SRH) (Karata *et al.*, 1999; Ogura and Wilkinson, 2001). The Walker A, Walker B and the SRH motifs are located within a 200-250 amino acid sequence that is termed the AAA module. The SRH motif, therefore, is the distinguishing feature between the AAA proteins and other ATPases (Karata *et al.*, 1999; Dougan *et al.*, 2002). Proteins contain either one or two copies of this module and are subsequently classified into Class II and Class I, respectively (Schirmer *et al.*, 1996; Karata *et al.*, 1999; Dougan *et al.*, 2002). It seems confusing, then, that this family exhibits such a diverse range of functions when there is such extreme conservation of the sequence within the AAA module.

Most of the AAA ATPase proteins take the form of hexameric rings, which are able to change their shape during the ATPase cycle (Vale, 2000). Vps4p, however, was the first example of an AAA-type ATPase for which a model linking ATP hydrolysis to the oligomeric state of the protein was proposed (Babst *et al.*, 1998). Vsp4p is a protein from *S. cerevisiae* important for the efficient transport of endocytic and biosynthetic material from the endosome to the lysosome-like vacuole. The oligomeric state of Vsp4p is determined by the ATPase activity, which drives the assembly and disassembly cycle of the Vsp4p dimers/oligomers (Babst *et al.*, 1998).

The AAA family and related ATPases are collectively designated AAA+ family members (Karata *et al.*, 1999). AAA+ proteins possess an extensive number of accessory domains and factors (Dougan *et al.*, 2002; Erzberger and Berger, 2006). They function by linking conformational changes during oligomerisation, mediated by nucleotide binding, to specific chemo-mechanical movements. These motions are then transduced to the target macromolecule (Erzberger and Berger, 2006). There is, therefore, a switch between at least two protein conformational states, which is defined by the ATP binding and hydrolysis cycle (Erzberger and Berger, 2006).

Regardless of whether they are Class I or Class II enzymes, the AAA+ proteins contain an extra domain usually at the N-terminus (N domain), which varies considerably between family members (Dougan *et al.*, 2002). There are a number of examples where the AAA+ proteins seem to be regulated by adaptor proteins, for example, *in vitro* the *E. coli* ClpAP mediates degradation of SsrA-tagged proteins, a role that is switched to aggregated proteins by the ClpS adaptor (Dougan *et al.*, 2002). It would appear that the most divergent region of the AAA+ ATPases, the N domain, provides a common region that promotes recognition of the diverse adaptor proteins (Dougan *et al.*, 2002). The adaptor protein may switch the activity of particular AAA+ protein or, alternatively, the attention of the AAA+ ATPase may be diverted to a particular substrate without having an influence on other activities of that enzyme (Dougan *et al.*, 2002).

The conformational coupling mechanism of the AAA+ proteins can be tuned for its role. Certain AAA+ ATPases act as bimodal switches that are activated upon binding ATP and reset after ADP release; the ATP binding and hydrolysis in this case is a slow, regulated event. The opposite extreme describes AAA+ proteins that use this mechanism to propagate cycles of nucleotide binding and release, resulting in a processive motor activity. Remodelling factors reside somewhere between these two groups. They possess intermediate levels of activity that can be tweaked for a given function. The final examples of AAA+ proteins are those that lack ATP-binding and hydrolysis activity, for example, the δ and δ' subunits of the bacterial clamp loader in replication (Erzberger and Berger, 2006). In this situation, these proteins typically act as amplifiers or modulators for ATPase subunits that do possess activity within the oligomeric complex.

Previous pull down experiments, carried out by Dr Jen Roberts, identified an interaction between the hypothetical protein Sso1289 and PCNA 2. Inspection of the amino acid sequence of Sso1289 highlighted a possible hydrophobic PCNA interacting motif (LLVFTL) (Roberts, 2004). Initial bioinformatics analysis using 3D-PSSM (http://www.sbg.bio.ic.ac.uk/~3dpssm), which compares predicted protein folds, identified the bacterial helicases RuvB and UvrB as having 14 % and 12 % identity to Sso1289, respectively (Roberts, 2004). A BLASTP search indicated that Sso1289 was a member of the AAA family of ATPases.

RuvB is an example of an AAA ATPase with a role in DNA replication. RuvB particularly, as part of the RuvABC complex, is involved in migrating and subsequently processing the Holliday junction (HJ) structure that results from the regression of a stalled replication fork or from homologous recombination (HR). Typical of AAA proteins, RuvB forms a hexameric ring. It encircles the DNA and exhibits ATP-dependent helicase activity, which promotes branch migration (Ogura and Wilkinson, 2001).

Branch migration is an important step in genetic recombination and describes the movement of the HJ structure or exchange point between two homologous DNA duplexes. If regions of DNA sequence homology flank the branch point, then the HJ can spontaneously migrate in either direction. This occurs as a result of the exchange of hydrogen bonds between the DNA bases in the homologous DNA strands (Panyutin and Hsieh, 1994). The amount of genetic information that is exchanged between the homologues is governed by the extent of HJ branch migration. During recombination, however, the exchange of strands occurs between similar but not identical duplexes, thereby presenting sequence heterology. Also, strand exchange can involve hundreds to thousands of base pairs. In both of these cases proteins are required to promote branch migration (Panyutin and Hsieh, 1994).

A homologue of Sso1289 was also identified in *S. solfataricus*, another hypothetical protein referred to as Sso1468 (figure 6.1). The amino acid sequence of Sso1468 was the subject of a BLASTP search. Proteins representing close matches were members of the AAA ATPase family. No apparent motor for branch migration of a HJ has been recognised in *S. solfataricus* (Roberts *et al.*, 2003), therefore the ability of Sso1289 and Sso1468 to migrate HJ was the subject of investigation.

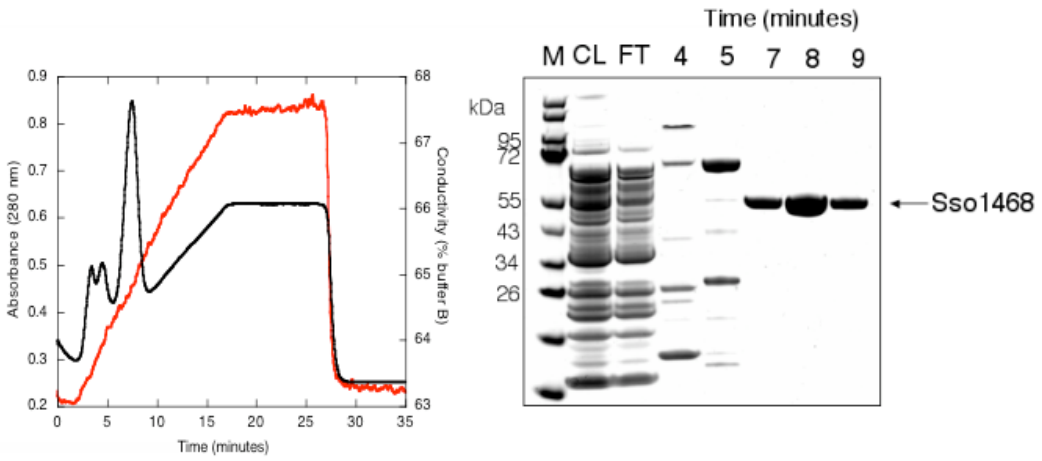
Figure 6.1. A Clustal W alignment of Sso1289 and Sso1468
The *S. solfataricus* hypothetical proteins Sso1289 and Sso1468 share 55 % similarity and 36.9 % identity with each other. The Walker A box is highlighted in red and the Walker B box is shown in blue.

No expression of Sso1289 was detected in any of the expression conditions investigated and so the initial characterisation of the hypothetical *S. solfataricus* protein, Sso1468, is described in this chapter.

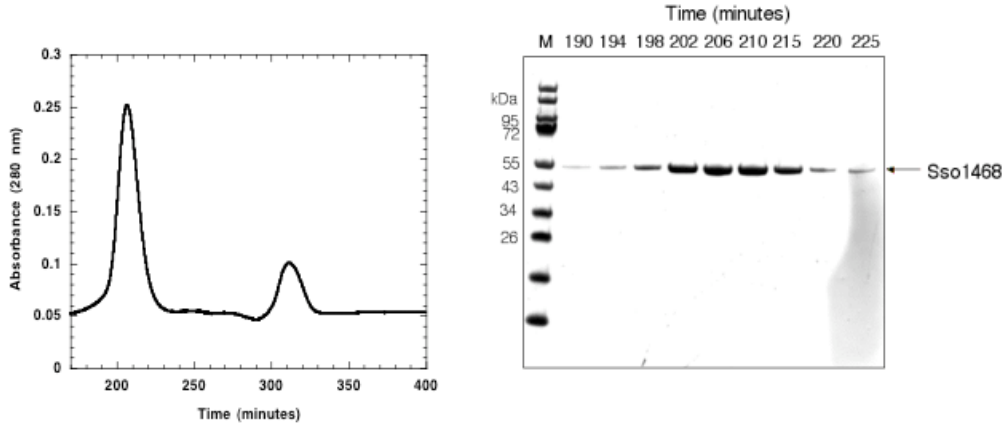
6.2 Expression and purification of Sso1468

The *sso1468* gene was cloned into the pDEST14 vector with and N-terminal polyhistidine tag, using the Gateway® Cloning system and was provided by Dr Huanting Liu (Biomolecular Sciences, St Andrews University). Sso1468 was purified by heat treatment and a combination of nickel column and gel filtration column chromatography as described in section 2.3.5 (figure 6.2).

A



B



C

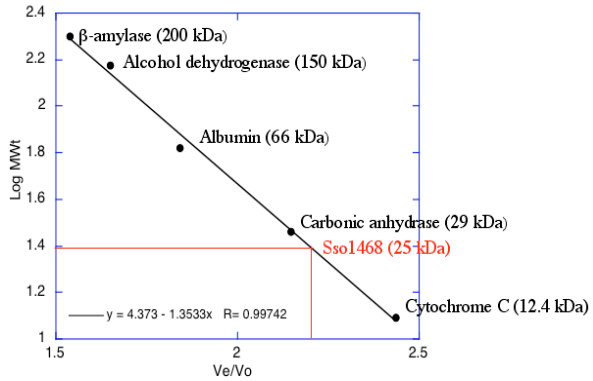


Figure 6.2. Purification of polyhistidine tagged Sso1468
Soluble *E. coli* cell lysate containing overexpressed Sso1489 was heated before being applied to A) the nickel column. Protein was eluted from the column across a 30 – 500 mM imidazol gradient. The peak fractions were analysed by SDS-PAGE and the relevant fractions were concentrated and passed down B) an equilibrated gel filtration column. Sso1468 was identified on an SDS-PAGE gel. The absorbance trace and SDS-PAGE gel for each column are shown. C) Size determination of Sso1468: the gel filtration column was calibrated with standards of known molecular weights. Sso1468 was estimated to have a molecular weight of 25 kDa compared to an expected value of 48,199.0 Da.

Fractions containing the protein were pooled and concentrated in dialysis tubing using PEG 8000. This seemed to maintain the protein in a soluble state in contrast to other methods, which resulted in precipitation of Sso1468. The Walker A box mutant Sso1468 K53A was prepared by site directed mutagenesis with Pfu Turbo DNA polymerase, the reaction was optimized to include a 1 minute 55 °C annealing step and a 10 minute elongation phase at 68 °C. This was considered as a negative control for the ATPase assays, since the invariant lysine of the Walker A box is crucial for ATP hydrolysis by ATPases (section 1.4.2). Sso1468 K53A was expressed and purified according to the protocol described for WT Sso1468, except that this mutant was not subjected to a heat step before column chromatography, since this decreased the yield to almost nothing. DNA sequencing and mass spectrometry confirmed the identity of both WT Sso1468 and the Walker A box mutant K53A. Gel filtration analysis indicated that Sso1468 was a monomer, it eluted from the calibrated gel filtration column with an estimated molecular weight of 25 kDa (figure 6.2C). This was very low compared to the expected molecular weight of 48,199.0 Da. As described for XPB2 (section 3.4), the late elution of Sso1468 may be due to tight folding of the protein or interactions with the column material (Morris *et al.*, 2001).

6.3 Sso1468 binds to a range of DNA substrates

The binding affinity of Sso1468 to a range of DNA substrates was investigated using electrophoretic mobility shift assays (EMSA), as described for XPB1 and XPB2. A range of DNA substrates were prepared, a 50 nt single strand (J150B), a 50 mer duplex (J150B + complement), a 50 nt duplex with a 3' 25 T overhang (J150B + complement with 3' 25 nt T tail) and a HJ (J150B, J150X, J150R and J150H) each arm of which was 25 bp long (figure 6.3).

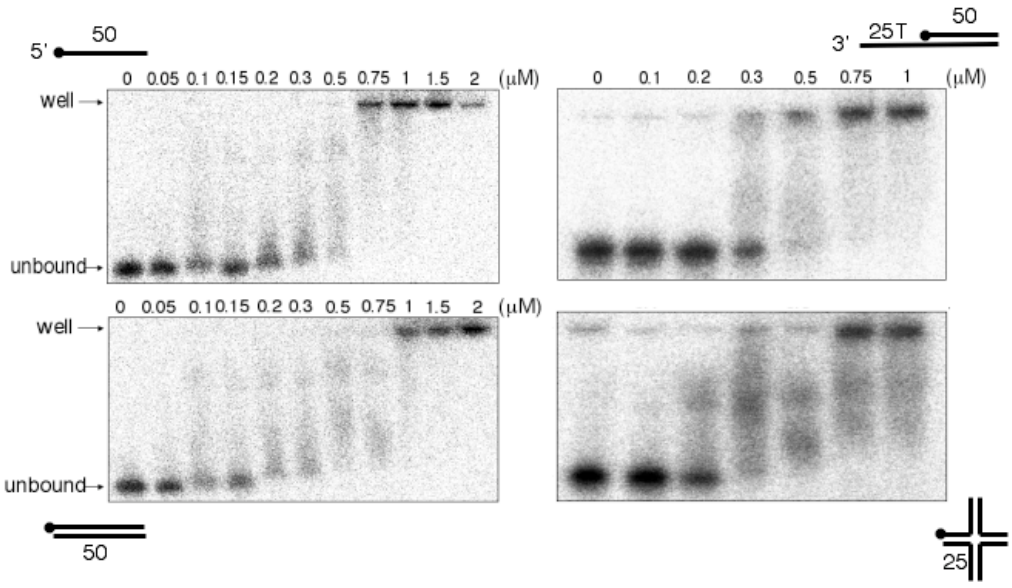


Figure 6.3. Electrophoretic Mobility Shift Assay with Sso1468
The DNA binding affinity of Sso1468 was investigated by incubating the protein at the concentrations shown with 10 nM ^{32}P-5'-radiolabelled (black circle) DNA substrate at room temperature for 10 minutes. The substrates used were a 50 nt single strand, a 50 nt duplex, a 50 nt duplex with a 25 nt 3' T tail and a HJ (each arm was 25 bp long). The samples were separated on 12 % polyacrylamide : TBE gels.

Sso1468 bound to all four of these DNA substrates, although the gels were difficult to quantify since the DNA-protein complexes remained in the wells. The gel pictures did not initially present an obvious difference in the binding affinity of Sso1468 for the four different substrates. With the proposal that Sso1468 could be a branch migration protein, Sso1468 was expected to show a binding preference for the HJ substrate. On closer inspection, binding of Sso1468 to the dsDNA, the HJ and even the 3' overhang substrates was evident at slightly lower concentrations of Sso1468 compared to the ssDNA substrate, although there was no significant difference (figure 6.3). The smears evident on the gel pictures represent various slow migrating protein-DNA complexes, as described in section 3.5.

6.4 ATP hydrolysis induces a conformational change in Sso1468

The ATPase activity of Sso1468 was investigated using the malachite green assay described for XPB1 and XPB2. Initial experiments were carried out exactly according to the XPB protocols, but no free phosphate was detected. The temperature, the protein

concentration, the DNA concentration and the incubation times were all factors that were varied in an attempt to optimise the assay conditions. There was still no evidence of ATP hydrolysis (data not shown).

This was reminiscent of RadB, a euryarchaeal Rad-51 like protein (Rashid *et al.*, 1996); *Haloferax volcanii* RadB did not seem to exhibit ATPase activity, despite the conserved ATPase fold (Guy *et al.*, 2006). Circular Dichroism (CD) Spectroscopy was used to investigate the binding of RadB to ATP. A conformational change was evident when RadB was incubated with ATP but not seen when incubated alone or with ADP, ATPγS (non-hydrolysable analogue of ATP), GTP, CTP or DNA.

In reference to this publication, CD spectroscopy was used to detect an interaction between Sso1468 and ATP. CD spectroscopy is a method by which the difference in absorbance between the two circularly polarized components (L and R) of plane polarised light is measured (reported in terms of ellipticity (θ) in degrees). R and L rotate in the clockwise and counter clockwise directions, respectively, and are of equal magnitude. If L and R are not absorbed, or are absorbed by the sample to the same extent, then recombining the two components results in polarised radiation in the original plane (Kelly *et al.*, 2005). Alternatively, the radiation generated if the L and R components are absorbed to different extents is described as possessing elliptical polarisation.

When a chromophore is optically active, otherwise referred to as chiral, a CD signal is observed. When studying proteins, a number of spectral regions can provide useful structural information. The peptide bond (absorption below 240 nm), aromatic amino acid side chains (absorption in the range 260 to 320 nm) and disulphide bonds (weak broad absorption bands centered around 260 nm) are all chromophores of interest (Kelly *et al.*, 2005). Absorption in the region below 240 nm (far UV) is particularly important for determining the secondary structure composition and is mainly due to the peptide bond. CD spectra in the far UV region are characteristic of the different secondary structures of proteins (figure 6.4). Algorithms are used to automatically estimate the secondary structure of the protein of interest, based on CD analysis of proteins with already known structures solved by X-ray crystallography.

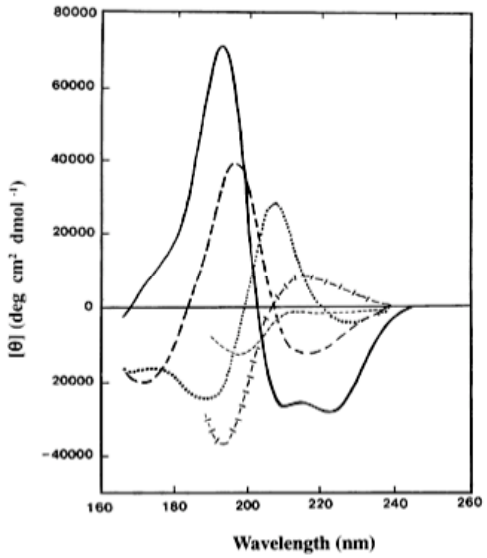


Figure 6.4. Characteristic CD spectra of typical secondary structures in the far UV range
Absorption in the region below 240 nm (far UV) is particularly important for determining the secondary structure composition and is mainly due to the peptide bond. The different secondary structures of proteins are responsible for particular CD spectra: α-helix – solid line; antiparallel β-sheet – long dashed line; type I β-turn – dotted line; extended 3_1-helix or poly (Pro) II helix – cross dashed line; irregular structure – short dashed line (Kelly *et al*., 2005).

The secondary structure of Sso1468 was visualised in the far UV region, as the different secondary structures give characteristic CD spectra (Kelly *et al*., 2005). Protein (3.3 μM) was incubated with 3.3 μM nucleotide and the results were corrected for buffer background absorbance.

CD spectroscopy analysis of Sso1468 produced a spectrum characteristic of a predominantly α-helical protein (Kelly *et al*., 2005); this same result was observed when Sso1468 was incubated with GTP, CTP and ATPγS (figure 6.5). Upon incubation with ATP, however, there was a shift in the spectrum representative of a subtle conformational change, similar to that seen by Guy *et al*. (2006). The walker A box mutant Sso1468 K53A was also analysed using CD spectroscopy. The yield after purification was low so it was incubated at a much lower concentration. Due to the reduced concentration the absorbance levels were much lower and, although it did not look to show a conformational change in the presence of ATP as expected, it was inconclusive. Such a change would be harder to detect at the lower concentrations as the shift in the spectra would be much more subtle.

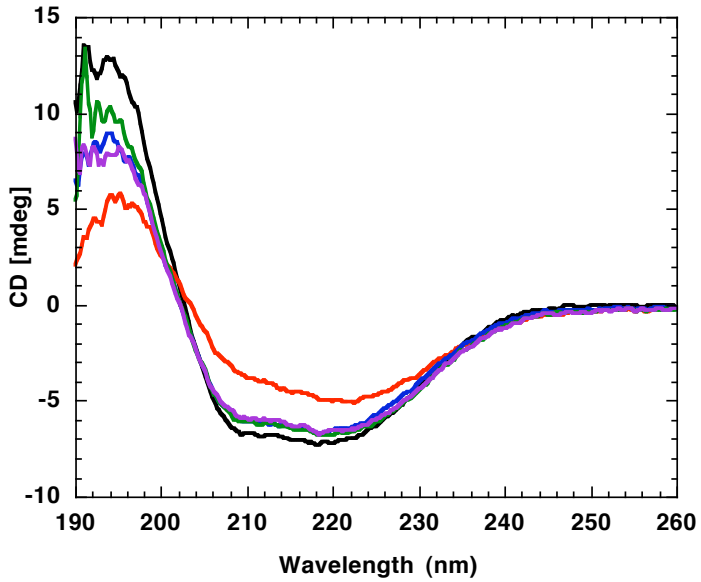


Figure 6.5. Circular Dichroism spectroscopy analysis of Sso1468 binding ATP
The secondary structure of Sso1468 was determined by CD spectroscopy and the absorbance in the far UV region (< 240nm) was recorded. Sso1468 was incubated with a range of nucleotides at equimolar concentrations (3.3 μM): black represents Sso1468 without a nucleotide; Sso1468 incubated with ATP is shown in red; blue indicates Sso1468 incubated with ATPγS; and green and purple represent Sso1468 incubated with GTP and CTP respectively.

This conformational change indicated that Sso1468 was in fact hydrolyzing ATP, since no change was apparent with ATPγS. This was not consistent with the ATPase assay results. Guy *et al*. (2006) proposed a possible auto-phosphorylation event since ATP was being hydrolysed yet free phosphate was not being detected. Sso1468 was incubated with ATP and the whole molecular mass was determined by mass spectrometry compared to that of the protein alone (figure 6.6).

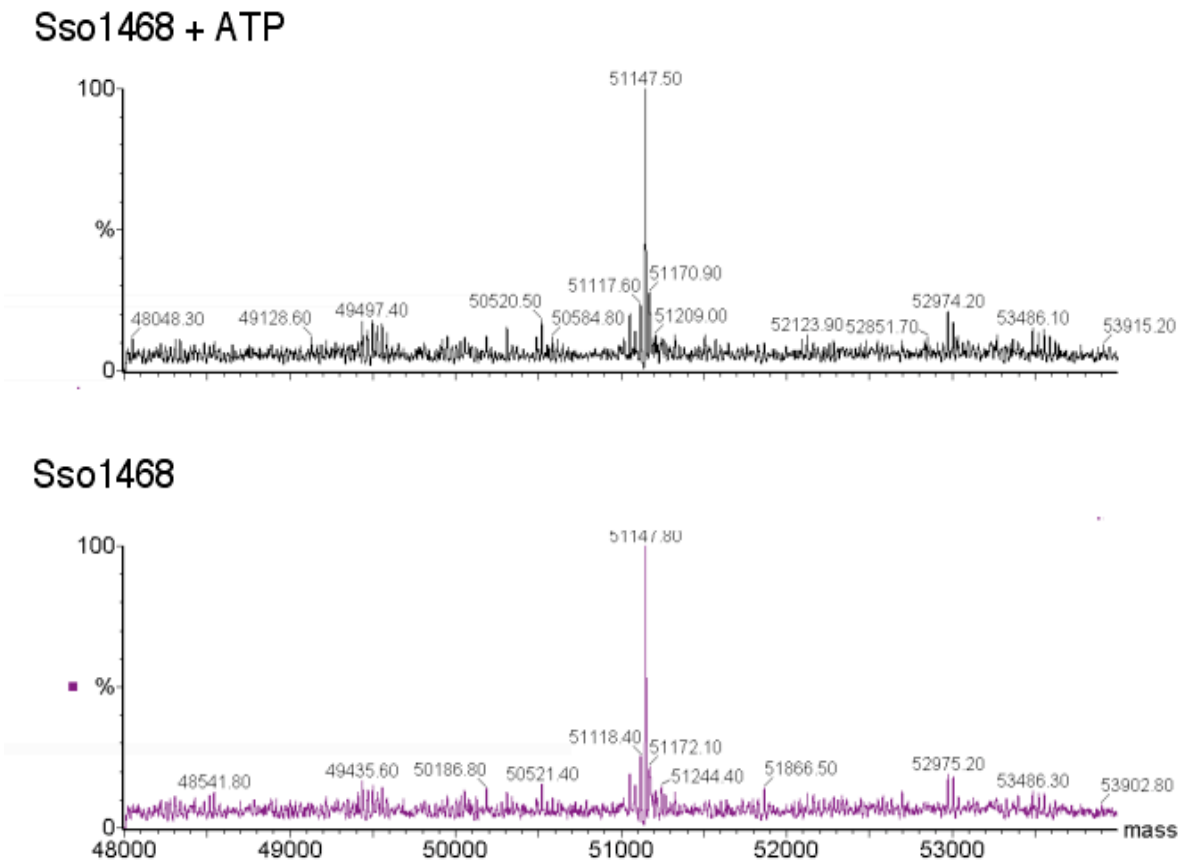


Figure 6.6. Mass spectrometry analysis of Sso1468 auto-phosphorylation
10 μM Sso1468 was incubated with 10 μM ATP for 30 minutes at room temperature (black). This was sent analysed by mass spectrometry to 0.1 Da. 10 μM 1468 was also analysed as a control (purple).

It was also incubated with radiolabelled ATP in order to be able to directly see the addition of phosphate onto an amino acid residue of Sso1468 (figure 6.7). Neither of these experiments provided evidence to support an auto-phosphorylation event.

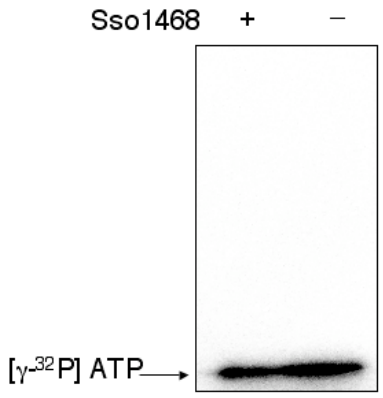


Figure 6.7. Incubation of Sso1468 with [γ-^{32}P] ATP
0.2 μM Sso1468 was incubated with 10 μCi [γ-^{32}P] ATP for 30 minutes at room temperature. A control sample was prepared without Sso1468. The reactions were analysed on a 12 % polyacrylamide: TBE gel.

The possibility that ATP hydrolysis may affect the quaternary structure was investigated by passing Sso1468 down the calibrated gel filtration column, incubated alone or in combination with ATP or ATPγS (Guy *et al.*, 2006). The results are not shown, but Sso1468 was eluted from the column with same retention time typical of a monomer, regardless of the nucleotide present.

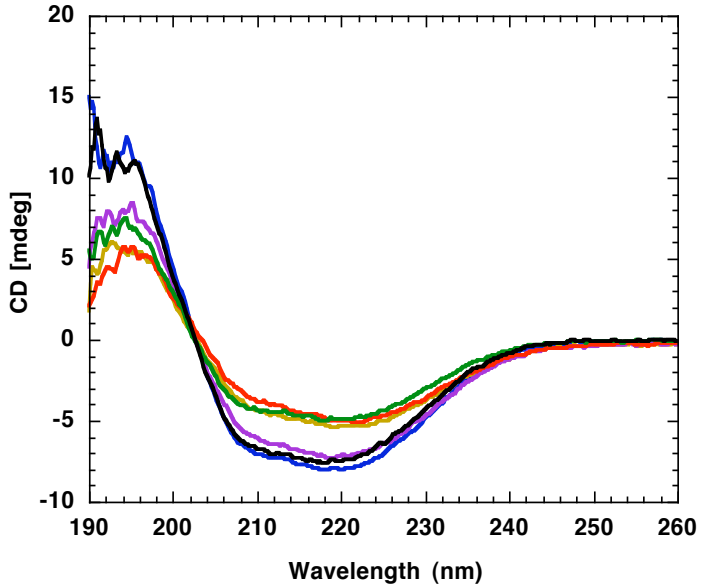


Figure 6.8. The effect of DNA on the ATP induced conformation change of Sso1468
The effect of DNA on the ATP induced conformational change in Sso1468 was investigated by incubating Sso1468 with DNA and ATP in different combinations. Both Sso1468 and ATP were used at a final concentration of 3.3 μM; an excess of DNA was used (2 μM). Black represents Sso1468 without a nucleotide; Sso1468 incubated with ATP is shown in red; blue represents Sso1468 incubated with DNA alone; Sso1468 was incubated with ATP prior to the addition of DNA this is shown in green; gold represents addition of ATP after a 15 minute pre incubation of Sso1468 with DNA; and a pre incubation of Sso1468 with DNA overnight before addition of ATP is shown in purple.

Guy *et al.* (2006) found that incubation of RadB with DNA prior to ATP addition abolished the conformational change and, therefore, the effect of DNA binding on Sso1468 was investigated (figure 6.8). Sso1468 was mixed with ATP and DNA in a number of combinations, the DNA was used at a final concentration of 2 μM to ensure Sso1468 was binding. The data indicated that DNA only seemed to inhibit ATP-induced conformational changes when incubated with Sso1468 for a prolonged period, otherwise the DNA had no effect. It seemed that once the DNA was bound, it blocked access of ATP to the catalytic site, yet if ATP was present it successfully competed for binding.

The data supports the original idea that Sso1468 hydrolyses ATP and it was proposed that ATP turnover was too low to be detected by malachite green. Therefore, the ATPase assays were repeated with much higher concentrations of Sso1468 in order to enable detection of low levels of free phosphate. The data presented in figure 6.9 is in good agreement with a very low turnover of ATP as catalysed by Sso1468, 38 ± 4.5 pmoles phosphate.min^{-1}.pmole enzyme^{-1}.

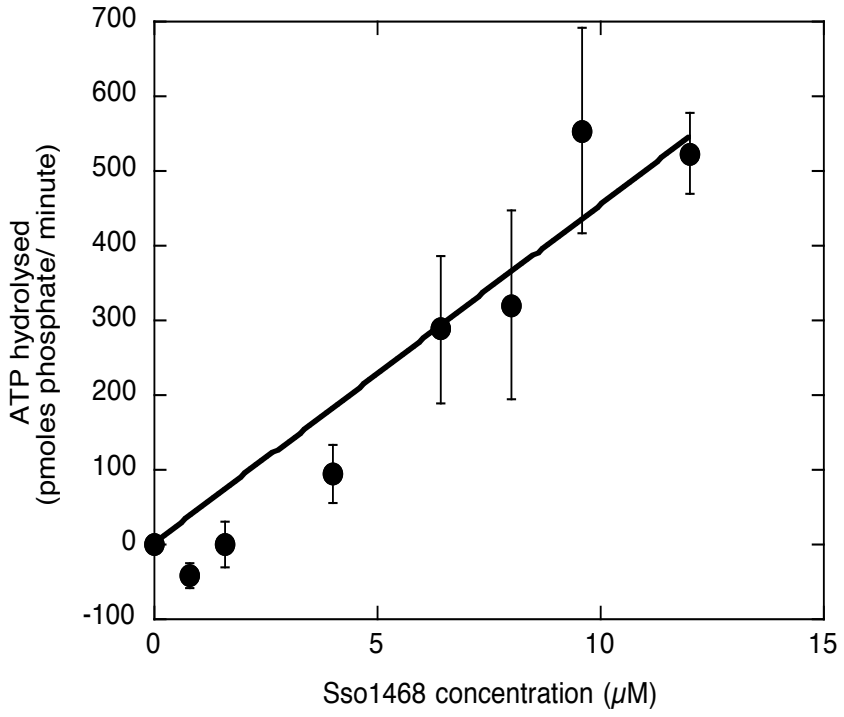


Figure 6.9. ATPase activity of Sso1468
Sso1468 was incubated at the concentrations plotted with 10 nM PhiX174 virion ssDNA substrate at 45 °C. Samples were taken after 15 minutes and added to chilled 0.3 M PCA to stop the reaction. The orange to green colour change of the malachite green was measured at 590 nm after a 12 minute incubation.

6.5 Sso1468 does not migrate Holliday junctions

Proteins involved in branch migration can also exhibit helicase activity; this enables them to migrate a static junction by unwinding the duplex portions. Those that do not have a helicase function, for example Rad54 (Petukhova *et al*., 1998; Bugreev *et al*., 2006), are only able to migrate junctions that have homologous regions and are, therefore, mobile (Bugreev *et al*., 2006). A mobile HJ (11M, 10, 21 and 22, see appendix 2 for sequences) was constructed in order to be able to detect any branch migration activity exhibited by Sso1468. The junctions were designed to incorporate a mismatch in order to prevent spontaneous branch migration (Bugreev *et al*., 2006).

Sso1468 was incubated with a large mobile HJ at 30 °C for 30 minutes (figure 6.10). The levels of spontaneous branch migration were kept to a minimum by addition of $Mg(CH_3COO)_2$ to the reaction buffer and $MgCl_2$ to the gel running buffer, magnesium has previously been shown to significantly reduce the levels of spontaneous branch migration (Panyutin and Hsieh, 1994). No activity was seen for Sso1468. The assay was lacking positive controls and, therefore, the data was inconclusive.

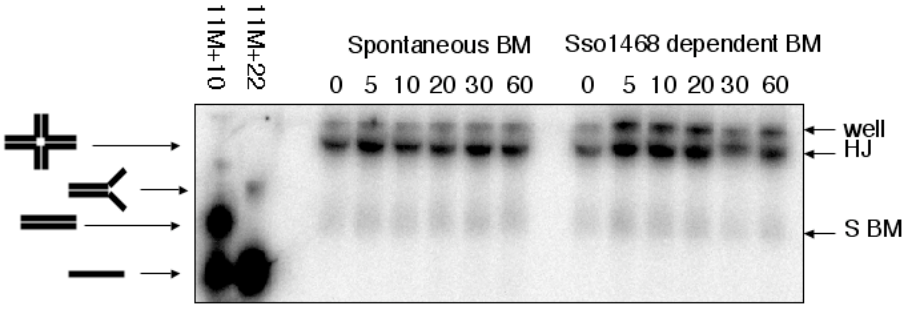


Figure 6.10. Electrophoretic analysis of the branch migration activity of Sso1468
To investigate the ability of Sso1468 to catalyse ATP-dependent migration of a mobile Holliday junction, 0.5 μM Sso1468 was incubated with 10 nM mobile HJ (^{32}P-5'-radiolabelled) at 30 °C for 30 minutes (final volume 70 μl). Samples were taken at the time points indicated and the reaction was stopped in STOP solution (materials and methods). The samples were run on 8 % polyacrylamide: TBE gel at 4 °C to prevent the HJ substrates dissociating. HJ represents the Holliday junction and the spontaneous branch migration is referred to as S BM.

6.6 No interaction with PCNA was detected

Sso1289 was identified to interact with PCNA2, it possesses a possible hydrophobic PCNA interacting motif (PIP box) in its N-terminal region (LLVFTL) (Roberts, 2004). Sso1468 is a homologue of Sso1289 and to follow up on the proposal

that Sso1468 might also interact with PCNA, Sso1468 was passed down a calibrated gel filtration column alone and in combination with heterotrimeric Sso PCNA, as described for XPB2 and Bax1. There was no indication of an interaction between PCNA and Sso1468 in solution since the two proteins eluted separately from the column (figure 6.11).

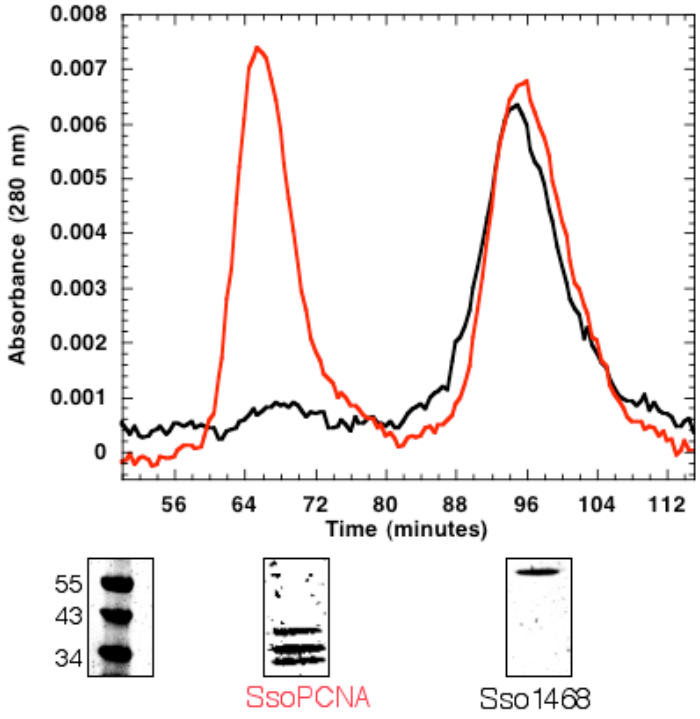


Figure 6.11. Gel filtration analysis of PCNA and Sso1468 in combination
Sso1468 and Sso PCNA were passed through a calibrated gel filtration column separately and in combination to identify an interaction between Sso1468 and PCNA. Sso1468 elution is shown in black and the red represents the combined Sso1468 and PCNA column run. The peak fractions were analysed by SDS-PAGE.

6.7 Discussion

AAA ATPase proteins have a large number of roles in the cell. BLAST P analysis identified Sso1468 as an AAA ATPase based on its protein sequence, although preliminary ATPase assays failed to detect any ATP hydrolysis. CD spectroscopy results indicated that Sso1468 was undergoing a conformational change upon incubation with ATP. This structural change was not observed when Sso1468 was incubated with GTP, CTP, ADP, DNA or ATPγS, the non-hydrolysable analogue of ATP, suggesting that Sso1468 was hydrolyzing the ATP consistent with the result seen with RadB from *H*.

volcanii (Guy *et al.*, 2006). An autophosphorytlation event was ruled out when neither mass spectrometry nor experiments with radioactive ATP could detect the addition of a phosphate onto the protein. Sso1468 hydrolyses ATP and releases the free phosphate, which was only detected when the ATPase assays were repeated with high protein concentration. This implies that Sso1468 exhibits a very low turnover of ATP.

Under the conditions used, Sso1468 was identified as a monomeric protein in the presence of DNA or nucleotide by gel filtration column analysis, suggesting that any changes seen upon ATP hydrolysis were the result of changes in the monomer conformation rather than changes to the oligomeric state of Sso1468 (Guy *et al.*, 2006). This is unusual as many of the AAA ATPase proteins form hexameric rings (Ogura and Wilkinson, 2001), although some are prone to disassembly upon ATP hydrolysis (Vale, 2000). The buffer conditions used for these experiments may not have been adequate to promote oligomerisation of Sso1468 and subsequent ATPase activity. The preliminary branch migration assays, although inconclusive due to the lack of positive controls, indicated that Sso1468 was unable to migrate a mobile HJ.

AAA ATPases are hypothesised to form hexameric structures, however, some AAA ATPases exist as monomers and exhibit low, but significant, ATPase activity. These proteins may form transient oligomers under ATPase assay conditions. *S. cerevisiae* Vsp4p is an example of an AAA ATPase that undergoes concentration dependent oligomerisation, which results in the dramatic increase of ATP hydrolysis (Babst *et al.*, 1998; Karata *et al.*, 1999). The ATPase activity drives a cycle of assembly and disassembly of the monomeric and oligomeric Vsp4p species.

The ATP binding and hydrolysis catalysed by Sso1468 could be a slow, but highly regulated event causing Sso1468 to act as a bimodal switch. In this case, ATP binding would activate Sso1468 and hydrolysis and release of ADP would reset Sso1468 ready for the next controlled cycle (Erzberger and Berger, 2006) Alternatively, Sso1468 may require homo- or hetero-protein interactions to stimulate its activity or the activity of the interacting protein. This is common to many proteins that possess the Walker A and B box motifs; they function as part of a complex or require activation as part of their control mechanism (Singleton *et al.*, 2000). An example of an ATP induced regulation mechanisms involves RadB, which interacts with and inhibits the archaeal HJ resolvase

Hjc in the absence of ATP. When bound to ATP, a conformational change in RadB causes it to dissociate from Hjc, which then becomes active and can catalyse HJ resolution (Komori *et al*., 2000). RadB and Hjc have been proposed to form a complex, possibly recruiting other proteins to stimulate branch migration. Only when a suitable cleavage site is reached does RadB dissociate, acting as a switch to negatively regulate Hjc (Komori *et al*., 2000; Guy *et al*., 2006). One possibility is that Sso1468, like RadB, is involved in branch migration as part of a protein complex, acting as a control mechanism to prevent premature junction resolution, for example. It would be interesting to identify protein partners for Sso1468 and to discover whether or not this hypothetical protein shares a similar role with that described for RadB (Komori *et al*., 2000).

Sso1289 was shown to interact with PCNA, possible via the LLVFTL sequence in the N-terminal region of the protein. Sso1468 did not exhibit any interaction with PCNA. This is consistent with the comparison of the two sequences (figure 6.1), since neither the N-terminal extension nor the PIP-box are conserved in the Sso1468 protein sequence.

It would be interesting to deduce whether it acts to regulate a partner protein or whether it relies on such interactions for its own activation. Further investigations are required to understand how Sso1468 might function *in vivo* and to identify the significance of the ATP hydolysis-induced conformation change.

CHAPTER 7: CONCLUSIONS AND FUTURE WORK

7.1 Conclusions and future work

The characterisation of the putative helicases XPB1, XPB2 and Hel308 and the hypothetical protein Sso1468 from *S. solfataricus* are described in this thesis.

XPB1 and XPB2 were shown to bind preferentially to ssDNA rather than a dsDNA substrate. XPB1 binds DNA with an affinity consistent with other SF2 helicases (Voloshin and Camerini-Otero, 2007), in contrast, XPB2 exhibited a very low DNA binding affinity for DNA. Both XPB1 and XPB2 contain the seven motifs characteristic of SF1 and SF2 helicase, although neither protein was able to displace streptavidin from a biotinylated probe. Furthermore, XPB1 and XPB2 were unable to catalyse DNA strand separation consistent with recently published studies with human XPB (Coin *et al.*, 2007).

It is known that human XPB binds the DNA downstream of the promoter, which positions it too far from the melted DNA to act as a conventional helicase and it is thought to act as an ATP-dependent conformational switch (Kim *et al.*, 2000; Lin *et al.*, 2005). It would be of great interest to pursue this observation further in the archaeal system and investigate the positioning of SsoXPB1 and/or SsoXPB2 at the melted promoter and the possibility that both SsoXPB1 and SsoXPB2 could act as a "ratchet wrench" (Kim *et al.*, 2000) to promote formation of the stable nascent transcription bubble.

Most archaeal genomes possess homologues of repair genes, such as XPB (Grogan, 2000; She *et al.*, 2001). The presence of two *xpb* genes seemed to be a feature of the crenarchaeal genomes inspected, a feature most likely the consequence of a duplication event. RT PCR and western blot analysis detected more *xpb1* mRNA and protein respectively, compared to that of XPB2 in the *S. solfataricus* cell. This suggested that the two proteins could fulfill separate functions within the cell, although this was by no means conclusive. Closer inspection of the archaeal genomes identified a gene in close association with the *xpb2* gene. Conservation of this pair implied that the two gene products might interact physically and functionally, similar to the interaction between human XPB and p52 (Coin *et al.*, 2007). SsoXPB2 and SacBax1 were shown to interact with each other. In contrast to p52, however, SacBax1 was unable to stimulate the activity of SsoXPB2, a result of a possible functional barrier presented by the heterologous origins of the two proteins. Alternatively, it would be interesting to

investigate whether Bax1 functions to recruit XPB2 in addition to a role in modulating the catalytic activity of XPB2 in archaeal transcription and/or repair. More information is required to understand the importance of the XPB2-Bax1 interaction; it would be invaluable to obtain recombinant XPB2 and Bax1 from the same organism. This would require either the recloning of the *S. solfataricus* or *S. acidocaldarius* genes into a range of expression vectors or cloning, expressing and purifying XPB2 and Bax1 from a different species.

In order to gain more of an insight onto the interaction between SsoXPB2 and SacBax1, mass spectrometry can be used to identify the interacting surface. The specific technique is referred to as isotope-coded protein label (ICPL) and is based on differential isotope labelling of specific amino groups on isolated intact proteins with either light or heavy tags (Schmidt *et al*., 2005). Quantitative analysis of the light and heavy labelled peptides is achieved by direct comparison of the relative signal intensities in the MS-spectra (Schmidt *et al*., 2005). To date, the results have been inconclusive (Wilkinson, 2007). This work, however, is ongoing and is being carried out by Dr Sally Shirran and Dr Catherine Botting (BMS Mass Spectrometry and proteomics facility, St Andrews University).

In order to confirm a role for XPB1 and XPB2 in NER, it would be interesting to carry out knock out studies in *S. solfataricus*. By knocking out both *xpb* genes, *xpb1* or *xpb2* and assessing colony survival after UV irradiation an involvement of XPB1 and XPB2 in NER could be identified. Knocking out one gene at a time would focus investigations on a plausible compensatory role of XPB1 for XPB2, or vice versa, which may provide a possible explanation for the presence of two *xpb* genes in the genome of *S. solfataricus* and other crenarchaea.

In stark contrast to the results presented for XPB1 and XPB2, the putative helicase Hel308 was able to unwind duplex DNA. Hel308 bound preferentially to ssDNA and was able to displace streptavidin from a biotinylated DNA probe. A number of arginine residues thought to be directly involved in DNA binding, based on the crystal structure, were separately mutated to alanine residues. In each case reduced DNA binding and, therefore, reduced catalytic activity was evident.

Unexpectedly, mutation of the RAR motif in domain 5 of Hel308 or the removal of this entire domain produced a more efficient helicase. Domain 5 was, therefore, identified as an autoinhibitory domain. This was an unexpected result since domain 5 consists of two helix-turn-helix motifs, such motifs are known to be involved in DNA binding (Luscombe *et al.*, 2000). Further research must be undertaken to understand the mechanism by which domain 5 inhibits the activity of the intact protein, its importance, and how it is relieved *in vivo*. There are a number of possibilities including protein-protein interactions, conformational changes or proteolysis. Identification of interacting protein partners could provide insight into the activation mechanism.

Hel308 is thought to be involved in clearing the lagging strand template at a stalled replication fork, either by displacing the lagging DNA strand or by the displacement of SSB, in order for RadA to load onto the DNA and promote strand exchange (as proposed in figure 5.18). Another possibility is that Hel308 may act to inhibit strand exchange by disruption of the RadA nucleofilament. Further investigations are required to identify the exact role of Hel308 at this stage of replication restart. This work on Hel308 is ongoing and is being carried out by Annie McRobbie (PhD student, Professor Malcolm White group, BMS, St Andrews University).

It will be interesting to identify any inhibitory effects of the DNA binding proteins SSB and Alba and the strand exchange protein RadA on the catalytic activities of Hel308. Since the block imposed by domain 5 must somehow be relieved, the effect of SSB, Alba and RadA on the activity of the truncation mutant could provide some insight into the inhibition release mechanism.

Nothing is known so far about the way in which Hel308 loads onto the DNA. The crystal structure shows that there are no covalent links between domains 2 and 4. This suggests that the open ring structure is flexible to allow Hel308 to load onto DNA that does not have a 3' end, such as a stalled replication fork. A loop connecting domains 2 and 3 has been proposed to act as the hinge for this movement (Richards *et al.*, 2007).

In order to investigate this movement further, spin labelling studies will be carried out. By incorporating a cysteine residue and subsequently a methanethiosulfonate label into domain 2 and domain 4 the distance between the domains can be measured. A spin label is an organic molecule that possesses an unpaired electron (usually on a nitrogen atom) and is, therefore, said to be paramagnetic. Spin labels are very sensitive to motion,

which is dictated by the local environment and has profound effects on the Electron Paramagnetic Resonance (EPR) spectrum. EPR is the technique used to study molecules with unpaired electrons. Any conformational change in the ring structure will result in a change in the local environment and to the distance between the two domains and subsequently between the two methanethiosulfonate labels (Gross *et al*., 1999; Steinhoff, 2002).

This technique will also be used to investigate the conformational changes to Hel308 during ATP hydrolysis and, therefore, spin labels have been incorporated into domains 1 and domains 2. These investigations are being carried out by Annie McRobbie (PhD student, Professor Malcolm White group, BMS, St Andrews University).

The hypothetical *S. solfataricu*s protein Sso1468 was thought to be an AAA ATPase, based on a BLASTP alignment search, with a role as a branch migration protein. Initially no ATP hydrolysis could be detected, however, a conformational change upon ATP (and not ATPγS) binding was identified. After an autophosphorylation event was ruled out, it was shown that Sso1468 exhibits a very low turnover of ATP. In contrast to many AAA ATPases, Sso1468 did not form a hexameric complex but was identified as a monomer in the conditions used. Some AAA ATPases do exist as monomers and exhibit low ATPase activity. It is thought that these enzymes undergo transient and concentration dependent oligomerisation under ATPase assay conditions (Babst *et al*., 1998; Karata *et al*., 1999). Alternatively Sso1468 may require protein interactions to stimulate its own activity or that of the interacting partner.

It would be interesting to identify proteins that interact with Sso1468. Mutagensis studies of Sso1468 could provide some insight into its function; for example, Sso1468 could possess an autoinhibitory domain similar to that described for Hel308. No branch migration activity was detected with Sso1468, the experiments, however, lacked positive controls. This is another area of interest that would need further investigation.

The research presented in this thesis covers a number of *S. solfataricus* proteins that function to maintain the genome. The study of XPB1, XPB2, Sso1468 and in particular Hel308 from the archaeal system provides some insights into the complicated environment of the eukaryotic cell.

References

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

Acharya, S., Foster, P.L., Brooks, P., and Fishel, R. (2003). The coordinated functions of the E. coli MutS and MutL proteins in mismatch repair. Mol Cell *12*, 233-246.

Ahnert, P., and Patel, S.S. (1997). Asymmetric interactions of hexameric bacteriophage T7 DNA helicase with the 5'- and 3'-tails of the forked DNA substrate. J Biol Chem *272*, 32267-32273.

Aravind, L., Walker, D.R., and Koonin, E.V. (1999). Conserved domains in DNA repair proteins and evolution of repair systems. Nucleic Acids Res *27*, 1223-1242.

Babst, M., Wendland, B., Estepa, E.J., and Emr, S.D. (1998). The Vps4p AAA ATPase regulates membrane association of a Vps protein complex required for normal endosome function. Embo J *17*, 2982-2993.

Barns, S.M., Delwiche, C.F., Palmer, J.D., and Pace, N.R. (1996). Perspectives on archaeal diversity, thermophily and monophyly from environmental rRNA sequences. Proc Natl Acad Sci U S A *93*, 9188-9193.

Barry, E.R., and Bell, S.D. (2006). DNA replication in the archaea. Microbiol Mol Biol Rev *70*, 876-887.

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

Bell, C.E. (2005). Structure and mechanism of Escherichia coli RecA ATPase. Mol Microbiol *58*, 358-366.

Bell, S.D., and Jackson, S.P. (1998). Transcription and translation in Archaea: a mosaic of eukaryal and bacterial features. Trends Microbiol *6*, 222-228.

Bell, S.P., and Dutta, A. (2002). DNA replication in eukaryotic cells. Annu Rev Biochem *71*, 333-374.

Bennett, R.J., and Keck, J.L. (2004). Structure and function of RecQ DNA helicases. Crit Rev Biochem Mol Biol *39*, 79-97.

Bernander, R. (2000). Chromosome replication, nucleoid segregation and cell division in archaea. Trends Microbiol *8*, 278-283.

Bianco, P.R., and Kowalczykowski, S.C. (2000). Translocation step size and mechanism of the RecBC DNA helicase. Nature *405*, 368-372.

Bjornson, K.P., Wong, I., and Lohman, T.M. (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*, 411-422.

Branzei, D., and Foiani, M. (2007). Interplay of replication checkpoints and repair proteins at stalled replication forks. DNA Repair (Amst) *6*, 994-1003.

Brendza, K.M., Cheng, W., Fischer, C.J., Chesnik, M.A., Niedziela-Majka, A., and Lohman, T.M. (2005). Autoinhibition of Escherichia coli Rep monomer helicase activity by its 2B subdomain. Proc Natl Acad Sci U S A *102*, 10076-10081.

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

Bugreev, D.V., Mazina, O.M., and Mazin, A.V. (2006). Rad54 protein promotes branch migration of Holliday junctions. Nature *442*, 590-593.

Burgers, P.M., Koonin, E.V., Bruford, E., Blanco, L., Burtis, K.C., Christman, M.F., Copeland, W.C., Friedberg, E.C., Hanaoka, F., Hinkle, D.C., Lawrence, C.W., Nakanishi, M., Ohmori, H., Prakash, L., Prakash, S., Reynaud, C.A., Sugino, A., Todo, T., Wang, Z., Weill, J.C., and Woodgate, R. (2001). Eukaryotic DNA polymerases: proposal for a revised nomenclature. J Biol Chem *276*, 43487-43490.

Buttner, K., Nehring, S., and Hopfner, K.P. (2007). Structural basis for DNA duplex separation by a superfamily-2 helicase. Nat Struct Mol Biol *14*, 647-652.

Byrd, A.K., and Raney, K.D. (2004). Protein displacement by an assembly of helicase molecules aligned along single-stranded DNA. Nat Struct Mol Biol *11*, 531-538.

Byrd, A.K., and Raney, K.D. (2006). Displacement of a DNA binding protein by Dda helicase. Nucleic Acids Res *34*, 3020-3029.

Cann, I.K., and Ishino, Y. (1999). Archaeal DNA replication: identifying the pieces to solve a puzzle. Genetics *152*, 1249-1267.

Caruthers, J.M., and McKay, D.B. (2002). Helicase structure and mechanism. Curr Opin Struct Biol *12*, 123-133.

Castella, S., Bingham, G., and Sanders, C.M. (2006). Common determinants in DNA melting and helicase-catalysed DNA unwinding by papillomavirus replication protein E1. Nucleic Acids Res *34*, 3008-3019.

Chen, L., Brugger, K., Skovgaard, M., Redder, P., She, Q., Torarinsson, E., Greve, B., Awayez, M., Zibat, A., Klenk, H.P., and Garrett, R.A. (2005a). The genome of Sulfolobus acidocaldarius, a model organism of the Crenarchaeota. J Bacteriol *187*, 4992-4999.

Chen, Y.J., Yu, X., Kasiviswanathan, R., Shin, J.H., Kelman, Z., and Egelman, E.H. (2005b). Structural polymorphism of Methanothermobacter thermautotrophicus MCM. J Mol Biol *346*, 389-394.

Chen, Y.Z., Zhuang, W., and Prohofsky, E.W. (1992). Energy flow considerations and thermal fluctuational opening of DNA base pairs at a replicating fork: unwinding consistent with observed replication rates. J Biomol Struct Dyn *10*, 415-427.

Coin, F., Auriol, J., Tapias, A., Clivio, P., Vermeulen, W., and Egly, J.M. (2004). Phosphorylation of XPB helicase regulates TFIIH nucleotide excision repair activity. Embo J *23*, 4835-4846.

Coin, F., Bergmann, E., Tremeau-Bravard, A., and Egly, J.M. (1999). Mutations in XPB and XPD helicases found in xeroderma pigmentosum patients impair the transcription function of TFIIH. Embo J *18*, 1357-1366.

Coin, F., Marinoni, J.C., Rodolfo, C., Fribourg, S., Pedrini, A.M., and Egly, J.M. (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*, 184-188.

Coin, F., Oksenych, V., and Egly, J.M. (2007). Distinct roles for the XPB/p52 and XPD/p44 subcomplexes of TFIIH in damaged DNA opening during nucleotide excision repair. Mol Cell *26*, 245-256.

Coin, F., Proietti De Santis, L., Nardo, T., Zlobinskaya, O., Stefanini, M., and Egly, J.M. (2006). p8/TTD-A as a repair-specific TFIIH subunit. Mol Cell *21*, 215-226.

Costa, R.M., Chigancas, V., Galhardo Rda, S., Carvalho, H., and Menck, C.F. (2003). The eukaryotic nucleotide excision repair pathway. Biochimie *85*, 1083-1099.

Cox, M.M., Goodman, M.F., Kreuzer, K.N., Sherratt, D.J., Sandler, S.J., and Marians, K.J. (2000). The importance of repairing stalled replication forks. Nature *404*, 37-41.

Crowley, D.J., Boubriak, I., Berquist, B.R., Clark, M., Richard, E., Sullivan, L., DasSarma, S., and McCready, S. (2006). The uvrA, uvrB and uvrC genes are required for repair of ultraviolet light induced DNA photoproducts in Halobacterium sp. NRC-1. Saline Systems *2*, 11.

Cubeddu, L., and White, M.F. (2005). DNA damage detection by an archaeal single-stranded DNA-binding protein. J Mol Biol *353*, 507-516.

Dangel, A.W., Shen, L., Mendoza, A.R., Wu, L.C., and Yu, C.Y. (1995). Human helicase gene SKI2W in the HLA class III region exhibits striking structural similarities to the yeast antiviral gene SKI2 and to the human gene KIAA0052: emergence of a new gene family. Nucleic Acids Res *23*, 2120-2126.

de Laat, W.L., Appeldoorn, E., Sugasawa, K., Weterings, E., Jaspers, N.G., and Hoeijmakers, J.H. (1998). DNA-binding polarity of human replication protein A positions nucleases in nucleotide excision repair. Genes Dev *12*, 2598-2609.

de Laat, W.L., Jaspers, N.G., and Hoeijmakers, J.H. (1999). Molecular mechanism of nucleotide excision repair. Genes Dev *13*, 768-785.

Della, M., Palmbos, P.L., Tseng, H.M., Tonkin, L.M., Daley, J.M., Topper, L.M., Pitcher, R.S., Tomkinson, A.E., Wilson, T.E., and Doherty, A.J. (2004). Mycobacterial Ku and ligase proteins constitute a two-component NHEJ repair machine. Science *306*, 683-685.

Dennis, C., Fedorov, A., Kas, E., Salome, L., and Grigoriev, M. (2004). RuvAB-directed branch migration of individual Holliday junctions is impeded by sequence heterology. Embo J *23*, 2413-2422.

Dillingham, M.S., Soultanas, P., Wiley, P., Webb, M.R., and Wigley, D.B. (2001). Defining the roles of individual residues in the single-stranded DNA binding site of PcrA helicase. Proc Natl Acad Sci U S A *98*, 8381-8387.

Dionne, I., Nookala, R.K., Jackson, S.P., Doherty, A.J., and Bell, S.D. (2003a). A heterotrimeric PCNA in the hyperthermophilic archaeon Sulfolobus solfataricus. Mol Cell *11*, 275-282.

Dionne, I., Robinson, N.P., McGeoch, A.T., Marsh, V.L., Reddish, A., and Bell, S.D. (2003b). DNA replication in the hyperthermophilic archaeon Sulfolobus solfataricus. Biochem Soc Trans *31*, 674-676.

DiRuggiero, J., Brown, J.R., Bogert, A.P., and Robb, F.T. (1999). DNA repair systems in archaea: mementos from the last universal common ancestor? J Mol Evol *49*, 474-484.

Doherty, A.J., and Jackson, S.P. (2001). DNA repair: how Ku makes ends meet. Curr Biol *11*, R920-924.

Donmez, I., and Patel, S.S. (2006). Mechanisms of a ring shaped helicase. Nucleic Acids Res *34*, 4216-4224.

Dougan, D.A., Mogk, A., Zeth, K., Turgay, K., and Bukau, B. (2002). AAA+ proteins and substrate recognition, it all depends on their partner in crime. FEBS Lett *529*, 6-10.

Dua, R., Levy, D.L., Li, C.M., Snow, P.M., and Campbell, J.L. (2002). In vivo reconstitution of Saccharomyces cerevisiae DNA polymerase epsilon in insect cells. Purification and characterization. J Biol Chem *277*, 7889-7896.

Duggin, I.G., and Bell, S.D. (2006). The chromosome replication machinery of the archaeon Sulfolobus solfataricus. J Biol Chem *281*, 15029-15032.

Dutta, A., and Bell, S.P. (1997). Initiation of DNA replication in eukaryotic cells. Annu Rev Cell Dev Biol *13*, 293-332.

Dvir, A., Conaway, R.C., and Conaway, J.W. (1997). A role for TFIIH in controlling the activity of early RNA polymerase II elongation complexes. Proc Natl Acad Sci U S A *94*, 9006-9010.

Eisen, J.A. (2000). Horizontal gene transfer among microbial genomes: new insights from complete genome analysis. Curr Opin Genet Dev *10*, 606-611.

Enemark, E.J., and Joshua-Tor, L. (2006). Mechanism of DNA translocation in a replicative hexameric helicase. Nature *442*, 270-275.

Erzberger, J.P., and Berger, J.M. (2006). Evolutionary relationships and structural mechanisms of AAA+ proteins. Annu Rev Biophys Biomol Struct *35*, 93-114.

Evans, E., Moggs, J.G., Hwang, J.R., Egly, J.M., and Wood, R.D. (1997). Mechanism of open complex and dual incision formation by human nucleotide excision repair factors. Embo J *16*, 6559-6573.

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

Fishel, R., and Kolodner, R.D. (1995). Identification of mismatch repair genes and their role in the development of cancer. Curr Opin Genet Dev *5*, 382-395.

Fitch, W.M. (2000). Homology a personal view on some of the problems. Trends Genet *16*, 227-231.

Flores, M.J., Bierne, H., Ehrlich, S.D., and Michel, B. (2001). Impairment of lagging strand synthesis triggers the formation of a RuvABC substrate at replication forks. Embo J *20*, 619-629.

Fousteri, M., Vermeulen, W., van Zeeland, A.A., and Mullenders, L.H. (2006). Cockayne syndrome A and B proteins differentially regulate recruitment of chromatin remodeling and repair factors to stalled RNA polymerase II in vivo. Mol Cell *23*, 471-482.

Friedman, R., and Hughes, A.L. (2001). Pattern and timing of gene duplication in animal genomes. Genome Res *11*, 1842-1847.

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

Fujikane, R., Komori, K., Shinagawa, H., and Ishino, Y. (2005). Identification of a novel helicase activity unwinding branched DNAs from the hyperthermophilic archaeon, Pyrococcus furiosus. J Biol Chem *280*, 12351-12358.

Fujikane, R., Shinagawa, H., and Ishino, Y. (2006). The archaeal Hjm helicase has recQ-like functions, and may be involved in repair of stalled replication fork. Genes Cells *11*, 99-110.

Fujita, M., Yamada, C., Goto, H., Yokoyama, N., Kuzushima, K., Inagaki, M., and Tsurumi, T. (1999). Cell cycle regulation of human CDC6 protein. Intracellular localization, interaction with the human mcm complex, and CDC2 kinase-mediated hyperphosphorylation. J Biol Chem *274*, 25927-25932.

Garcia-Vallve, S., Romeu, A., and Palau, J. (2000). Horizontal gene transfer in bacterial and archaeal complete genomes. Genome Res *10*, 1719-1725.

Gardner, M.J., Hall, N., Fung, E., White, O., Berriman, M., Hyman, R.W., Carlton, J.M., Pain, A., Nelson, K.E., Bowman, S., Paulsen, I.T., James, K., Eisen, J.A., Rutherford, K., Salzberg, S.L., Craig, A., Kyes, S., Chan, M.S., Nene, V., Shallom, S.J., Suh, B., Peterson, J., Angiuoli, S., Pertea, M., Allen, J., Selengut, J., Haft, D., Mather, M.W., Vaidya, A.B., Martin, D.M., Fairlamb, A.H., Fraunholz, M.J., Roos, D.S., Ralph, S.A., McFadden, G.I., Cummings, L.M., Subramanian, G.M., Mungall, C., Venter, J.C., Carucci, D.J., Hoffman, S.L., Newbold, C., Davis, R.W., Fraser, C.M., and Barrell, B. (2002). Genome sequence of the human malaria parasite Plasmodium falciparum. Nature *419*, 498-511.

Garg, P., and Burgers, P.M. (2005). DNA polymerases that propagate the eukaryotic DNA replication fork. Crit Rev Biochem Mol Biol *40*, 115-128.

Giglia-Mari, G., Coin, F., Ranish, J.A., Hoogstraten, D., Theil, A., Wijgers, N., Jaspers, N.G., Raams, A., Argentini, M., van der Spek, P.J., Botta, E., Stefanini, M., Egly, J.M., Aebersold, R., Hoeijmakers, J.H., and Vermeulen, W. (2004). A new, tenth subunit of TFIIH is responsible for the DNA repair syndrome trichothiodystrophy group A. Nat Genet *36*, 714-719.

Gomez-Llorente, Y., Fletcher, R.J., Chen, X.S., Carazo, J.M., and San Martin, C. (2005). Polymorphism and double hexamer structure in the archaeal minichromosome maintenance (MCM) helicase from Methanobacterium thermoautotrophicum. J Biol Chem *280*, 40909-40915.

Gorbalenya, A.E., and Koonin, E.V. (1993). Helicases: amino acid sequence comparisons and structure-function relationship. Curr Opin Struct Biol *3*, 419-429.

Gorbalenya, A.E., Koonin, E.V., Donchenko, A.P., and Blinov, V.M. (1989). Two related superfamilies of putative helicases involved in replication, recombination, repair and expression of DNA and RNA genomes. Nucleic Acids Res *17*, 4713-4730.

Gregg, A.V., McGlynn, P., Jaktaji, R.P., and Lloyd, R.G. (2002). Direct rescue of stalled DNA replication forks via the combined action of PriA and RecG helicase activities. Mol Cell *9*, 241-251.

Grogan, D.W. (2000). The question of DNA repair in hyperthermophilic archaea. Trends Microbiol *8*, 180-185.

Gross, A., Columbus, L., Hideg, K., Altenbach, C., and Hubbell, W.L. (1999). Structure of the KcsA potassium channel from Streptomyces lividans: a site-directed spin labeling study of the second transmembrane segment. Biochemistry *38*, 10324-10335.

Gruz, P., Shimizu, M., Pisani, F.M., De Felice, M., Kanke, Y., and Nohmi, T. (2003). Processing of DNA lesions by archaeal DNA polymerases from Sulfolobus solfataricus. Nucleic Acids Res *31*, 4024-4030.

Guy, C.P., and Bolt, E.L. (2005). Archaeal Hel308 helicase targets replication forks in vivo and in vitro and unwinds lagging strands. Nucleic Acids Res *33*, 3678-3690.

Guy, C.P., Haldenby, S., Brindley, A., Walsh, D.A., Briggs, G.S., Warren, M.J., Allers, T., and Bolt, E.L. (2006). Interactions of RadB, a DNA repair protein in archaea, with DNA and ATP. J Mol Biol *358*, 46-56.

Guzder, S.N., Sung, P., Bailly, V., Prakash, L., and Prakash, S. (1994). RAD25 is a DNA helicase required for DNA repair and RNA polymerase II transcription. Nature *369*, 578-581.

Hacker, K.J., and Johnson, K.A. (1997). A hexameric helicase encircles one DNA strand and excludes the other during DNA unwinding. Biochemistry *36*, 14080-14087.

Hall, M.C., and Matson, S.W. (1999). Helicase motifs: the engine that powers DNA unwinding. Mol Microbiol *34*, 867-877.

Harris, P.V., Mazina, O.M., Leonhardt, E.A., Case, R.B., Boyd, J.B., and Burtis, K.C. (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*, 5764-5771.

Heller, R.C., and Marians, K.J. (2005). The disposition of nascent strands at stalled replication forks dictates the pathway of replisome loading during restart. Mol Cell *17*, 733-743.

Henning, K.A., Li, L., Iyer, N., McDaniel, L.D., Reagan, M.S., Legerski, R., Schultz, R.A., Stefanini, M., Lehmann, A.R., Mayne, L.V., and Friedberg, E.C. (1995). The Cockayne syndrome group A gene encodes a WD repeat protein that interacts with CSB protein and a subunit of RNA polymerase II TFIIH. Cell *82*, 555-564.

Hess, M.T., Schwitter, U., Petretta, M., Giese, B., and Naegeli, H. (1997). Bipartite substrate discrimination by human nucleotide excision repair. Proc Natl Acad Sci U S A *94*, 6664-6669.

Hey, T., Lipps, G., and Krauss, G. (2001). Binding of XPA and RPA to damaged DNA investigated by fluorescence anisotropy. Biochemistry *40*, 2901-2910.

Heyer, W.D., Li, X., Rolfsmeier, M., and Zhang, X.P. (2006). Rad54: the Swiss Army knife of homologous recombination? Nucleic Acids Res.

Hiasa, H., and Marians, K.J. (1994). Primase couples leading- and lagging-strand DNA synthesis from oriC. J Biol Chem *269*, 6058-6063.

Hoeijmakers, J.H. (2001). Genome maintenance mechanisms for preventing cancer. Nature *411*, 366-374.

Hoeijmakers, J.H., and Bootsma, D. (1994). DNA repair. Incisions for excision. Nature *371*, 654-655.

Huber, H., Hohn, M.J., Rachel, R., Fuchs, T., Wimmer, V.C., and Stetter, K.O. (2002). A new phylum of Archaea represented by a nanosized hyperthermophilic symbiont. Nature *417*, 63-67.

Hurles, M. (2004). Gene duplication: the genomic trade in spare parts. PLoS Biol *2*, E206.

Jain, R., Rivera, M.C., and Lake, J.A. (1999). Horizontal gene transfer among genomes: the complexity hypothesis. Proc Natl Acad Sci U S A *96*, 3801-3806.

James, J.A., Aggarwal, A.K., Linden, R.M., and Escalante, C.R. (2004). Structure of adeno-associated virus type 2 Rep40-ADP complex: insight into nucleotide recognition and catalysis by superfamily 3 helicases. Proc Natl Acad Sci U S A *101*, 12455-12460.

Jawhari, A., Laine, J.P., Dubaele, S., Lamour, V., Poterszman, A., Coin, F., Moras, D., and Egly, J.M. (2002). p52 Mediates XPB function within the transcription/repair factor TFIIH. J Biol Chem *277*, 31761-31767.

Jeong, Y.J., Levin, M.K., and Patel, S.S. (2004). The DNA-unwinding mechanism of the ring helicase of bacteriophage T7. Proc Natl Acad Sci U S A *101*, 7264-7269.

Johnson, A., and O'Donnell, M. (2005). Cellular DNA replicases: components and dynamics at the replication fork. Annu Rev Biochem *74*, 283-315.

Karata, K., Inagawa, T., Wilkinson, A.J., Tatsuta, T., and Ogura, T. (1999). Dissecting the role of a conserved motif (the second region of homology) in the AAA family of ATPases. Site-directed mutagenesis of the ATP-dependent protease FtsH. J Biol Chem *274*, 26225-26232.

Kawaoka, J., Jankowsky, E., and Pyle, A.M. (2004). Backbone tracking by the SF2 helicase NPH-II. Nat Struct Mol Biol *11*, 526-530.

Kelly, S.M., Jess, T.J., and Price, N.C. (2005). How to study proteins by circular dichroism. Biochim Biophys Acta *1751*, 119-139.

Kelly, S.M., and Price, N.C. (1997). The application of circular dichroism to studies of protein folding and unfolding. Biochim Biophys Acta *1338*, 161-185.

Kelman, Z., and White, M.F. (2005). Archaeal DNA replication and repair. Curr Opin Microbiol *8*, 669-676.

Kerr, I.D., Wadsworth, R.I., Cubeddu, L., Blankenfeldt, W., Naismith, J.H., and White, M.F. (2003). Insights into ssDNA recognition by the OB fold from a structural and thermodynamic study of Sulfolobus SSB protein. Embo J *22*, 2561-2570.

Kim, J.L., Morgenstern, K.A., Griffith, J.P., Dwyer, M.D., Thomson, J.A., Murcko, M.A., Lin, C., and Caron, P.R. (1998). Hepatitis C virus NS3 RNA helicase domain with a bound oligonucleotide: the crystal structure provides insights into the mode of unwinding. Structure *6*, 89-100.

Kim, J.S., Krasieva, T.B., Kurumizaka, H., Chen, D.J., Taylor, A.M., and Yokomori, K. (2005). Independent and sequential recruitment of NHEJ and HR factors to DNA damage sites in mammalian cells. J Cell Biol *170*, 341-347.

Kim, T.K., Ebright, R.H., and Reinberg, D. (2000). Mechanism of ATP-dependent promoter melting by transcription factor IIH. Science *288*, 1418-1422.

Komori, K., Miyata, T., DiRuggiero, J., Holley-Shanks, R., Hayashi, I., Cann, I.K., Mayanagi, K., Shinagawa, H., and Ishino, Y. (2000). Both RadA and RadB are involved in homologous recombination in Pyrococcus furiosus. J Biol Chem *275*, 33782-33790.

Konneke, M., Bernhard, A.E., de la Torre, J.R., Walker, C.B., Waterbury, J.B., and Stahl, D.A. (2005). Isolation of an autotrophic ammonia-oxidizing marine archaeon. Nature *437*, 543-546.

Koonin, E.V. (1993). A common set of conserved motifs in a vast variety of putative nucleic acid-dependent ATPases including MCM proteins involved in the initiation of eukaryotic DNA replication. Nucleic Acids Res *21*, 2541-2547.

Korolev, S., Hsieh, J., Gauss, G.H., Lohman, T.M., and Waksman, G. (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*, 635-647.

Krejci, L., Chen, L., Van Komen, S., Sung, P., and Tomkinson, A. (2003). Mending the break: two DNA double-strand break repair machines in eukaryotes. Prog Nucleic Acid Res Mol Biol *74*, 159-201.

Kuhl, P.W. (1994). Excess-substrate inhibition in enzymology and high-dose inhibition in pharmacology: a reinterpretation [corrected]. Biochem J *298 (Pt 1)*, 171-180.

Kulaksiz, G., Reardon, J.T., and Sancar, A. (2005). Xeroderma pigmentosum complementation group E protein (XPE/DDB2): purification of various complexes of XPE and analyses of their damaged DNA binding and putative DNA repair properties. Mol Cell Biol *25*, 9784-9792.

Kvaratskhelia, M., and White, M.F. (2000). Two Holliday junction resolving enzymes in Sulfolobus solfataricus. J Mol Biol *297*, 923-932.

Laine, J.P., Mocquet, V., and Egly, J.M. (2006). TFIIH enzymatic activities in transcription and nucleotide excision repair. Methods Enzymol *408*, 246-263.

Lakowicz, J.R. (1983). Principles of Fluorescence Spectroscopy, Plenum Press, New York.

LeBowitz, J.H., and McMacken, R. (1986). The Escherichia coli dnaB replication protein is a DNA helicase. J Biol Chem *261*, 4738-4748.

Lee, J.Y., and Yang, W. (2006). UvrD helicase unwinds DNA one base pair at a time by a two-part power stroke. Cell *127*, 1349-1360.

Liao, J.C., Jeong, Y.J., Kim, D.E., Patel, S.S., and Oster, G. (2005). Mechanochemistry of t7 DNA helicase. J Mol Biol *350*, 452-475.

Lin, Y.C., Choi, W.S., and Gralla, J.D. (2005). TFIIH XPB mutants suggest a unified bacterial-like mechanism for promoter opening but not escape. Nat Struct Mol Biol *12*, 603-607.

Lin, Y.C., and Gralla, J.D. (2005). Stimulation of the XPB ATP-dependent helicase by the beta subunit of TFIIE. Nucleic Acids Res *33*, 3072-3081.

Lindahl, T., and Wood, R.D. (1999). Quality control by DNA repair. Science *286*, 1897-1905.

Lohman, T.M., and Bjornson, K.P. (1996). Mechanisms of helicase-catalyzed DNA unwinding. Annu Rev Biochem *65*, 169-214.

Lundblad, J.R., Laurance, M., and Goodman, R.H. (1996). Fluorescence polarization analysis of protein-DNA and protein-protein interactions. Mol Endocrinol *10*, 607-612.

Luscombe, N.M., Austin, S.E., Berman, H.M., and Thornton, J.M. (2000). An overview of the structures of protein-DNA complexes. Genome Biol *1*, REVIEWS001.

Mackintosh, S.G., and Raney, K.D. (2006). DNA unwinding and protein displacement by superfamily 1 and superfamily 2 helicases. Nucleic Acids Res *34*, 4154-4159.

Malta, E., Moolenaar, G.F., and Goosen, N. (2006). Base flipping in nucleotide excision repair. J Biol Chem *281*, 2184-2194.

Malta, E., Moolenaar, G.F., and Goosen, N. (2007). Dynamics of the UvrABC nucleotide excision repair proteins analyzed by fluorescence resonance energy transfer. Biochemistry *46*, 9080-9088.

Marini, F., and Wood, R.D. (2002). A human DNA helicase homologous to the DNA cross-link sensitivity protein Mus308. J Biol Chem *277*, 8716-8723.

Martins-Pinheiro, M., Galhardo, R.S., Lage, C., Lima-Bessa, K.M., Aires, K.A., and Menck, C.F. (2004). Different patterns of evolution for duplicated DNA repair genes in bacteria of the Xanthomonadales group. BMC Evol Biol *4*, 29.

Mazina, O.M., and Mazin, A.V. (2004). Human Rad54 protein stimulates DNA strand exchange activity of hRad51 protein in the presence of Ca2+. J Biol Chem *279*, 52042-52051.

McCaffrey, R., St Johnston, D., and Gonzalez-Reyes, A. (2006). Drosophila mus301/spindle-C encodes a helicase with an essential role in double-strand DNA break repair and meiotic progression. Genetics *174*, 1273-1285.

McDougal, V.V., and Guarino, L.A. (2001). DNA and ATP binding activities of the baculovirus DNA helicase P143. J Virol *75*, 7206-7209.

McGlynn, P., and Lloyd, R.G. (2000). Modulation of RNA polymerase by (p)ppGpp reveals a RecG-dependent mechanism for replication fork progression. Cell *101*, 35-45.

McGlynn, P., and Lloyd, R.G. (2002). Recombinational repair and restart of damaged replication forks. Nat Rev Mol Cell Biol *3*, 859-870.

McGlynn, P., Lloyd, R.G., and Marians, K.J. (2001). Formation of Holliday junctions by regression of nascent DNA in intermediates containing stalled replication forks: RecG stimulates regression even when the DNA is negatively supercoiled. Proc Natl Acad Sci U S A *98*, 8235-8240.

McHenry, C.S. (2003). Chromosomal replicases as asymmetric dimers: studies of subunit arrangement and functional consequences. Mol Microbiol *49*, 1157-1165.

Mellon, I. (2005). Transcription-coupled repair: a complex affair. Mutat Res *577*, 155-161.

Moreland, R.J., Tirode, F., Yan, Q., Conaway, J.W., Egly, J.M., and Conaway, R.C. (1999). A role for the TFIIH XPB DNA helicase in promoter escape by RNA polymerase II. J Biol Chem *274*, 22127-22130.

Morimatsu, K., and Kowalczykowski, S.C. (2003). RecFOR proteins load RecA protein onto gapped DNA to accelerate DNA strand exchange: a universal step of recombinational repair. Mol Cell *11*, 1337-1347.

Morris, P.D., and Raney, K.D. (1999). DNA helicases displace streptavidin from biotin-labeled oligonucleotides. Biochemistry *38*, 5164-5171.

Morris, P.D., Tackett, A.J., Babb, K., Nanduri, B., Chick, C., Scott, J., and Raney, K.D. (2001). Evidence for a functional monomeric form of the bacteriophage T4 DdA helicase. Dda does not form stable oligomeric structures. J Biol Chem *276*, 19691-19698.

Mott, M.L., and Berger, J.M. (2007). DNA replication initiation: mechanisms and regulation in bacteria. Nat Rev Microbiol *5*, 343-354.

Nick McElhinny, S.A., Snowden, C.M., McCarville, J., and Ramsden, D.A. (2000). Ku recruits the XRCC4-ligase IV complex to DNA ends. Mol Cell Biol *20*, 2996-3003.

O'Donnell, M.E., Elias, P., and Lehman, I.R. (1987). Processive replication of single-stranded DNA templates by the herpes simplex virus-induced DNA polymerase. J Biol Chem *262*, 4252-4259.

Obmolova, G., Ban, C., Hsieh, P., and Yang, W. (2000). Crystal structures of mismatch repair protein MutS and its complex with a substrate DNA. Nature *407*, 703-710.

Ochman, H., Lawrence, J.G., and Groisman, E.A. (2000). Lateral gene transfer and the nature of bacterial innovation. Nature *405*, 299-304.

Ogrunc, M., Becker, D.F., Ragsdale, S.W., and Sancar, A. (1998). Nucleotide excision repair in the third kingdom. J Bacteriol *180*, 5796-5798.

Ogura, T., and Wilkinson, A.J. (2001). AAA+ superfamily ATPases: common structure--diverse function. Genes Cells *6*, 575-597.

Olsen, G.J., and Woese, C.R. (1997). Archaeal genomics: an overview. Cell *89*, 991-994.

Opresko, P.L., Cheng, W.H., and Bohr, V.A. (2004). Junction of RecQ helicase biochemistry and human disease. J Biol Chem *279*, 18099-18102.

Orren, D.K., and Sancar, A. (1990). Formation and enzymatic properties of the UvrB.DNA complex. J Biol Chem *265*, 15796-15803.

Panyutin, I.G., and Hsieh, P. (1994). The kinetics of spontaneous DNA branch migration. Proc Natl Acad Sci U S A *91*, 2021-2025.

Park, C.H., Bessho, T., Matsunaga, T., and Sancar, A. (1995). Purification and characterization of the XPF-ERCC1 complex of human DNA repair excision nuclease. J Biol Chem *270*, 22657-22660.

Park, J.S., Marr, M.T., and Roberts, J.W. (2002). E. coli Transcription repair coupling factor (Mfd protein) rescues arrested complexes by promoting forward translocation. Cell *109*, 757-767.

Pastwa, E., and Blasiak, J. (2003). Non-homologous DNA end joining. Acta Biochim Pol *50*, 891-908.

Pause, A., and Sonenberg, N. (1992). Mutational analysis of a DEAD box RNA helicase: the mammalian translation initiation factor eIF-4A. Embo J *11*, 2643-2654.

Perkins, G., and Diffley, J.F. (1998). Nucleotide-dependent prereplicative complex assembly by Cdc6p, a homolog of eukaryotic and prokaryotic clamp-loaders. Mol Cell *2*, 23-32.

Petukhova, G., Stratton, S., and Sung, P. (1998). Catalysis of homologous DNA pairing by yeast Rad51 and Rad54 proteins. Nature *393*, 91-94.

Pitcher, R.S., Tonkin, L.M., Green, A.J., and Doherty, A.J. (2005a). Domain structure of a NHEJ DNA repair ligase from Mycobacterium tuberculosis. J Mol Biol *351*, 531-544.

Pitcher, R.S., Wilson, T.E., and Doherty, A.J. (2005b). New insights into NHEJ repair processes in prokaryotes. Cell Cycle *4*, 675-678.

Pufall, M.A., and Graves, B.J. (2002). Autoinhibitory domains: modular effectors of cellular regulation. Annu Rev Cell Dev Biol *18*, 421-462.

Ramsden, D.A., and Gellert, M. (1998). Ku protein stimulates DNA end joining by mammalian DNA ligases: a direct role for Ku in repair of DNA double-strand breaks. Embo J *17*, 609-614.

Rashid, N., Morikawa, M., and Imanaka, T. (1996). A RecA/RAD51 homologue from a hyperthermophilic archaeon retains the major RecA domain only. Mol Gen Genet *253*, 397-400.

Reid, S.L., Parry, D., Liu, H.H., and Connolly, B.A. (2001). Binding and recognition of GATATC target sequences by the EcoRV restriction endonuclease: a study using fluorescent oligonucleotides and fluorescence polarization. Biochemistry *40*, 2484-2494.

Richards, J.D., Johnson, K.A., Liu, H., McMahon, S., Oke, M., Carter, L., Naismith, J.H., and White, M.F. (2007). Structure of the DNA repair helicase Hel308 reveals DNA binding and autoinhibitory domains. Journal of Biological Chemistry (In press).

Riedl, T., Hanaoka, F., and Egly, J.M. (2003). The comings and goings of nucleotide excision repair factors on damaged DNA. Embo J *22*, 5293-5303.

Roberts, J.A. (2004). Characterisation of Sulfolobus solfataricus XPF An archaeal DNA repair endonuclease. PhD Thesis, University of St Andrews.

Roberts, J.A., Bell, S.D., and White, M.F. (2003). An archaeal XPF repair endonuclease dependent on a heterotrimeric PCNA. Mol Microbiol *48*, 361-371.

Roberts, J.A., and White, M.F. (2005). An archaeal endonuclease displays key properties of both eukaryal XPF-ERCC1 and Mus81. J Biol Chem *280*, 5924-5928.

Robinson, N.P., and Bell, S.D. (2005). Origins of DNA replication in the three domains of life. Febs J *272*, 3757-3766.

Rossi, M., Ciaramella, M., Cannio, R., Pisani, F.M., Moracci, M., and Bartolucci, S. (2003). Extremophiles 2002. J Bacteriol *185*, 3683-3689.

Rudolf, J. (2006). Characterisation of XPD from *Sulfolous acidocaldarius*: an Iron-Sulphur Cluster Containing DNA Repair Helicase. PhD Thesis, University of St Andrews.

Rudolf, J., Makrantoni, V., Ingledew, W.J., Stark, M.J., and White, M.F. (2006). The DNA repair helicases XPD and FancJ have essential iron-sulfur domains. Mol Cell *23*, 801-808.

Sandler, S.J., Satin, L.H., Samra, H.S., and Clark, A.J. (1996). recA-like genes from three archaean species with putative protein products similar to Rad51 and Dmc1 proteins of the yeast Saccharomyces cerevisiae. Nucleic Acids Res *24*, 2125-2132.

Schefe, J.H., Lehmann, K.E., Buschmann, I.R., Unger, T., and Funke-Kaiser, H. (2006). Quantitative real-time RT-PCR data analysis: current concepts and the novel "gene expression's CT difference" formula. J Mol Med *84*, 901-910.

Schirmer, E.C., Glover, J.R., Singer, M.A., and Lindquist, S. (1996). HSP100/Clp proteins: a common mechanism explains diverse functions. Trends Biochem Sci *21*, 289-296.

Schmidt, A., Kellermann, J., and Lottspeich, F. (2005). A novel strategy for quantitative proteomics using isotope-coded protein labels. Proteomics *5*, 4-15.

Seeberg, E., Eide, L., and Bjoras, M. (1995). The base excision repair pathway. Trends Biochem Sci *20*, 391-397.

Seitz, E.M., Haseltine, C.A., and Kowalczykowski, S.C. (2001). DNA recombination and repair in the archaea. Adv Appl Microbiol *50*, 101-169.

Seki, M., Marini, F., and Wood, R.D. (2003). POLQ (Pol theta), a DNA polymerase and DNA-dependent ATPase in human cells. Nucleic Acids Res *31*, 6117-6126.

Sharma, S., Doherty, K.M., and Brosh, R.M., Jr. (2006). Mechanisms of RecQ helicases in pathways of DNA metabolism and maintenance of genomic stability. Biochem J *398*, 319-337.

She, Q., Singh, R.K., Confalonieri, F., Zivanovic, Y., Allard, G., Awayez, M.J., Chan-Weiher, C.C., Clausen, I.G., Curtis, B.A., De Moors, A., Erauso, G., Fletcher, C., Gordon, P.M., Heikamp-de Jong, I., Jeffries, A.C., Kozera, C.J., Medina, N., Peng, X., Thi-Ngoc, H.P., Redder, P., Schenk, M.E., Theriault, C., Tolstrup, N., Charlebois, R.L., Doolittle, W.F., Duguet, M., Gaasterland, T., Garrett, R.A., Ragan, M.A., Sensen, C.W., and Van der Oost, J. (2001). The complete genome of the crenarchaeon Sulfolobus solfataricus P2. Proc Natl Acad Sci U S A *98*, 7835-7840.

Shuttleworth, G., Fogg, M.J., Kurpiewski, M.R., Jen-Jacobson, L., and Connolly, B.A. (2004). Recognition of the pro-mutagenic base uracil by family B DNA polymerases from archaea. J Mol Biol *337*, 621-634.

Sijbers, A.M., de Laat, W.L., Ariza, R.R., Biggerstaff, M., Wei, Y.F., Moggs, J.G., Carter, K.C., Shell, B.K., Evans, E., de Jong, M.C., Rademakers, S., de Rooij, J., Jaspers, N.G., Hoeijmakers, J.H., and Wood, R.D. (1996). Xeroderma pigmentosum group F caused by a defect in a structure-specific DNA repair endonuclease. Cell *86*, 811-822.

Singleton, M.R., Dillingham, M.S., and Wigley, D.B. (2007). Structure and mechanism of helicases and nucleic acid translocases. Annu Rev Biochem *76*, 23-50.

Singleton, M.R., Sawaya, M.R., Ellenberger, T., and Wigley, D.B. (2000). Crystal structure of T7 gene 4 ring helicase indicates a mechanism for sequential hydrolysis of nucleotides. Cell *101*, 589-600.

Singleton, M.R., and Wigley, D.B. (2002). Modularity and specialization in superfamily 1 and 2 helicases. J Bacteriol *184*, 1819-1826.

Sladek, F.M., Munn, M.M., Rupp, W.D., and Howard-Flanders, P. (1989). In vitro repair of psoralen-DNA cross-links by RecA, UvrABC, and the 5'-exonuclease of DNA polymerase I. J Biol Chem *264*, 6755-6765.

Smith, A.J., Szczelkun, M.D., and Savery, N.J. (2007). Controlling the motor activity of a transcription-repair coupling factor: autoinhibition and the role of RNA polymerase. Nucleic Acids Res *35*, 1802-1811.

Steinhoff, H.J. (2002). Methods for study of protein dynamics and protein-protein interaction in protein-ubiquitination by electron paramagnetic resonance spectroscopy. Front Biosci *7*, c97-110.

Stephanou, N.C., Gao, F., Bongiorno, P., Ehrt, S., Schnappinger, D., Shuman, S., and Glickman, M.S. (2007). Mycobacterial nonhomologous end joining mediates mutagenic repair of chromosomal double-strand DNA breaks. J Bacteriol *189*, 5237-5246.

Story, R.M., Weber, I.T., and Steitz, T.A. (1992). The structure of the E. coli recA protein monomer and polymer. Nature *355*, 318-325.

Stratagene. (2005). Methods and Application Guide, Introduction to Quantitative PCR.

Studier, F.W. (2005). Protein production by auto-induction in high density shaking cultures. Protein Expr Purif *41*, 207-234.

Sugasawa, K., Ng, J.M., Masutani, C., Iwai, S., van der Spek, P.J., Eker, A.P., Hanaoka, F., Bootsma, D., and Hoeijmakers, J.H. (1998). Xeroderma pigmentosum group C protein complex is the initiator of global genome nucleotide excision repair. Mol Cell *2*, 223-232.

Sung, P., Higgins, D., Prakash, L., and Prakash, S. (1988). Mutation of lysine-48 to arginine in the yeast RAD3 protein abolishes its ATPase and DNA helicase activities but not the ability to bind ATP. Embo J *7*, 3263-3269.

Svoboda, D.L., Taylor, J.S., Hearst, J.E., and Sancar, A. (1993). DNA repair by eukaryotic nucleotide excision nuclease. Removal of thymine dimer and psoralen monoadduct by HeLa cell-free extract and of thymine dimer by Xenopus laevis oocytes. J Biol Chem *268*, 1931-1936.

Tada, S. (2007). Cdt1 and geminin: role during cell cycle progression and DNA damage in higher eukaryotes. Front Biosci *12*, 1629-1641.

Tapias, A., Auriol, J., Forget, D., Enzlin, J.H., Scharer, O.D., Coin, F., Coulombe, B., and Egly, J.M. (2004). Ordered conformational changes in damaged DNA induced by nucleotide excision repair factors. J Biol Chem *279*, 19074-19083.

Theis, K., Chen, P.J., Skorvaga, M., Van Houten, B., and Kisker, C. (1999). Crystal structure of UvrB, a DNA helicase adapted for nucleotide excision repair. Embo J *18*, 6899-6907.

Theis, K., Skorvaga, M., Machius, M., Nakagawa, N., Van Houten, B., and Kisker, C. (2000). The nucleotide excision repair protein UvrB, a helicase-like enzyme with a catch. Mutat Res *460*, 277-300.

Thornton, J.W., and DeSalle, R. (2000). Gene family evolution and homology: genomics meets phylogenetics. Annu Rev Genomics Hum Genet *1*, 41-73.

Tornaletti, S., Reines, D., and Hanawalt, P.C. (1999). Structural characterization of RNA polymerase II complexes arrested by a cyclobutane pyrimidine dimer in the transcribed strand of template DNA. J Biol Chem *274*, 24124-24130.

Tuteja, N., and Tuteja, R. (2004). Unraveling DNA helicases. Motif, structure, mechanism and function. Eur J Biochem *271*, 1849-1863.

Umezu, K., Chi, N.W., and Kolodner, R.D. (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*, 3875-3879.

Vale, R.D. (2000). AAA proteins. Lords of the ring. J Cell Biol *150*, F13-19.

van Hoffen, A., Balajee, A.S., van Zeeland, A.A., and Mullenders, L.H. (2003). Nucleotide excision repair and its interplay with transcription. Toxicology *193*, 79-90.

Van Houten, B. (1990). Nucleotide excision repair in Escherichia coli. Microbiol Rev *54*, 18-51.

Vasquez, K.M., Christensen, J., Li, L., Finch, R.A., and Glazer, P.M. (2002). Human XPA and RPA DNA repair proteins participate in specific recognition of triplex-induced helical distortions. Proc Natl Acad Sci U S A *99*, 5848-5853.

Veaute, X., Delmas, S., Selva, M., Jeusset, J., Le Cam, E., Matic, I., Fabre, F., and Petit, M.A. (2005). UvrD helicase, unlike Rep helicase, dismantles RecA nucleoprotein filaments in Escherichia coli. Embo J *24*, 180-189.

Velankar, S.S., Soultanas, P., Dillingham, M.S., Subramanya, H.S., and Wigley, D.B. (1999). Crystal structures of complexes of PcrA DNA helicase with a DNA substrate indicate an inchworm mechanism. Cell *97*, 75-84.

Verhoeven, E.E., van Kesteren, M., Moolenaar, G.F., Visse, R., and Goosen, N. (2000). Catalytic sites for 3' and 5' incision of Escherichia coli nucleotide excision repair are both located in UvrC. J Biol Chem *275*, 5120-5123.

Verhoeven, E.E., Wyman, C., Moolenaar, G.F., and Goosen, N. (2002). The presence of two UvrB subunits in the UvrAB complex ensures damage detection in both DNA strands. Embo J *21*, 4196-4205.

Vindigni, A. (2007). Biochemical, biophysical, and proteomic approaches to study DNA helicases. Mol Biosyst *3*, 266-274.

Volker, M., Mone, M.J., Karmakar, P., van Hoffen, A., Schul, W., Vermeulen, W., Hoeijmakers, J.H., van Driel, R., van Zeeland, A.A., and Mullenders, L.H. (2001). Sequential assembly of the nucleotide excision repair factors in vivo. Mol Cell *8*, 213-224.

Voloshin, O.N., and Camerini-Otero, R.D. (2007). The DinG protein from Escherichia coli is a structure-specific helicase. J Biol Chem *282*, 18437-18447.

Wadsworth, R.I., and White, M.F. (2001). Identification and properties of the crenarchaeal single-stranded DNA binding protein from Sulfolobus solfataricus. Nucleic Acids Res *29*, 914-920.

Wang, H., DellaVecchia, M.J., Skorvaga, M., Croteau, D.L., Erie, D.A., and Van Houten, B. (2006). UvrB domain 4, an autoinhibitory gate for regulation of DNA binding and ATPase activity. J Biol Chem *281*, 15227-15237.

Weeda, G., van Ham, R.C., Vermeulen, W., Bootsma, D., van der Eb, A.J., and Hoeijmakers, J.H. (1990). A presumed DNA helicase encoded by ERCC-3 is involved in the human repair disorders xeroderma pigmentosum and Cockayne's syndrome. Cell *62*, 777-791.

Weinreich, M., Liang, C., and Stillman, B. (1999). The Cdc6p nucleotide-binding motif is required for loading mcm proteins onto chromatin. Proc Natl Acad Sci U S A *96*, 441-446.

White, M.F. (2003). Archaeal DNA repair: paradigms and puzzles. Biochem Soc Trans *31*, 690-693.

Wilkinson, S.K. (2007). Isotope-coded protein labelling and enrichment of phosphorylated peptides in proteomics. Masters Degree Thesis, University of St Andrews.

Winkler, G.S., Araujo, S.J., Fiedler, U., Vermeulen, W., Coin, F., Egly, J.M., Hoeijmakers, J.H., Wood, R.D., Timmers, H.T., and Weeda, G. (2000). TFIIH with inactive XPD helicase functions in transcription initiation but is defective in DNA repair. J Biol Chem *275*, 4258-4266.

Woese, C.R., and Fox, G.E. (1977). Phylogenetic structure of the prokaryotic domain: the primary kingdoms. Proc Natl Acad Sci U S A *74*, 5088-5090.

Woese, C.R., Kandler, O., and Wheelis, M.L. (1990). Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. Proc Natl Acad Sci U S A *87*, 4576-4579.

Wong, I., Chao, K.L., Bujalowski, W., and Lohman, T.M. (1992). DNA-induced dimerization of the Escherichia coli rep helicase. Allosteric effects of single-stranded and duplex DNA. J Biol Chem *267*, 7596-7610.

Wood, R.D. (1997). Nucleotide excision repair in mammalian cells. J Biol Chem *272*, 23465-23468.

Yang, W. (2000). Structure and function of mismatch repair proteins. Mutat Res *460*, 245-256.

Yarranton, G.T., and Gefter, M.L. (1979). Enzyme-catalyzed DNA unwinding: studies on Escherichia coli rep protein. Proc Natl Acad Sci U S A *76*, 1658-1662.

Ye, J., Osborne, A.R., Groll, M., and Rapoport, T.A. (2004). RecA-like motor ATPases--lessons from structures. Biochim Biophys Acta *1659*, 1-18.

Yong, Y., and Romano, L.J. (1995). Nucleotide and DNA-induced conformational changes in the bacteriophage T7 gene 4 protein. J Biol Chem *270*, 24509-24517.

Yoon-Robarts, M., Blouin, A.G., Bleker, S., Kleinschmidt, J.A., Aggarwal, A.K., Escalante, C.R., and Linden, R.M. (2004). Residues within the B' motif are critical for DNA binding by the superfamily 3 helicase Rep40 of adeno-associated virus type 2. J Biol Chem *279*, 50472-50481.

Zechner, E.L., Wu, C.A., and Marians, K.J. (1992). Coordinated leading- and lagging-strand synthesis at the Escherichia coli DNA replication fork. II. Frequency of primer synthesis and efficiency of primer utilization control Okazaki fragment size. J Biol Chem *267*, 4045-4053.

Zhang, L., Gaut, B.S., and Vision, T.J. (2001). Gene duplication and evolution. Science *293*, 1551.

Zou, Y., and Van Houten, B. (1999). Strand opening by the UvrA(2)B complex allows dynamic recognition of DNA damage. Embo J *18*, 4889-4901.

A1 Primer sequences

A1.1 Gateway primers:

The TEV cleavage site is highlighted in red; the 5' recombination site (*att*B) on the megaprimer is shown in green; the 3' recombination site (*att*B) is highlighted in blue; and the sequence for the polyhistidine tag is in bold font. The portion that is complementary to the gene sequence is underlined.

A1.2 All primers

Table 6. Primer sequences for gene amplification, mutagenesis and RT PCR.

Name	Primer sequence 5' to 3'
XPB1	CGTCGGATCCCCATGGCCTCGAGGACTTTCTATATCAAACAATGG
XPB1-r	CCGGGGATCCGTCGACTTAATCTCTGTCGATATTTTCGGAGG
XPB1 K192A	CCGACAGGCGCTGGAGCGACAGTAGTTGG
XPB1 K192A-r	Complement of 5' sequence
XPB2	CGTCGGATCCCCATGGTACGATTACGATACTTTAAATG
XPB2-r	CCGGGGATCCGTCGACTTAGAGGTTAGCATTTTGTCTTCGCC
XPB2 K96A	CTGGAGGTGGAGCAACTGTAATAGG
XPB2 K96A-r	Complement of 5' sequence
Sso0475 (Bax1) pET28c	CGTCGGATCCCCATGGTAACCTCTGATCTAGCTAGG
Sso0475-r	CCGGGGATCCGTCGACTTAAACCTCTTTGATCTTTTCTAC
Sso0475 (Bax1) pET Duet	GCCAACTCGCCATATGCTAACCTCTGATCTAGCTAGG
Sso0475-r	GCGCACTGGACGTCTTAAACCTCTTTGATCTTTTCTACAATTGCATC
Sso1468 (Gateway)	CCGAAAACCTGTATTTTCAGGGCATGTATATTGAAGAAATAAAGAGA GTTCTTAG
Sso1468-r (Gateway)	GGGGACCACTTTGTACAAGAAAGCTGGGTCCTATTCAAGAAGGAACC TGTATAGCG

Universal Megaprimer (Gateway)	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCGAAGGAGATATACATATGTCGTACTAC**CATCACCATCACCATCAC**GATTACGATATCCCAACGACCGAAAACCTGTATTTTCAGGGC
Sso1468 K56A	GAAGATCCGGGGCATCGACACTTTC
Sso1468 K56A-r	Complement of 5' sequence
Sac1656 (Gateway)	CCGAAAACCTGTATTTTCAGGGCATGTTACCGTGGGAATTAGCGAGATT
Sac1656-r (Gateway)	GGGGACCACTTTGTACAAGAAAGCTGGGTCCTATTTCTTTACTACGACAG CGTCAG
PBL2025 Hel308 (Gateway)	CCGAAAACCTGTATTTTCAGGGCATGAGTTTAGAATTAGAGTGGATGTC
PBL2025 Hel308-r (Gateway)	GGGGACCACTTTGTACAAGAAAGCTGGGTCCTAATGAAATCTATTAAGTA ATCTTG
PBL2025 Hel308 K52A	GGCTCCGGGACGACACTAATAGCC
PBL2025 Hel308 K52A-r	Complement of 5' sequence
PBL2025 Hel308 R155A	GACCCCGAAGCGGGTCCAGTG
PBL2025 Hel308 R155A-r	Complement of 5' sequence
PBL2025 Hel308 R255A	AGGATTTCTGCGAAAATGGCTGAG
PBL2025 Hel308 R255A-r	Complement of 5' sequence
PBL2025 Hel308 R320A	AAGGCATTAGCGGATTTAATAGAGGAAGGA
PBL2025 Hel308 R320A-r	Complement of 5' sequence

PBL2025 Hel308 K646STOP	GTGAGAGATGGTATATAGGAAGAGCTA
PBL2025 Hel308 K646STOP-r	Complement of 5' sequence
PBL2025 Hel308 R662A	GTGGGTAGGAAGGCGGCTAGACTATTA
PBL2025 Hel308 R662A-r	Complement of 5' sequence
PBL2025 Hel308 Domain 5	CACCGGTATAAAGGAAGAGCTATTG
PBL2025 Hel308 Domain 5-r	CTAATGAAATCTATTAAGTAATCTTGC
PBL2025 Hel308 internal sequencing primer	GATCCTGAAGCAGGGCCAGTA
PBL2025 Hel308 internal sequencing primer-r	CAAATCTCCACAACCTATGTT
RT PCR XPB1	TGAATGCAGGGGTTCTTGTT
RT PCR XPB1-r	AGTTTTGCTTGCTTGCCATT
RT PCR XPB2	AGTGATTCCCACTGGAGGTG
RT PCR XPB2-r	CCGCCTCAACTCCAAGTAAC

A2 Substrate oligonucleotide sequences

Table 7. The sequences of the oligonucleotides used to construct DNA substrates.

Name	Oligonucleotide sequence 5' to 3'
10	CTTTAGCTGCATATTTACAACATGTTGACCTACAGCACCAGATTCAG<u>**C**</u>AATTAAGCTCTAAGCCATCCGCAAAAATGACCTCTTATCAAAAGGA
11M	TCCTTTTGATAAGAGGTCATTTTTGCGGATGGCTTAGAGCTTAATTA<u>**T**</u>AAAATCTGGTGCTGTAGGTCAACATGTTGTAAATATGCAGCTAAAG
21	AGGCAACTGAGCCGTAATTCACACCGCATATCGAGCACTGACTGAACCTGAATCTGGTGCTGTAGGTCAACATGTTGTAAATATGCAGCTAAAG
22	CTTTAGCTGCATATTTACAACATGTTGACCTACAGCACCAGATTTAGGTTCAGTCAGTGCTCGATATGCGGTGTGAATTACGGCTCAGTTGCTT
J150B	CCTCGAGGGATCCGTCCTAGCAAGCCGCTGCTACCGGAAGCTTTGGACC
J150H	GGTCCAGAAGCTTCCGGTAGCAGCGAGAGCGGTGGTTGAATTCCTCGACG
J150R	CGTCGAGGAATTCAACCACCGCTCTTCTCAACTGCAGTCTAGACTCGAGC
J150X	GCTCGAGTCTAGACTGCAGTTGAGAGCTTGCTAGGACGGATCCCTCGAGG
Fluorescein labelled J150B	CCTCGAGGGATCCGTCCTAGCAAGCCGC[Fl~dT]GCTACCGGAAGCTTTGGACC
J150B comp	GGTCCAGAAGCTTCCGGTAGCAGCGGCTTGCTAGGACGGATCCCTCGAGG
J150B comp + 3' 25T tail	GGTCCAGAAGCTTCCGGTAGCAGCGGCTTGCTAGGACGGATCCCTCGAGGTTTTTTTTTTTTTTTTTTTTTTT
5' 25T tail + J150B comp	TTTTTTTTTTTTTTTTTTTTTTTTTGGTCCAGAAGCTTCCGGTAGCAGCGGCTTGCTAGGACGGATCCCTCGAGG
B1-25	CCTCGAGGGATCCGTCCTAGCAAGC

H1-25	GGTCCAGAAGCTTCCGGTAGCAGCG
R1-25	CGTCGAGGAATTCAACCACCGCTCT
R26-50	TCTCAACTGCAGTCTAGACTCGAGC
X26-50	GCTTGCTAGGACGGATCCCTCGAGG
15mer (anisotropy)	TCGGAGTACAGTGGG
15mer complement (anisotropy)	CCCACTGTACTCCGA

A3 Growth Media

A3.1 Auto-induction media

ZYM-5052

2 mM $MgSO_4$, 1 x trace metals, 1 x 5052, 1 x M

50 x M

1.25 M Na_2HPO_4, 1.25 M KH_2PO_4, 2.5 M NH_4Cl, 0.25 M Na_2SO_4

50 x 5052

25% glycerol, 2.5% glucose, 10% α-lactose monohydrate

ZY

1% N-Z-amine AS (or tryptone, other enzyme digests of casein)

0.5% aspartic acid

1000 x Trace Metals Mixture

50 mM $FeCl_3$, 20 mM $CaCl_2$, 10 mM $MnCl_2$-$4H_2O$, 10 mM $ZnSO_4$-$7H_2O$, 2 mM $CoCl_2$-$6H_2O$, 2 mM $CuCl_2$-$2H_2O$, 2 mM $NiCl_2$-$6H_2O$, 2 mM $NaMoO_4$-$2H_2O$, 2 mM Na_2SeO_3-$5H_2O$, 2 mM H_3BO_3.

A3.2 *Sulfolobus* Media

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; and 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 950ml of water; adjust the pH to 1.5 with H_2SO_4 and adjust volume to 1 L before autoclaving. Inoculum should be mid logarithmic (between OD_{600} 0.35-1.0) and 5-15%. Incubate the cultures at 80°C 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 °C. Harvest and resuspend mid logarithmic cultures in 0.01 volume of *Sulfolobus* medium (pH 5.5) and 15 % glycerol; and freeze in liquid nitrogen.

A4 Gel filtration column calibration

A4.1 Molecular weight markers

Table 8. Molecular weight markers and their respective elution volumes from the gel filtration column. The elution volume of blue dextran was regarded as the void volume.

Protein marker	Molecular weight (kDa)	Log (Mwt) (2sf)	Elution volume (ml) (2sf)	Log (Mwt)/Void vol (36.16 ml) (2sf)
Blue dextran	2000	3.3	36	-
β-amylase	200	2.3	56	1.5
Alcohol dehydrogenase	150	2.2	60	1.6
Albumin	66	1.8	67	1.8
Carbonic anhydrase	29	1.5	78	2.1
Cytochrome C	12.4	1.1	88	2.4

A4.2 Standard curve

The void volume (blue dextran, 36 ml) divided by the individual elution volumes was plotted against log of the molecular weights of each size standard.