Genetic Analysis of the Hel308 Helicase in the Archaeon *Haloferax volcanii*.

Rebecca Joy Gamble-Milner, MRes.

Thesis submitted to the University of Nottingham for the degree of Doctor of Philosophy

May 2016

Contents

Abstract

Hel308 is a RecQ family DNA helicase that is conserved in metazoans and archaea but is absent from bacteria and fungi. Hel308 family helicases are implicated in DNA repair, homologous recombination and genome stability, but the exact role of Hel308 is largely unknown. Strains deleted for *hel308* are sensitive to DNA inter-strand crosslinks, which are potent blocks to DNA replication. In this study, the archaeon *Haloferax volcanii* was used as a model organism to study the role of Hel308.

In archaea, homologous recombination is catalysed by polymerisation of the RadA recombinase onto ssDNA; the mediator RadB assists this process. Strains deleted for *radB* exhibit decreased levels of recombination and an increased sensitivity to DNA damaging agents. In this study, strains deleted for *hel308* in combination with *radB* exhibited in an improvement in both these phenotypes, suggesting that Hel308 acts as an anti-recombinase to antagonise RadA filament formation.

Genetic analysis of point mutants in Hel308 revealed that that the helicase activity of Hel308 is separate to its role in the regulation of recombination, which appears to rely heavily on the correct structural conformation of Hel308. Analysis of these point mutations suggests that Hel308 may act in regulating the pathway choice for the resolution of homologous recombination intermediates

This study showed that *H. volcanii* contains a second Hel308 helicase named Hel308b, which lacks the 'auto-inhibitory' domain 5 found in canonical Hel308 helicases. Deletion of *hel308b* does not lead to sensitivity to DNA inter-strand crosslinks but does result in defects in homologous recombination.

Acknowledgements

Firstly, I would like to thank Thorsten for his guidance and patience throughout my PhD. Thank you for the opportunity to learn from you, for such a fascinating project and for such an interesting 4 years!

Also, thank you to all past and present members of the lab for all your help, advice, in particular thank you to Kayleigh, Hannah, Laura and Jaime. Thank you to all members of D119 and beyond, thank you to everyone in the PhD and postdoc office for your friendship and keeping me sane!

A huge thank you to my amazing family, particularly to my Mam and Dad who have always believed in me. Thank you for and supporting me through all the difficult times and putting up with many teary phone calls over the years. Without you none of this would have been possible. Thank you to my big brother Jonathon for always having my back and to Megan for being the best sister anyone could ask for. Thank you to my Gran for cheering me on throughout my PhD (I have finally done it!).

And finally, to my wife Aimee who deserves the biggest thank you for her love and companionship and for always supporting me through everything I do.

Abbreviations

5-FOA	5-fluoroorotic Acid
Amp	Ampicillin
ATP	Adenosine 5'–triphosphate
BER	Base excision repair
BIR	Break induced replication
BLAST	Basic Local Alignment Search Tool
bp	Base pair
DEPC	Diethylpyrocarbonate
dHJ	Double- Holliday junction
DMSO	Dimethyl sulfoxide
DNA	Deoxyribose nucleic acid
dNTP	Deoxynucleotide
DSB	Double strand break
DSBR	Double strand break repair
dsDNA	Double-stranded DNA
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
HJ	Holliday junction
HR	Homologous recombination
kb	Kilobase
LB	Lysogeny broth
Leu	Leucine
Mb	Megabase
MMEJ	Micro-homology mediated end-joining
MMC	Mitomycin C
NER	Nucleotide excision repair
NHEJ	Non-homologous end joining
PAGE	Polyacrylamide gel electrophoresis
PCR	Polymerase chain reaction
PMSF	phenylmethanesulfonylfluoride
RNA	Ribonucleic acid
SDS	Sodium dodecyl sulphate
SDSA	Synthesis dependent strand annealing
ssDNA	Single-stranded deoxyribose nucleic acid
SSPE	Saline sodium phosphate EDTA
TAE	Tris/Acetic acid/EDTA
TBE	Tris/Borate/EDTA
TE	Tris/EDTA
TEMED	Tetramethylethylenediamine
TFF	Tangential flow filtration
Thy	Thymidine
Trp	Tryptophan
Ura	Uracil
UV	Ultraviolet light
v/v	Volume per volume
w/v	Weight per volume
WT	Wild-type
X-gal	5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside

Chapter 1: Introduction

1.1 Archaea

1.1.1 Discovery of Archaea

In 1977 Carl Woese and George Fox first identified Archaebacteria through nucleotide sequence analysis of small subunit ribosomal RNAs (16S rRNA) from a wide range of organisms (Woese & Fox 1977). rRNA is common to all living organisms and displays a high level of sequence conservation. This allowed Woese and Fox to carry out a sophisticated phylogenetic analysis that did not rely on cellular morphology, physiology or pathology. Archaebacteria are prokaryotic but it was later revealed that they are more closely related to Eukaryotes than Bacteria and therefore the Archaebacteria were renamed Archaea. From here, a model phylogenetic tree was developed whereby Archaea, Bacteria and Eukaryotes constitute the three domains of life, Figure 1.1 (Woese et al 1990).

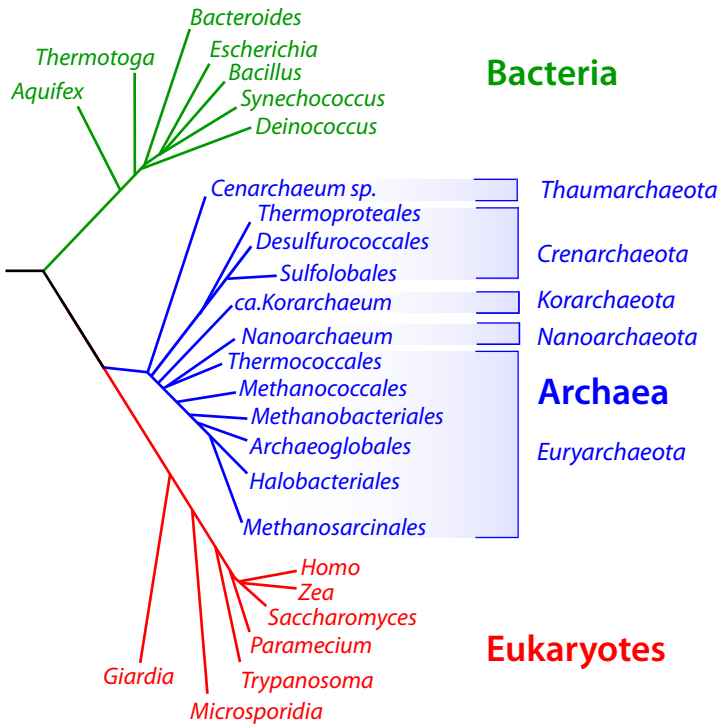


Figure 1.1: **Tree of Life based on 16S rRNA sequencing.** Bacteria, Archaea and Eukaryotes represent the three domains of life. Selected phyla and genera are shown. Figure adapted from (Allers & Mevarech 2005).

Further analysis of archaeal rRNA nucleotide sequences indicated that archaea cluster into two main phyla; the Crenarchaeota and Euryarchaeota (Winker & Woese 1991). More recently, four other phyla have been identified: Thaumarchaeota, which were previously thought to be mesophilic Crenarchaeota (Brochier-Armanet et al 2008, Pester et al 2011), Korarchaeota believed to be an ancient division that diverged from the Archaea lineage before the separation of Crenarchaeota and Euryarchaeota. (Auchtung et al 2006, Elkins et al 2008). Nanoarchaeota that contains only one member; the nanosized *Nanoarchaeum equitans* (Di Giulio 2007, Huber et al 2002, Huber et al 2003) and Aigarchaeota (Nunoura et al 2011).

As shown in Figure 1.1, Woese proposed a three domain tree where Eukaryotes are a sister group with Archaea and are derived from a common ancestor to the exclusion of Bacteria (Pace 1997). However, evidence from comparative genomics is accumulating to suggest a two domain tree where Eukaryotes have emerged from a deep branch within the TACK (Thaumarchaeota, Aigarchaeota, Crenarchaeota, Korarchaeota) superphylum of the Archaea, Figure 1.2 (Koonin & Yutin 2014).

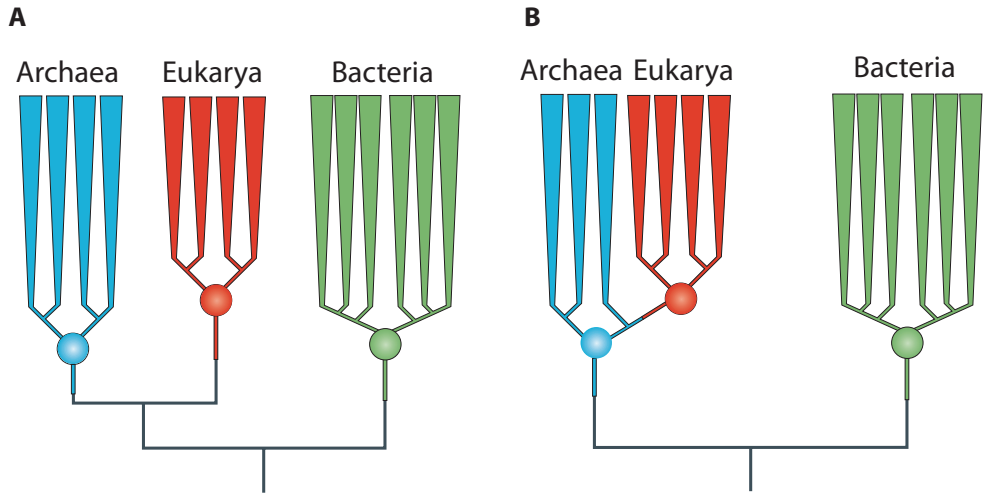


Figure 1.2: **Phylogenetic relationship between the Eukarya and the Archaea. (A)** The classical Woesean three domains of life tree, Eukarya and Archaea are two distinct sister lineages, implying that they share an ancestor. Branch lengths and number of lineages within each domain are arbitrary. For simplicity, the root of the universal tree of life has been placed in the bacterial branch. **(B)** The new proposed two domains tree of life; the Archaea and the Bacteria are the two primary domains, whereas the Eukarya is a secondary domain that arose from within the archaea. Figure adapted from (Gribaldo et al 2010).

The evidence for this hypothesis is that a dispersal of eukaryotic components such as ubiquitin (Ub) signaling, RNA interference (RNAi), actin and tubulin based cytoskeleton can be seen across the Archaea, Figure 1.3. The genome of *Ca. Caldiarchaeum subterraneum*, the only known representative of the Aigarchaeota, contains a predicted operon containing Ub-like protein, a deubiquitinating enzyme and homologs of all three Ub ligase subunits (Nunoura et al 2011). All these proteins show high similarity to eukaryotic counterparts. Recently, a distinct cell division system homologous to the ESCRT-III membrane remodeling complex has been characterized within the *Sulfolobales* of the Crenarchaeota and the Thaumarchaeon *Nitrosopumilus maritimus* (Lindas et al 2008, Makarova et al 2010, Pelve et al 2011, Samson et al 2008). The eukaryotic hallmark RNAi system that is involved in antivirus defense and regulation of gene expression has mixed archaeal and bacterial origins, with the archaeal components apparently derived from Euryarchaeota (Makarova et al 2009).

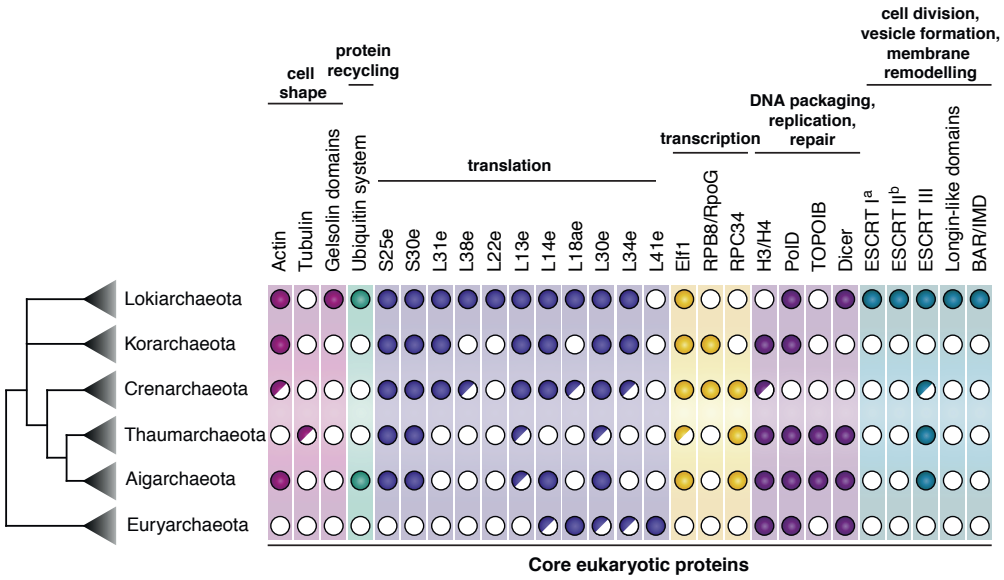


Figure 1.3: **Taxonomic distribution of archaeal orthologs of eukaryotic signature proteins**. Phylogenetic relationships are depicted as a consensus of recent phylogenomic analyses. Colour filled circles indicate the presence of homologues in all members of a lineage, whereas half filled and white circles denote patchy distribution and absence of homologues, respectively. Adapted from (Eme & Doolittle 2015)

The robustness of the two-domain hypothesis has been subjected to lively debate; seven recent large-scale phylogenomic studies have investigated the

tree of life to address the issue of the relationship between the Archaea and Eukarya. Three studies support the three-domain hypothesis that the Archaea and the Eukarya are two independent monophyletic lineages (Ciccarelli et al 2006, Harris et al 2003, Koonin & Yutin 2014). Conversely four studies support the two-domain theory that shows a relationship between the Eukarya and a particular archaeal lineage (Cox et al 2008, Foster et al 2009, Pisani et al 2007, Rivera & Lake 2004). Although, progress has been made to understand the relationship between the Archaea and Eukarya, no individual analysis is definitive. The Archaea are central to understanding the origin of Eukarya, and metagenomic mining of new Archaeal lineages in the future could provide the answers to this phylogenetic problem.

1.1.2 Characteristics of Archaea

Archaea are typically thought of as extremophiles capable of living in harsh environments such as hypersaline brines, hydrothermal vents and acidic/alkaline waters. However, archaea are found to be widespread in non-extreme habitats such as soil, fresh water sediments and oceans where they represent more than 20% of all microbial cells present (DeLong & Pace 2001, Pace 1997).

Archaea possess features in common with both bacteria and eukaryotes, see Table 1.1 for details. Morphologically, archaea are more closely related to bacteria than eukaryotes but in relation to DNA replication, transcription and translation in archaea resemble eukaryotic systems, however they are less complex (Barry & Bell 2006, Grabowski & Kelman 2003).

Table 1.1: Properties of the three domains of life.

Trait	Bacteria	Archaea	Eukaryotes	Reference
Nucleus	No	No	Yes	(Lake 1989)
Chromosome	Mostly circular	Circular	Linear	(Tan & Tomkins 2015)
Origins	Single	Single or Multiple	Multiple	(Kelman & Kelman 2004, Wu et al 2014)
Chromatin Proteins	None	Histones (some species)	Histones	(Pereira & Reeve 1998)
Organelles	No	No	Yes	(Gray 1989)
Lipids	Ester-linked to an *sn*-glycerol-3-phosphate backbone	Ether-linked to an *sn*-glycerol-1-phosphate backbone	Ester-linked to an *sn*-glycerol-3-phosphate backbone	(Albers & Meyer 2011)
Cell wall	Yes (peptidoglycan)	Some species (pseudopeptidoglycan)	No	(Albers & Meyer 2011)
Cell size	<5μm	<5μm	10 -100 μm	(Tan & Tomkins 2015)

Adapted from (Tan & Tomkins 2015)

Genome Organisation

Like bacteria and unlike eukaryotes, archaea have circular chromosomes and contain operons (Olsen & Woese 1997). However, archaeal genomes tend to have a higher gene density than bacterial or eukaryotic genomes (Koonin & Wolf 2008). Archaea also possess single or multiple origins of replication, this is discussed in further detail in Section 1.3.1.

Similar to eukaryotes, the majority of Euryarchaeal, Thaumarchaeal and Nanoarchaeal genomes are packaged by histone proteins. In eukaryotes, DNA is compacted by tight wrapping around octamers of four core histones; H2A, H2B, H3 and H4 to form a nucleosome. Eukaryotic histones contain N- and C-terminal tails that undergo extensive posttranslational modifications such as acetylation and methylation, which represses transcription and therefore gene expression (Reeve et al 1997, White & Bell 2002). Archaeal histones are shorter and lack these N- and C-terminal tails suggesting that regulation of transcription is not mediated through histones. Furthermore archaeal histones can form both homodimers and heterodimers and the number of histones varies from species to species. For example *Methanothermus fervidus* contains two histone-coding genes whereas *Methanobacterium* species has three, *Methanococcus jannaschii* has five and *Haloferax volcanii* has one (Ammar et al 2012, Bult et al 1996, Grayling et al 1996).

In addition to histones, Crenarchaea, Thaumarchaea, Nanoarchaea and some Euryarchaea also contain the DNA binding protein Alba. Excluding the Crenarchaea, Alba is always found alongside a histone (Sandman & Reeve 2005, White & Bell 2002). Alba induces negative supercoiling and protects DNA from nucleases. Species such as methanogens that lack Alba contain the DNA binding protein MC1 that significantly distorts and compacts circular DNA (Cam et al 1999, Toulme et al 1995).

DNA Replication and Repair

The mechanisms and proteins involved in archaeal DNA replication and DNA repair is more eukaryotic than bacterial, albeit simpler. DNA replication and repair across all three domains of life is discussed in further detail in Sections 1.3 and 1.4.

Transcription and Translation

The core apparatus of transcription, RNA polymerase, is universal in distribution across all domains of life, but the subunits of the archaeal and eukaryotic RNA polymerase (RNAP) are more alike than the bacterial counterpart, Figure 1.4 (Bell & Jackson 2001, Hirata & Murakami 2009, Langer et al 1995). The TATA box binding protein (TBP) and transcription factors IIB and IIIB (also known as TFB) involved in the initiation of transcription are also highly conserved across eukaryotes and archaea (Gietl et al 2014). However, transcription in archaea is regulated by bacterial proteins such as the leucine-responsive regulatory protein (Lrp), a major global regulator found in *Escherichia coli* (Peeters & Charlier 2010).

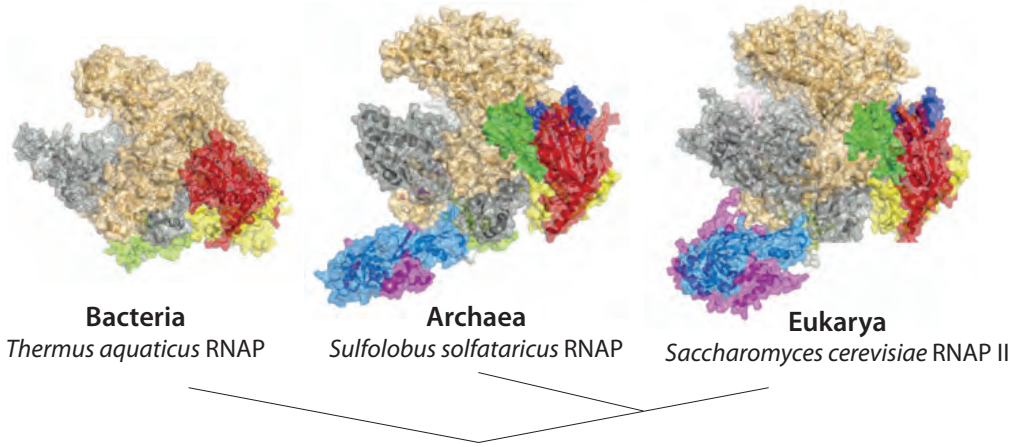


Figure 1.4: **RNA polymerase structures across the three domains of life.** Crystal structure of bacterial, archaeal and eukaryotic RNAPs. Archaeal and eukaryotic RNAPs are more closely related containing equivalent subunits that are not present in the bacterial RNAP. Orthologous subunits are shown in the same colour. Figure adapted from (Albers et al 2013)

The majority of archaeal translation initiation factors (IFs) are universally conserved across all domains of life (IFs 4-6), however the factors a/eIF2 and aIF6 are shared only with eukaryotes. No archaea-specific initiation factors have been identified (Londei 2005).

1.1.3 Halophiles

Archaea are highly diverse and are found in a wide range of moderate and extreme locations for example oceans and soils to acidic pools, hydrothermal vents, hot springs and hypersaline pools (Chaban et al 2006). Organisms that live in hypersaline conditions are known as halophiles and for these organisms to survive in such conditions, one of two mechanisms are employed to adapt to high salt conditions. The first mechanism used largely by halophilic bacteria and eukaryotes is the 'salt out' mechanism. Here, salts are pumped out from the cell and the cytoplasm is packed with organic solutes such as glycerol or glycine betaine to maintain the osmotic balance (Christian & Waltho 1962, Oren 1999, Oren 2008). Halophilic archaea and some bacteria accumulate high levels of salt in the cytoplasm to maintain osmotic balance, this is a 'salt in' mechanism approach (Oren et al 2002). Proteins in halophilic archaea have adapted to function in high salt and low water conditions by several different strategies, which will be discussed in further detail in Chapter 6: *in vitro Analysis of Hel308,* Section 6.1.1: *Halophilic Proteins*.

1.2 *Haloferax volcanii*

Haloferax volcanii is a halophilic euryarchaeon that was originally isolated from the Dead Sea. *H. volcanii* is coccid, pigmented red due to the presence of carotenoids and the outer cell surface is comprised of a glycoprotein S layer (Hartman et al 2010, Mullakhanbhai & Larsen 1975). *H. volcanii* is aerobic and can be easily in cultured at an optimum growth temperature of 45ºC in NaCl concentrations of 1.7 - 2.5 M, with a generation time of 2.5-3 hours. Due to its ease of culture and large range of genetic tools available (discussed in section 1.2.1), *H. volcanii* is a model organism for the study of archaeal genetics.

There is no evidence for a defined cell cycle within *H. volcanii* (Iain Duggin, personal communication). This is corroborated by DNA replication profiles generated through deep sequencing, which showed that the genome copy number maximum:minimum ratio in *H. volcanii* is >2:1. This demonstrates that concurrent rounds of replication are occurring within the cell and is evidence against the presence of a defined S-phase (Hawkins et al 2013).

The complete genome of *H. volcanii* is 4.2 Mb in size, comprising a 2.85 Mb main chromosome and three megaplasmids: pHV1 at 86 kb, pHV3 at 442 Kb and pHV4 at 690 Kb (Charlebois et al 1991). The 6Kb plasmid, pHV2 was cured from the laboratory strain but is present in the wild-type isolate DS2 (Wendoloski et al 2001). Additionally, in the laboratory strain H26 the pHV4 megaplasmid has integrated onto the main chromosome (Hawkins et al 2013). *H. volcanii* has a genome with a %GC of around 65%, furthermore the genome of *H. volcanii* is highly polyploid with up to 20 genome copies present per cell (Breuert et al 2006). Finally, a full genome sequence is available for *H. volcanii* (Hartman et al 2010).

1.2.1 Genetic tools for *H. volcanii*

Selectable markers

The use of antibiotic selection is available within *H. volcanii*. Mutations in the *gyrB* and *hmgA* genes gives resistance to Novobiocin, which is an inhibitor of DNA gyrase and mevinolin an inhibitor of the HMG-CoA reductase, respectively (Holmes et al 1991, Lam & Doolittle 1989). However, these markers were isolated as mutant alleles of essential genes and therefore show near-complete homology to the chromosomal allele. Therefore homologous recombination may occur between the plasmid-borne resistance marker and the genome, resulting in the acquisition of antibiotic resistance by the host strain.

To overcome this problem, several gene deletions a have been made in *H. volcanii* to allow for the use of auxotrophic selectable markers. Selectable markers relevant to this study are described in Table 1.2.

Table 1.2: **Auxotrophic selection methods available for use in *H. volcanii*.**

Gene	Encodes	Involved in	Reference
pyrE2	Orotate phosphoribosyl transferase	Uracil biosynthesis	(Bitan-Banin et al 2003)
trpA	Tryptophan synthase	Tryptophan biosynthesis	(Allers et al 2004)
leuB	3-isopropylmalate dehydrogenase	Leucine biosynthesis	(Allers et al 2004)
hdrB	Dihydrofolate reductase	Thymidine biosynthesis	(Ortenberg et al 2000)

Reporter genes

Two reporter genes are widely used within *H. volcanii*, the first is a ß-galactosidase gene which allows for the detection of blue colonies after strains have been treated with X-gal (Holmes & Dyall-Smith 2000). This marker can be utilised in growth competition assays and recombination assays (Delmas et al 2009). The second reporter gene is green fluorescent protein (GFP) that has been adapted by three amino acid substitutions to allow for use in high salt conditions (Crameri et al 1996, Reuter & Maupin-Furlow 2004)

Transformation

H. volcanii can be easily transformed with plasmid DNA by the removal of the S-layer using EDTA and the addition of polyethylene glycol with facilitates the uptake of DNA (Cline et al 1989). This is described in further detail in Chapter 2: *Materials and Methods,* Section 2.2.2: *General Haloferax volcanii Microbiology*.

Gene Deletion/Replacement

A gene deletion and replacement/knock out system has been developed in *H. volcanii* that utilises the *pyrE2* marker during pop-in and pop-out steps. This is described in further detail in Chapter 2: *Materials and Methods,* Section 2.2.5: *Genetic Manipulation of Haloferax volcanii*.

Gene Expression

In *H. volcanii,* the overexpression of genes on episomal plasmids is induced by the addition of tryptophan. Overexpression plasmids contain the gene of interest downstream of the tightly-controlled tryptophan-inducible promoter p.*tnaA* (Large et al 2007), which originates from the *H. volcanii* tryptophanase gene *tnaA*. This promoter has been used for the overexpression of proteins natively in *H. volcanii* (Allers et al 2010). This will be discussed in further detail in Chapter 6: *in vitro Analysis of Hel308*.

1.2.2 Biochemical tools for *H. volcanii*

Halophilic proteins are typically insoluble and misfold in low salt conditions; therefore it is inadvisable to express proteins from *H. volcanii* in heterologous hosts such as *E. coli*.

Prior to this study, biochemical tools were developed to allow for the expression and purification of histidine tagged proteins in *H. volcanii*. The deletion of the restriction enzyme Mrr allows for efficient transformation of

protein expression plasmids into *H. volcanii* without the need to isolate the plasmid DNA from *E. coli dam-* strains. *H. volcanii* host strains were generated that reduce the co-purification of naturally histidine-rich proteins with *his_6-tagged* recombinant proteins, this required the removal of a histidine rich linker from PitA and the truncation of the Cdc48d protein to remove a histidine rich domain (Allers et al 2010). However, even after removal of these major contaminants, other histidine-rich contaminating proteins still co-purify during protein purification. For this reason, one aim of this study was to develop new biochemical methods for native protein expression and purification from *H. volcanii*. This will be discussed in detail in Chapter 6: *in vitro Analysis of Hel308*.

1.3 DNA replication

The complete and accurate replication of DNA is integral to the proliferation of life in all three domains. Without DNA replication genetic information cannot be transmitted to the next generation. The replication of DNA is tightly regulated and must occur before cell division. Across all three domains of life, DNA replication can be divided into three distinct stages: initiation, elongation and termination.

1.3.1 Initiation

Initiation of DNA replication occurs at defined regions on the genome called origins of replication. Bacteria typically contain a single origin of replication called the *oriC*, however bacteria have circular chromosomes and typically undergo concurrent rounds of replication (O'Donnell et al 2013). In *E. coli* the *oriC* is approximately 250 bp in length and contains an array of 9 bp repeat elements known as DnaA boxes that are located within the DNA unwinding element (DUE). These are sequence-specific binding sites for the initiator protein DnaA, an AAA+ family ATPase that forms a nucleoprotein filament at the *oriC* (Ozaki & Katayama 2009). Assembly of DnaA at the DnaA boxes results in stretching and therefore melting of the DNA within the DUE, generating single stranded DNA onto which the replicative helicase DnaB is

loaded (Erzberger et al 2006). DnaB is a hexameric helicase that uses the hydrolysis of ATP to unwind double stranded DNA. Loading of DnaB is mediated by the AAA+ family ATPase helicase loader DnaC, which loads DnaB on to single stranded DNA at the *oriC* by breaking the hexameric ring of DnaB (Arias-Palomo et al 2013). To prevent secondary structure formation and degradation by nucleases, the ssDNA generated by the helicase unwinding action of DnaB is bound by the ssDNA-binding protein (SSB) to form a nucleoprotein filament. In *E. coli* SSB is a homotetramer that can bind ssDNA via an oligosaccharide binding fold (OB-fold) and form protein-protein interactions via its C-terminal domain (Shereda et al 2008).

Due to larger genome sizes, eukaryotes have multiple origins of replication across a linear genome (O'Donnell et al 2013). The initiator machinery in eukaryotes is called the origin recognition complex (ORC), which is a six subunit heteromeric protein composed of the proteins Orc1-6 (Bell & Dutta 2002). ORC recruits the replication factor Cdc6, an AAA+ family ATPase and Cdt1, which function as a helicase loader and so in turn recruit the heterohexameric replicative MCM2-7 helicase to the origin. Together these proteins form the pre-replicative complex (pre-RC) and the origin is now licenced for replication. MCM forms a protein complex with Cdc45 and GINS which is also known as the CMG complex (Onesti & MacNeill 2013). The formation of the CMG complex transforms the pre-replicative complex into the pre-initiation complex; this complex is closely regulated by phosphorylation (Sclafani & Holzen 2007). To prevent DNA damage, the ssDNA generated by the DNA unwinding of the CMG complex is bound by replication protein A (RPA) which is the eukaryotic equivalent to SSB in bacteria (Nguyen et al 2014).

Archaea have either one or multiple origins of replication, for example *Pyrococcus abyssi* contains a single origin of replication whereas *H. volcanii* contains four origins of replication on its main chromosome (*oriC-1, 2,3 and ori-pHV4*) (Hawkins et al 2013, Kelman & Kelman 2004). Note that *ori-pHV4* is located on the chromosome as a result of the integration of the pHV4

megaplasmid onto the chromosome in the laboratory strain H26 of *H. volcanii*, Figure 1.5.

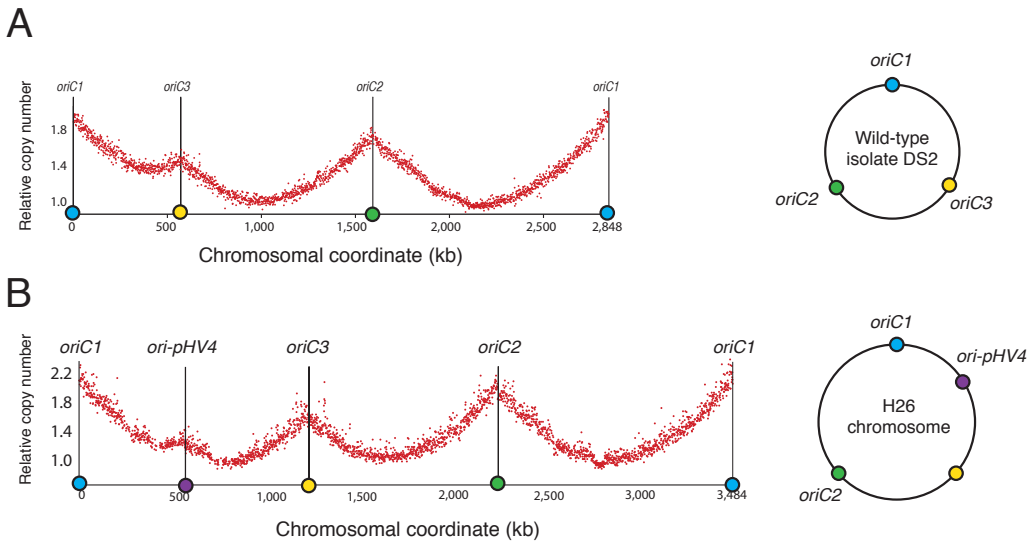


Figure 1.5: **Replication profile for the main chromosome of *H. volcanii*.** (A) Replication profile for the wild-type isolate DS2, the main chromosome has 3 origins of replication. (B) Replication profile for the lab strain H26. pHV4 has integrated into the main chromosome, resulting in 4 origins on the main chromosome. Figure adapted from (Hawkins et al 2013).

Archaea contain at least one homolog of the eukaryotic Orc1 or Cdc6 initiator proteins (Barry & Bell 2006), Many archaeal species contain multiple Cdc6/Orc1 proteins, *Pyrococcus* has only one, *Sulfolobus* has three and *H. volcanii* contains 16 *orc1/cdc6* genes. Not all of these genes are located next to origins of replication and at least two of these genes are not involved in DNA replication (Norais et al 2007). From deletion studies in *H. volcanii*, it appears that many of the *orc1/cdc6* genes have overlapping functions. The Cdc6/Orc1 proteins bind to origin recognition boxes (ORBs) and recruit the MCM replicative helicase. In contrast to eukaryotes, archaeal genomes only appear to have one MCM homologue, which forms homohexamers *in vitro*, but like the eukaryotic MCM proteins and bacterial DnaB, archaeal MCM is an AAA+ protein. Furthermore, no archaeal homologues to bacterial or eukaryotic helicase loading factors have been found suggesting that MCM interacts directly with Orc1/Cdc6 at the origin of replication (Barry & Bell 2006). Like eukaryotes, once recruited to the replication origin, MCM interacts with GINS to form the CMG complex (MacNeill 2010). But unlike eukaryotes the GINS complex also associates with a GINS-associated nuclease (GAN) which is a RecJ family nuclease and an ancient homologue of Cdc45 (Bell 2011,

Makarova et al 2012). The ssDNA generated by the CMG helicase activity is bound by a single stranded DNA binding protein that is either like the eukaryotic RPA or the bacterial SSB depending on the species (Barry & Bell 2006). The crenarchaeal single stranded binding protein from *Sulfolobus solfataricus* shows a similar structure to bacterial SSB (Wadsworth & White 2001). Euryarchaea have single stranded binding proteins more similar to the eukaryotic RPAs. The formation of RPA complexes varies between archaeal species. In *H. volcanii* RPA1 and RPA3 interact with the archaea specific RPA-associated proteins RPAP1 and RPAP3 but are not essential, whereas the essential RPA2 protein acts alone (Stroud et al 2012).

After the initiator protein, replicative helicase and other replicative factors have been recruited, bi-directional DNA synthesis is initiated at the origin of replication. A replication fork showing replication factors from all three domains of life is shown below, Figure 1.6.

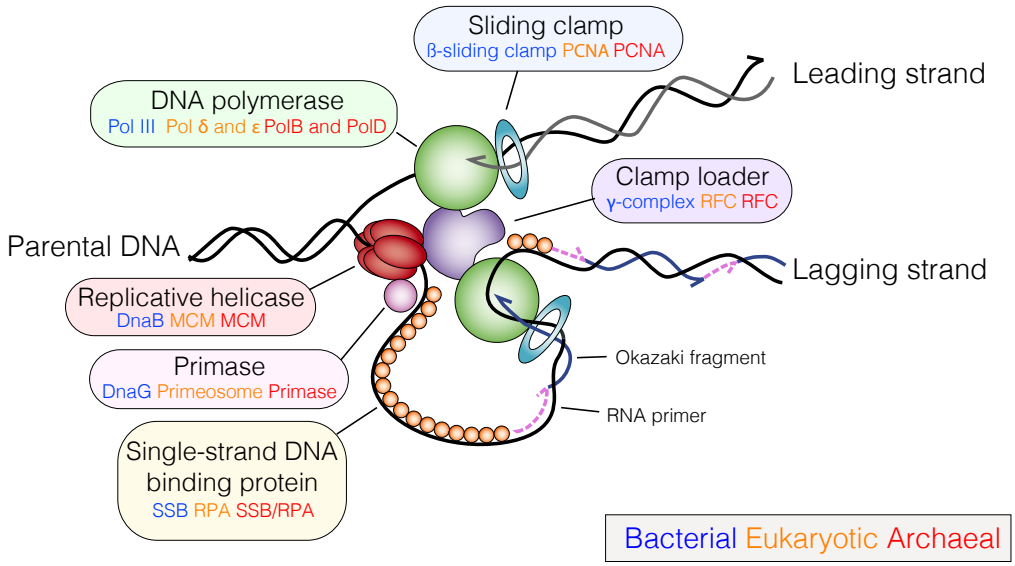


Figure 1.6 **Components of a replication fork.** A replication fork showing equivalent components from all three domains of life, bacteria (blue, eukaryotes, (orange) and archaea (red). Figure adapted from (McGlynn & Lloyd 2002b).

1.3.2 Elongation

The elongation phase of DNA replication can be divided into two stages, priming and DNA synthesis, both stages are discussed below.

Priming

DNA polymerases are unable to initiate DNA synthesis *de novo*, they require a short primer from which they can extend. An RNA polymerase known as a primase synthesises short RNA primers, a DNA polymerase then elongates these RNA primers and so carries out DNA synthesis. On the leading DNA strand (5'-3'), DNA synthesis is continuous and DNA synthesis on the lagging strand (3'-5) is discontinuous, leading to the formation of Okazaki fragments.

In bacteria, the short RNA primers are synthesised by the primase, DnaG. The initiation of the DnaG requires interaction with DnaB, the generation of a primer occurs in the opposite direction to helicase unwinding by DnaB (Fang et al 1999). RNA primers synthesised by DnaG can range from 10 to 60 nucleotides in length but are typically around 11 nucleotides long (Frick & Richardson 2001).

The primase in eukaryotes is a dimer consisting of PriS and PriL subunits, which associate with Pol α and the accessory B subunit to form a Pol α/primase complex. The primase synthesizes RNA primers of 8 to 12 nucleotides, which are then elongated to around 30 nucleotides by Pol α to produce a DNA-RNA hybrid before handing over to the replicative polymerase (Arezi & Kuchta 2000, Frick & Richardson 2001, Kuchta & Stengel 2010).

Archaea possess homologs of the eukaryotic PriS and PriL primases but lack the Pol α and the accessory B subunits, Archaeal primeases typically synthesise RNA primers 7-10 nucleotides in length (Barry & Bell 2006).

DNA synthesis

Once RNA primers have been synthesised by primases, the primer is then extended by DNA polymerases. Additional members of the replisome that are essential for DNA synthesis are circular sliding clamps and clamp loaders, which will be discussed below.

DNA is synthesised by DNA polymerase III holoenzyme in bacteria, this enzyme comprises two DNA polymerases in complex with nine other subunits (Kelman & O'Donnell 1995). The polymerase core consists of the DNA polymerases α and ε, a 3'-5' exonuclease and θ, these subunits are responsible for DNA polymerisation and proof-reading. DNA polymerase III is loaded onto the DNA by a ß clamp, which is able to encircle and slide along the DNA. The ß clamp is itself loaded by the γ complex, and the ß clamp increases processivity of the DNA polymerase by stabilising the complex (Kuriyan & O'Donnell 1993, Turner et al 1999).

Eukaryotic DNA synthesis is performed by two B-family DNA polymerases Polε and Polδ, which synthesise DNA on the leading and lagging strand respectively (O'Donnell et al 2013). The proliferating cell nuclear antigen (PCNA) is a sliding clamp which has a similar structure to ß clamp in bacteria. PCNA aids Polε and Polδ in DNA replication by tethering replicative enzymes to the replication fork. PCNA itself is loaded onto DNA by the clamp loader replication factor C (RFC), which is similar in structure to the bacterial clamp loader (Moldovan et al 2007).

Archaea contain a family B polymerase, and Euryarchaea and Thaumarchaeota contain an extra family D polymerase that is composed of the DP1 and DP2 subunits. DP1, which is related to the eukaryotic Polδ, exhibits 3'-5' exonuclease activity and proofreading ability (Barry & Bell 2006, Cann et al 1998). Like eukaryotes, the DNA polymerases are assisted by the sliding clamp PCNA that is loaded by the clamp loader RFC (Cann et al 1999).

DNA synthesis on the lagging strand is discontinuous and leads to the formation of Okazaki fragment flap structures separated by RNA primers. To generate a continuous nascent strand of DNA, primer removal, gap filling and ligation are necessary. In bacteria the Okazaki fragment flaps are removed by the RNase H endonuclease before ligation (Reyes-Lamothe et al 2012). In eukaryotes Okazaki fragment maturation is carried out by polymerase δ, RNase H, flap endonuclease 1 (FEN1), Dna2 and DNA ligase I, all of which

are co-ordinated by PCNA (Zheng & Shen 2011). In archaea, the RNA primers are removed FEN1 and RNase H2, which degrades the RNA primer. The gap in the DNA is joined by the DNA ligase LigA which is recruited to the nick site by PCNA (Barry & Bell 2006). In addition to LigA, *H. volcanii* possesses the DNA ligase LigN, which have been suggested to have an overlapping function in Okazaki fragment maturation (Giroux & MacNeill 2015, Zhao et al 2006).

1.3.3 Termination

Bacteria have circular genomes and in *E.coli* termination occurs opposite to the *oriC* at *Ter* sequences where the bidirectional replication forks meet. Ten oppositely-orientated *Ter* sequences, also known as a replication trap, block replication forks from traveling through them in a specific direction (Duggin et al 2008). This occurs when the *Ter* sites are bound by the terminator protein Tus. The Tus-*Ter* complex acts by blocking the action of the replicative DnaB helicase, but details of the mechanism are uncertain (Neylon et al 2005).

Unlike bacteria, eukaryotes do not have defined termination sites. Termination in eukaryotes occurs when two replication forks meet and the two nascent DNA strands are ligated together. This occurs randomly in the region between a pair of replication origins. However, in highly transcribed regions such as ribosomal operons site-specific fork barriers that prevents the collision of replication and transcription machinery results in the termination of replication (Eydmann et al 2008).

Like eukaryotes, archaea appear not to have defined termination sites, termination is likely to occur due to the collision of replication forks from multiple origins of replication. Replication profiles in *Sulfolobus spp*. shows that termination occurs asynchronously (Lundgren et al 2004). In *H. volcanii* termination is seen to happen over broad regions of the genome, indicated by smooth valleys as opposed to sharp troughs on marker frequency profiles, Figure 1.5 (Hawkins et al 2013).

1.3.4 Lesions in the DNA template

The replication machinery accurately copies undamaged template DNA but is unable to copy DNA faithfully in the presence of DNA lesions. Lesions on the leading strand of DNA can lead to replication fork stalling or collapse. Since DNA synthesis is discontinuous, lesions in the lagging strand can be bypassed by a new priming event and repaired at a later time (Pages & Fuchs 2003). Unrepaired DNA lesions can lead to mutations or wide-scale genome aberrations.

1.4 DNA repair

1.4.1 DNA damage

Direct chemical damage to the template DNA comes from both exogenous (environmental) and endogenous (spontaneous) sources and can lead to a variety of lesions, Figure.1.7. Some of these DNA lesions will be discussed in further detail below.

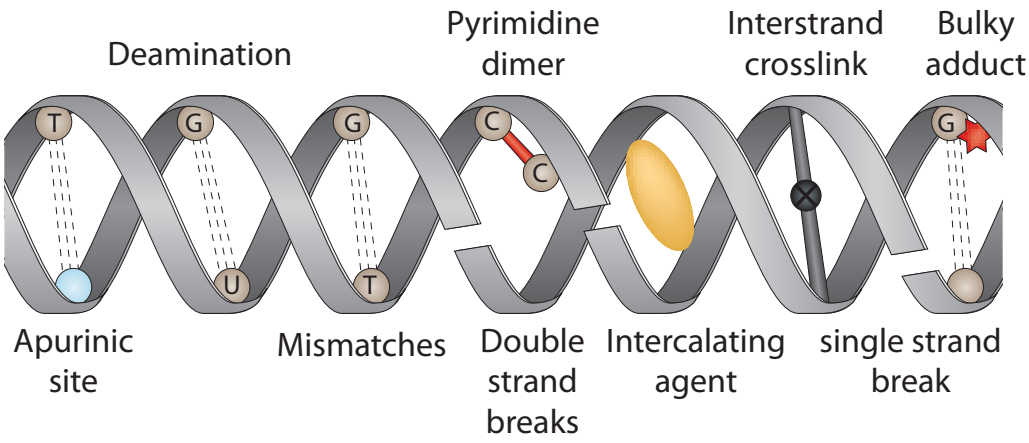


Figure 1.7: **Types of DNA damage.** Common types of DNA damage caused by exogenous and endogenous sources. Figure adapted from (Helleday et al 2014)

1.4.1.1 Endogenous DNA damage

Endogenous DNA damage occurs from chemical species that arise as a result of normal metabolic processes within the cell.

Oxidative DNA lesions

Free radicals such as reactive oxygen species or nitrogen oxide species are generated endogenously as by-products of normal cellular metabolism (Helleday et al 2014). These species can cause over 25 different oxidative base lesions, one of the most studied is 8-oxo-2′-deoxyguanosine, which forms hydrogen bonds with adenine and leads to G·C to T·A base pair transitions (Evans et al 2004). Oxidative species can also lead to breakages in the DNA backbone.

Deamination

Deamination occurs spontaneously in all DNA bases that contain primary amines, the principle repair mechanism for deamination is base excision repair (BER). Common deamination reactions include the hydrolytic deamination of 5-methylcytosine to thymine, which results in a C·G base pair to be converted to T·A during the next round of DNA replication (Lutsenko & Bhagwat 1999). Deamination of a cytosine to uracil can occur spontaneously and can also be catalysed by members of the cytidine deaminase family such as AID which converts cytidine to uridine to initiate the hypermutation process during immunoglobulin maturation (Teng & Papavasiliou 2007). Deamination of adenine to hypoxanthine gives rise to A·T to base pair G·C transitions (Lindahl 1993).

1.4.1.2 Exogenous DNA damage

Exogenous DNA damaged is caused by external chemical or physical agents, examples relevant to this thesis are described below.

Ultraviolet light

A major source of DNA damage is ultraviolet (UV) radiation that can cause several different types of DNA lesions such as pyrimidine dimers, 6-4 photoproducts, and single and double strand DNA breaks.

Upon excitation from high-energy UV radiation the C5 and C6 double bonds on neighbouring pyrimidine bases can become covalently crosslinked to form cyclobutane pyrimidine dimers (CPD), Figure 1.8. UV radiation can also induce pyrimidine (6-4) photoproducts where a single bond is formed between the C6 and C4 of adjacent pyrimidines (Ravanat et al 2001, Sinha & Hader 2002). These 'bulky' lesions distort the backbone of DNA and if unrepaired can lead to C·G to T·A or CC·GG to TT·AA substitutions during DNA replication (Helleday et al 2014). In organisms other than placental mammals, CPDs and (6-4) photoproducts can be repaired directly by photolyases in a process called photoreactivation. Bulky DNA lesions can also be removed by nucleotide excision repair (NER), here a section of DNA containing a lesion is removed, filled in and ligated (Sinha & Hader 2002). In humans NER is the primary mechanism in repairing UV-induced lesions. Both of these repair mechanisms will be discussed in further detail in Sections 1.4.2.1 and 1.4.2.2.

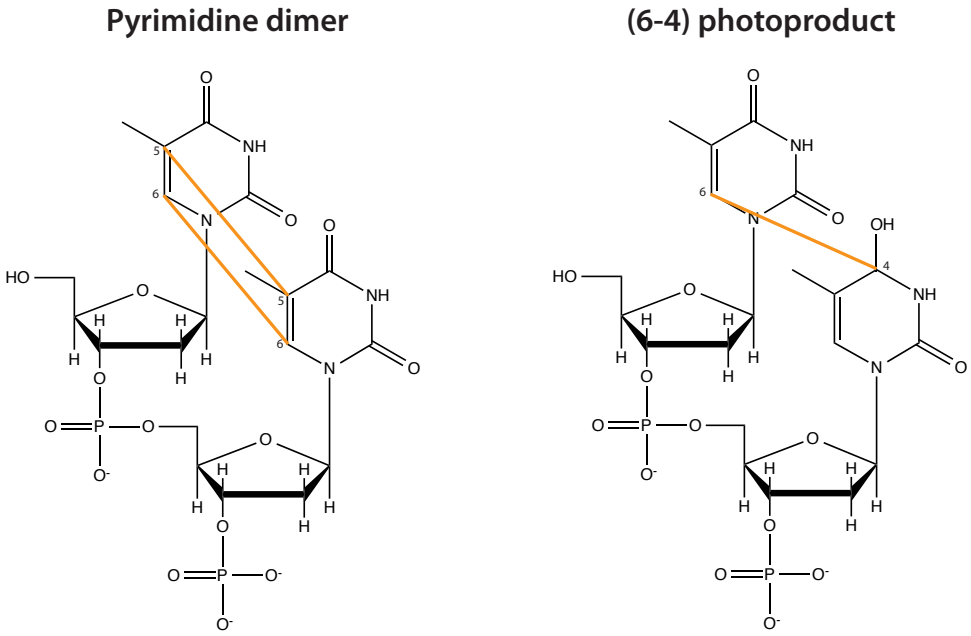


Figure 1.8**: UV-irradiation results in the formation of pyrimidine dimers and (6-4) photoproducts.** UV light is absorbed by pyrimidine C=C double bonds, causing them to open and allowing them react with neighbouring molecules. Adjacent thymindines can be covalently crosslinked to form pyrimidine dimers at the C=C double bonds. Alternatively, (6-4) photoproducts form in which the C6 and C4 of adjacent pyrimidines covalently link.

UV radiation also causes the production of reactive oxygen species that can result in single strand DNA breaks (Cadet et al 2005, Lankinen et al 1996). Double strand DNA breaks (DSB) can arise if two single strand breaks are in

close proximity or if a replication fork encounters a single strand break on the leading strand. DSBs can be repaired by either by homologous recombination, non-homologous end joining (NHEJ) or microhomology-mediated end joining (MMEJ), all of which will be discussed in further detail in Section 1.4.2.3.

DNA Crosslinking Agents

Many chemical agents can form crosslinks within DNA molecules but due to its relevance to this thesis, only mitomycin C (MMC) will be discussed here. MMC is naturally synthesised by *Streptomyces caespitosus* and is a chemical mutagen that is commonly used as a chemotherapeutic anti-tumour agent (Tomasz 1995). MMC monoalkylates DNA at guanine bases leading to monoadduct formation between MMC and DNA. Alkylation of a second guanine base generates a bisadduct and therefore crosslinking of the DNA. Crosslinking can occur between the two strands of DNA or between the same strand to generate inter-strand crosslinks and intra-strand crosslinks respectively. Interstrand crosslinks can be repaired by NER and homologous recombination, if unrepaired DNA crosslinks lead to blockages in DNA replication due the replication forks inability to pass through the crosslinked region.

1.4.2 DNA Repair Pathways

1.4.2.1 Direct repair

The majority of organisms contain mechanisms to repair DNA by direct reversal of chemical damage.

The first mechanism involves the photoreversal of UV-induced cyclobutane pyrimidine dimers and (6-4) photoproducts by DNA pyrimidine dimer photolyases and 6-4 photolyases respectively. Photolyases in *E. coli* are monomeric and contain the chromophore cofactor folate that absorbs violet/blue light photons, which excites a second cofactor, flavin (Li et al 1991). The excited flavin transfers an electron to the cyclobutane pyrimidine dimer to generate a dimer radical anion, which in turn splits the two

pyrimidines and thus reversing the damage (Sancar et al 2004) . Placental mammals lack photolyases and so are unable to perform photoreversal of pyrimidine dimers.

A second mechanism involves the removal of the O^6-methyl group from O^{6-} methylguanine (O^6MeGua) DNA bases by methylguanine DNA methyltransferase. The methylguanine methyltransferase flips out the O^6MeGua DNA base where the methyl group is then transferred to an active site cysteine in the enzymes active site (Olsson & Lindahl 1980). The enzyme becomes inactive after one catalytic event and so dissociates from the now repaired DNA. Mice lacking the O^6MeGua methyltransferase *MGMT* gene are highly susceptible to tumorigenesis by DNA alkylating agents (Kawate et al 1998).

1.4.2.2 Excision repair

Excision repair involves the removal of patches of DNA containing damaged bases or nucleotides via incisions on the damaged strand at either side of the lesion. The gap is then filled via DNA synthesis, replacing the damaged segment of DNA using the complementary strand as a template for repair. Defects in excision repair leads to a higher incidence of cancer susceptibility.

Base excision repair

Base excision repair (BER) corrects small DNA lesions resulting from oxidation, deamination, methylation and alkylation that do not significantly distort the DNA helix. BER can also remove and replace misincorporated uracil. The main components of BER are well conserved across all three domains of life (Sartori & Jiricny 2003). Defects in BER have been linked to age related diseases such as cancer and neurodegeneration (Wilson & Bohr 2007).

BER is initiated by a lesion-specific DNA glycosylase that flips out the damaged base and cleaves the N-glycosylic bond between the damaged base

and the deoxyribose, thus removing the base and leaving an apurinic or apyrimidinic (AP) site, Figure 1.9. An AP endonuclease cuts the sugar-phosphate backbone to the 5' side of the AP site, giving rise to a single strand break. The upstream fragment is terminated with a free 3' hydroxyl group and the downstream has a baseless deoxyribose-phosphate (dRP) residue at its 5' terminus. In *E. coli* baseless dRP is removed by a 5' phosphodiesterase such as RecJ (Piersen et al 2000). Some DNA glycosylases have a second function in which they display beta lyase activity that cleaves the phosphodiester bond 3' to the AP site. In mammalian cells the processing of the AP sites can occur through two different pathways; the preferred 'short-patch' BER which results in the replacement of a single nucleotide and 'long-patch' BER where the repair tract is 2-6 nucleotides in length (Dogliotti et al 2001, Sartori & Jiricny 2003). DNA polymerase β fills in the nucleotide gap and removes the dRP moiety by elimination and is ligated to seal the resulting nick (Matsumoto & Kim 1995). In 'long-patch' BER a longer 'flap' oligonucleotide is produced in the gap filling process, the overhanging flap is excised by the flap-endonuclease (FEN1) which is present in eukaryotes and archaea, and the resulting nick is sealed by DNA ligase (Lindahl 1993).

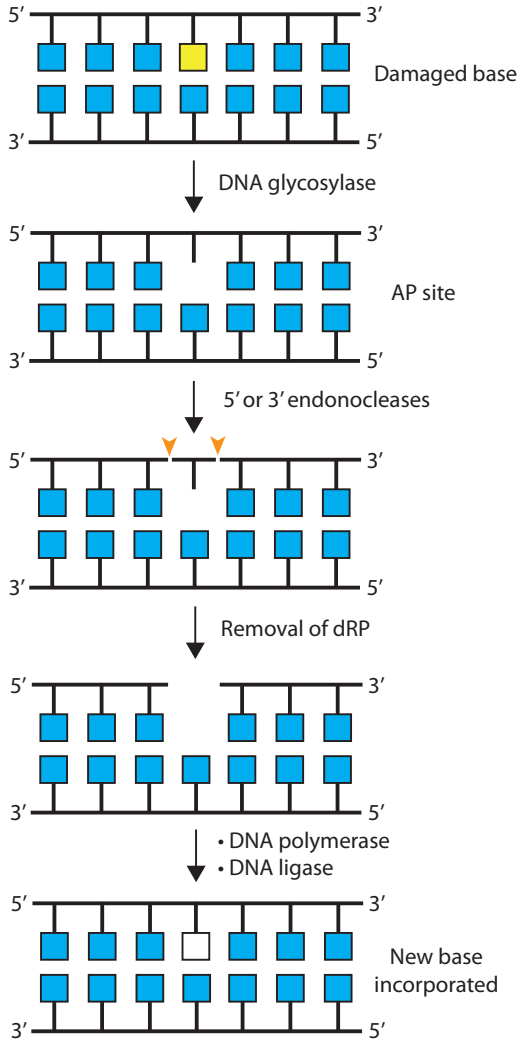


Figure 1.9: **Base excision repair.** Glycosylases remove damaged bases resulting in an apurinic or apyrimidinic (AP) site. AP endonuclease cleave the sugar-phosphate backbone to remove the baseless deoxyribose-phosphate (dRP). Gap filling is carried out by DNA polymerase using the complementary strand of DNA as a template. The nick is sealed by DNA ligase.

Nucleotide excision repair

Nucleotide excision repair (NER) is generally considered to be an 'error free' and highly versatile DNA damage removal pathway that counteracts the deleterious effects of an array of DNA lesions. Lesions repaired by NER tend to cause significant distortion to the DNA helix (de Laat et al 1999). Defects in NER results in the extreme photosensitivity and predisposition to cancer as seen in the syndrome xeroderma pigmentosum (XP) Cockayne syndrome (CS) and the photosensitive form of the brittle hair disorder trichothiodystrophy (TTD) (de Boer & Hoeijmakers 2000).

The general stages of NER are conserved across all three domains of life: lesion recognition, 3' and 5' sequential dual incisions, unwinding, patch repair synthesis and ligation, Figure 1.10. However, the bacterial and eukaryotic proteins involved show very little homology and many steps in mammalian NER are more complex than in other species.

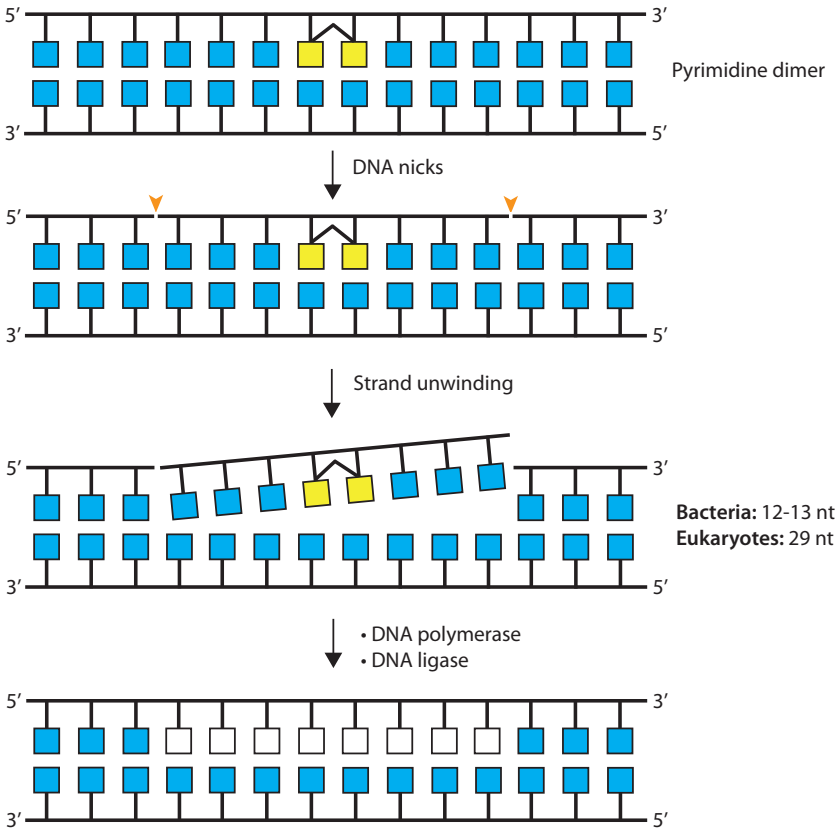


Figure 1.10: **Nucleotide excision repair. Nucleases** make incisions 3' and 5' to the DNA lesion shown here as a pyrimidine dimer. Helicases unwind the oligonucleotide 12-13 nucleotides in bacteria and 29 nucleotides in eukaryotes. Gap filling is carried out by DNA polymerase using the complementary strand of DNA as a template. The nick is sealed by DNA ligase.

In *E. coli,* a DNA lesion is first identified by the UvrA-UvrB protein complex, which has an A_2B_1 stoichiometry. Upon binding of the DNA lesion by the UvrA subunit, the helicase UvrB locally kinks and unwinds approximately 5 bp around the damaged DNA. UvrB binds tightly to the ssDNA lesion via a beta hairpin and isomerises and UvrA dissociates from the complex. UvrC then binds to the UvrB-DNA complex and makes first a 3' and then a 5' incision in the DNA, resulting in a 12-13 nucleotide oligomer. UvrD (helicase II) is

recruited to the nick sites, resulting in dissociation of the oligomer and UvrC. This gives access to DNA polymerase I to the 5' site hydroxyl residue, allowing it to synthesises DNA and repairs the gap; DNA ligase then seals the resultant nick (Petit & Sancar 1999).

NER in eukaryotes has the same overall steps as bacteria however a more complex set of proteins are involved. Eukaryotes also contain two sub pathways of NER; global genomic repair (GGR), which is used throughout the genome, and transcription coupled repair (TCR), which specifically deals with lesions that arrest the RNA polymerase in the transcribed strand of expressed genes (Hanawalt 1994). TCR is also found in bacteria.

The UvrABC counterpart in humans is composed of 16-17 proteins that are organised into 6 repair factors: XPA, RPA, TFIIH, XPC-RAD23B, XPG and ERCC1-XPF (Scharer 2013). In TCR, the NER recognition signal is a stalled RNA polymerase upstream (5') of the DNA lesion and it is thought that the CSA and CSB proteins (which are mutated in Cockayne syndrome) are responsible for recruiting the rest of the NER factors (Costa et al 2003). By contrast in GGR, XPC-RAD23B is the initial damage recognition factor with the XPC subunit recognising thermodynamically destabilised duplex DNA. Additionally, DDB1/DDB2 can recognise UV damaged DNA which in turn activates ubiquitin ligase that recruits XPC. RAD23B acts to stabilise XPC and dissociates from XPC upon damage recognition (Sugasawa et al 1996). The TFIIH factor consisting of XPB, p52, p8, p62, p34, p44 and CAK (cyclin activated kinase) subunits is then recruited to the XPC-RAD23B site. The two helicase subunits XPB and XPD are involved in unwinding the DNA around the lesion in a 3' to 5' and 5' to 3' direction respectively (Oksenych & Coin 2010, Tapias et al 2004). The association of XPD with the lesion allows for the assembly of the pre-incision complex consisting of XPA, RPA and XPG. XPA is thought to be a regulatory factor that maintains the other proteins in the complex while the single-strand DNA binding protein RPA co-ordinates excision and repair synthesis events. The endonuclease XPG makes a 3' incision generating a hydroxyl group that is used to initiate repair synthesis,

followed by a 5' incision by ERCC1-XPF. The dual incision results in a 24-32mer oligonucleotide containing the damaged DNA which then disassociates. The DNA is then filled by DNA polymerases δ and ε and assisted by the sliding clamp PCNA and the clamp loader RFC, and the resulting nick is then sealed by DNA ligase I.

Genomic sequencing has revealed that most archaea possess homologues of the eukaryotic NER proteins such as XPF, XPG, XPB, and XPD (Grogan 2000, White 2003). However, there are examples of archaea missing one or both of the XPB and XPD, suggesting that these two helicases do not have an essential cooperation during archaeal NER. Archaeal and eukaryotic XBP proteins share around 25% to 30% sequence identity and most crenarchaea contain multiple copies (Richards et al 2008a). The archaeal nuclease Bax1 co-transcribes with XPB and is thought to be the archaeal equivalent of the eukaryotic XPG. Two forms of archaeal XPF have been found: a long form containing nuclease and helicase domains known as Hef that is specific to the euryarchaea, and a short form found in crenarchaea and thaumarchaea that contains a nuclease domain only. (Komori et al 2004, Nishino et al 2005). The crenarchaeal XPF is dependent on the sliding clamp PCNA for its catalytic activity (Roberts et al 2003). The flap endonuclease NucS has also been suggested to play a role in archaeal NER, however its distribution amongst the archaea is sparser than that of XPF (Ren et al 2009).

Some mesophilic methanogens such as *Methanobacterium thermoautotrophicum* and some halophiles such as the *H. volcanii* possess orthologues of the bacterial UvrABC system (Costa et al 2003, Lestini et al 2010). Furthermore, some archaea such as *Methanosarcina mazei*, have a mixture of bacterial UvrABC and eukaryotic XPF orthologues (White 2003). A possible explanation for this observation is that the original NER pathway in archaea was eukaryotic in nature but has been replaced in some species by a more bacterial system through lateral gene transfer (Kelman & White 2005). This could mean that two NER pathways are acting in tandem or that archaeal NER proteins have additional functions within the cell. However, since an

archaea-specific NER system has not been demonstrated *in vivo* or *in vitro*, NER in the archaea is largely unknown.

It is unclear how DNA damage is detected in the archaeal NER systems that do not possess the UvrABC system, since no XPC-RAD23B homologues have been identified. It has been suggested that the single stranded binding protein SSB (also known as RPA in eukarya and some archaea) could fulfil this role. RPA is seen to be upregulated in *Halobacterium* after treatment with UV radiation, and SSB in *S. solfataricus* is capable of recognising DNA damage such as mismatches, photoproducts and bulky lesions *in vitro* (Cubeddu & White 2005, McCready et al 2005).

Mismatch repair

DNA mismatch repair (MMR) maintains genomic stability by correcting mismatched base pairs that arise mainly from errors during DNA replication. The overall mechanism and proteins involved in MMR are highly conserved across most species of bacteria and eukaryotes, but only in a limited set of archaeal species. Post-replicative MMR is typified by 'long-patch' mechanism where a long oligonucleotide is excised during repair. MMR in *Escherichia coli* increases the accuracy of DNA replication by 20 to 400 times (Schaaper 1993). In humans, mutations and epigenetic silencing of the MMR proteins have been linked to hereditary non-polyposis colon cancers (Jiricny 2006).

Two types of MMR mechanisms are known; the first mechanism appears to be specific to *E. coli* and other closely related bacteria; mismatched DNA in this system is recognized by the absence of DNA methylation at GATC sequences to restrict DNA repair to the newly synthesised DNA (Fukui 2010). Immediately following DNA synthesis, the daughter strand remains unmethylated for up to 2 minutes, due to the slow rate of methylation by the Dam methylase. The MutS homodimer recognises the mismatched DNA, which in turn recruits MutL to activate the restriction endonuclease MutH. MutH nicks the unmethylated DNA strand at a hemimethylated GATC site and generates an entry point for the excision reaction to occur (Smith & Modrich 1996). The error-containing strand is nicked by the exonucleases ExoVII or

RecJ at 5' to the DNA damage or by ExoI or ExoX at 3' to the DNA damage, the oligonucleotide is then unwound by the UvrD helicase (Li 2008). The new oligonucleotide is synthesised by DNA polymerase III and the nick is sealed by DNA ligase. Homologues of the *E. coli* MutS and MutL are present in most bacteria and eukaryotes, however no widespread presence of the MutH homologue has been identified.

The second MMR mechanism is employed by eukaryotes and the majority of bacteria, in this mechanism errors in the mismatched DNA duplex are recognised by strand discontinuities. In eukaryotes, the MutSα sliding clamp recognises base-base mismatches and insertions/deletions of up to two nucleotides, and the sliding clamp MutSβ can identify larger nucleotide insertions and deletions (Jiricny 2006). MutLα then nicks at either the 3' or 5' side of the mismatch on the discontinuous DNA strand (Kadyrov et al 2006, Modrich 2006). This DNA segment is then excised by EXO1 exonuclease in concert with ssDNA binding protein RPA, a new oligonucleotide is synthesised by DNA polymerase δ and the gap is sealed by DNA ligase 1 (Genschel & Modrich 2009). A similar process occurs in bacteria lacking MutH; mispaired bases and short insertion and deletion loops are recognised by MutS and nicked by MutL. The error-containing DNA is removed by the helicases such as UvrD and RecJ, and ExoI exonucleases with cooperation from single stranded DNA binding protein SSB. DNA polymerase III fills the gap and the nick is sealed by DNA ligase.

Several hyperthermophilic archaea such as *Sulfolobus spp*. exhibit effective mutation-avoidance mechanisms in response to DNA damaging agents, however hyperthermophiles on a whole lack any MMR machinery with homology to the eukaryotic and bacterial MutS and MutL system (Grogan 2004). This suggests that an alternative repair pathway is present in these archaea. In fact phylogenetic analysis has detected MutS and MutL orthologues in only nine halophilic and methanogenic archaeal species from the phylum Euryarcheaota, which appear to have arisen from bacteria via horizontal gene transfer (Lin et al 2007). The newly discovered endonuclease EndoMS which

is also known as NucS from *Pyrococcus furiosus* is capable of cleaving mismatched bases, suggesting the presence of a novel archaeal mismatch repair mechanism initiated by double strand breaks (Ishino et al 2016). *H. volcanii* encodes two homologs of MutL and four of MutS, which have been shown to function in MMR in *H. volcanii* (Adit Noar, Tel Aviv University and Stéphane Delmas, Sorbonne University, Paris, personal communication).

1.4.2.3 Double strand break repair

DNA double strand breaks (DSBs) can be induced by ionising radiation or by naturally-occurring metabolic products and reactive oxygen species that accumulate within the cell. DSBs are also intentionally formed during meiosis in eukaryotes and during V(D)J recombination for the development of T-cells and B-cells in the vertebrate immune system (Davis & Chen 2013). Failure to correctly repair DSBs results in chromosome breakage and rearrangements, which leads to cellular death and predisposition to cancers and immune system disorders (Hefferin & Tomkinson 2005).

Three repair pathways are involved in the repair of double strand breaks: non-homologous end-joining (NHEJ), microhomology-mediated end joining (MMEJ) and homologous recombination (HR), Figure 1.11. End joining is a rapid way to repair DNA breaks but these mechanisms are inherently error-prone. Converesely, HR is a slower method of DNA repair but is more accurate as an undamaged homologous DNA sequence is used as a template for repair.

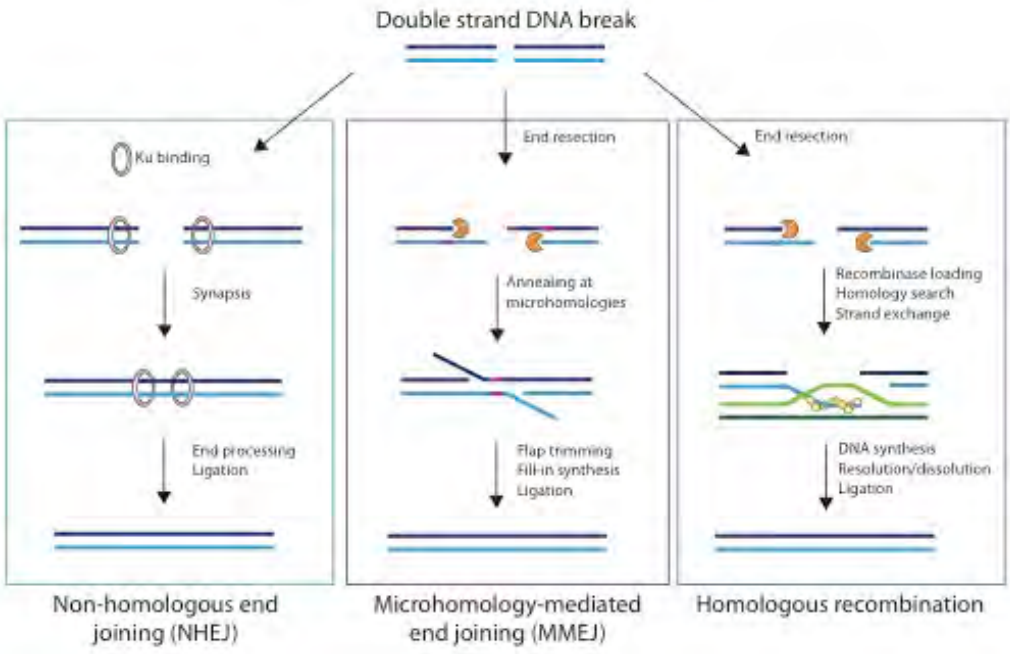


Figure 1.11: **Double-strand break repair pathways.** Three pathways exist to repair double strand breaks. During non-homologous end joining (NHEJ) Ku binds to the dsDNA ends at the site of a double strand break and recruits bridging factors as well as polymerases and exonucleases to process the DNA ends. DNA ends are then ligated to repair the DSB. During microhomology-mediated end joining (MHEJ) DNA ends are processed to reveal microhomologies, which then align and anneal. Gaps are filled by translesion polymerases and then the gap is ligated. During homologous recombination DNA ends are resected and recombinases are loaded onto the ssDNA. The recombinase-DNA filament invades an intact homologous DNA molecule and uses it as a template for DNA repair.

Double-strand break repair pathway choice

In eukaryotes the choice between HR and NHEJ is largely influenced by the cell cycle. In S and G2 phase, HR predominates because following DNA replication, a homologous sister chromatid available as a template for repair. NHEJ occurs in all stages of the cell cycle but unlike HR and MMEJ, NHEJ does not involve end resection and so it is the preferred mechanism in G1 phase when end resection activity is low within the cell (Ira et al 2004). The choice between NHEJ and MMEJ is largely influenced by the NHEJ DNA tethering protein Ku and the cell cycle; Ku has a higher affinity for DNA ends than the MMEJ equivalent PARP1, explaining the predominance of NHEJ over MMEJ.

The Mre11-Rad50 complex also plays a role in DNA repair pathway choice. Mre11 is a single stranded DNA nuclease and Rad50 is an ATPase with a long

coiled domain; this coiled domain contains a zink hook motif at the tip allowing for intermolecular interactions between other Rad50 molecules, DNA or chromosomes (de Jager et al 2001, Hopfner et al 2002, Paull & Gellert 2000). In vertebrates the Mre11-Rad50 complexes with Xrs2 or with Nbs1 in yeast. Archaea contain a homologue of Mre11Rad50 homologue but not Xrs2. The Xrs2/Nbs1 protein mediates interactions between the Mre11Rad50 complex and other DNA repair proteins and acts to recruit Mre11-Rad50 to sites of DNA damage through interactions with histone γH2AX (Kobayashi et al 2002). The Mre11 complex tethers DSBs and controls end-resection; if end-resection occurs, then homologous recombination or MMEJ is promoted and limited end processing promotes NHEJ (Symington & Gautier 2011). In *H. volcanii* Mre11-Rad50 restrains recombination at double strand breaks allowing time for pathways such as MMEJ to repair the DNA ends rather than HR (Delmas et al 2009).

In mammalian cells, the ATM (ataxia-telangiectasia mutated) and ATR (ATM- and Rad3-Related) proteins alongside protein kinases can also direct DNA repair pathways. ATM detects DNA double strand breaks leading to the recruitment of the Mre11-Rad50-Nsb1 complex (Marechal & Zou 2013). ATR interacts with ATRIP (ATR-interacting protein) to sense ssDNA generated by processing of DSBs, as well as ssDNA present at stalled replication forks. Both ATM and ATR initiate signalling cascades that involve the checkpoint kinases Chk1 and Chk2 that initiate a secondary wave of phosphorylation events in a large singling network (Matsuoka et al 2007). One phosphorylation target is BRCA1 that is a promoter of homologous recombination (Sancar et al 2004).

Non-homologous end-joining

Non-homologous end-joining (NHEJ) has the potential to be less accurate at repairing DSBs. This pathway often results in minor changes of the DNA sequence at the break site and occasionally leads to the joining of previously unlinked DNA molecules, leading to chromosomal rearrangements (Davis & Chen 2013). Conserved across eukaryotes, some bacteria and a very limited number of archaea, the basic mechanism of NHEJ involves the recognition and

bringing together of broken ends of DNA followed by a processing and ligation of the ends and disassembly of the NHEJ complex.

In eukaryotes, DSBs are first recognised and bound by the Ku heterodimer, composed of Ku70 and Ku80 subunits which form a ring shaped structure that encircles the dsDNA molecule (Walker et al 2001). At both ends of the DSB, Ku70/80 forms a scaffold to which the NHEJ factors are recruited (Davis & Chen 2013). The first NHEJ factor is the DNA-dependent protein kinase catalytic subunit (DNA-PKcs) that is involved in forming a synaptic complex, which brings both of the DNA ends together. DNA polymerases μ and λ fill in or exonucleases such as Artemis remove single stranded non-compatible overhangs, this is the step where loss of nucleotides can occur. The ligase IV and XRCC4 complex then catalyses the ligation of the processed ends which is aided by the XRCC4-like factor (XLF).

Ku is a major marker for the presence of a NHEJ pathway, however Ku genes are not present in all bacteria and are absent from the *E. coli* K12 strain. No obvious phylogenetic pattern is observed between bacterial species that possess or lack NHEJ apparatus, suggesting that NHEJ systems are acquired by horizontal gene transfer events (Bowater & Doherty 2006). At 30-40 kDa, bacterial Ku proteins are smaller than the eukaryotic counterparts and are predominantly homodimeric (Aravind & Koonin 2001, Doherty et al 2001). Bacterial Ku genes are usually found in operons alongside DNA dependent ligases such as LigD, forming a species-specific NHEJ complex (Weller et al 2002). The Ku associated ligases have shown to contain Pol X family polymerase domains and exonuclease activity, and are suggested to play a role in processing non-compatible DNA ends (Della et al 2004, Pitcher et al 2005).

Some NHEJ like genes have been identified in archaea, homologues of bacterial LigD phosphoesterase domain have been observed in seven species of the Euryarchaeota (Nair et al 2010, Smith et al 2011). Homologues of Ku have also been identified in several species of archaea such as *Archaeoglobus*

fulgidus, however the NHEJ apparatus appears to be absent in *H. volcanii* (Aravind & Koonin 2001, Bowater & Doherty 2006).

Microhomology-mediated end joining

An alternative mechanism of NHEJ is known as microhomology-mediated end joining (MMEJ), which involves the alignment and annealing of microhomologous sequences at the ends of a DNA break before ligation. MMEJ is error prone and can lead to deletions, insertions and chromosomal translocations at the DNA break point.

Similar to the role of Ku in NHEJ, poly ADP-ribose polymerase 1 (PARP1) tethers DNA ends at the site of DSBs, which are then resected in a 5' to 3' direction to expose microhomologous sequences. In mammalian and *Saccharomyces cerevisiae* cells, the Mre11–Rad50–Nbs1 and Mre11–Rad50–Xrs2/Sae2 complexes are responsible for the initiation of end resection by nicking the strand at the site of a DSB and carrying out 3'-5' resection, a ssDNA tract is then created by the 5'-3' exonuclease Exo1 (Symington & Gautier 2011). In mammalian cells an 1-18 nt ssDNA tract is formed during end resection and the ssDNA is maintained by binding of RPA (Sfeir & Symington 2015). Following annealing of resected ends by complementary base pairing, ssDNA gaps are filled in by the translesion DNA polymerase Polθ. Polθ, encoded by the gene *POLQ*, contains a C-terminal proof reading deficient polymerase domain and an N-terminal helicase domain that shows homology with the human Hel308 helicase HelQ. The helicase domain has shown to be critical in the regulation of DNA end joining (Ceccaldi et al 2015, Chan et al 2010, Marini & Wood 2002). Hel308 will be discussed in further detail in Section 1.7. After DNA synthesis, heterologous 3' flaps are removed by the endonuclease XPF–ERCC1 complex and the DNA ends are ligated by DNA ligases Lig1 and Lig3.

In some bacteria, the multidomain protein LigD composing of polymerase, nuclease and ligation domains carries out the majority of MMEJ (Pitcher et al 2007). MMEJ is observed in *H. volcanii* where the Mre11-Rad50 complex

appears to prevent the repair of double strand breaks by homologous recombination and this promoting MMEJ (Delmas et al 2009). Compared to eukaryotes, MMEJ is largely understudied in prokaryotes

Homologous recombination

Homologous recombination (HR) is considered to be an error-free mechanism to repair DNA double strand breaks. During homologous recombination a damaged DNA strand uses an homologous sister chromosome (usually in S or G2 phase cells) as a template. The understanding of HR is essential to this study, and the mechanism and role that HR plays in the repair of DNA lesions including DSBs and genomic stability will be discussed in further detail below.

1.5 Homologous recombination

Homologous recombination (HR) involves the genetic exchange between homologous DNA sequences and is regarded as an error-free mode of DNA repair. HR is critical for the repair of many types of DNA lesions such as double strand breaks, single strand DNA gaps and interstrand DNA crosslinks. HR is involved in the restart of stalled or broken replication forks, as well as for chromosomal pairing and exchange during meiosis (Michel et al 2001, Neale & Keeney 2006). The inability to repair DNA damage and resolve DNA replication stress in a timely fashion leads to genomic instability, which is a contributing factor to the development of cancer. For example, mutations in the BRCA1 and BRCA2 genes leads to a predisposition to breast and ovarian cancers (Venkitaraman 2002). Fanconi anaemia is a cancer predisposition syndrome, which is typified by the inability to repair DNA interstrand crosslinks and is caused by failures in FA proteins such as FancD1, FancD2 and FancJ (Takata et al 2006). Mutations in the Bloom's syndrome RecQ helicase BLM promotes excessive crossing over of sister chromatids leading to elevated levels of chromosomal rearrangements, genomic instability and a predisposition to cancer (Wu & Hickson 2003). Mutations in the Werner's syndrome RecQ helicase WRN results in defects in HR resolution and cell division, leading to genetic instability, cancer susceptibility and premature aging (Saintigny et al 2002). Since unrestrained HR can lead to undesirable DNA rearrangements, many regulatory mechanisms have evolved to ensure

that HR occurs accurately and at the correct time and place within the cell, this is discussed later in Section 1.4.3.6.

The overall mechanism of HR is conserved across all three domains of life and can be broken down in to three main stages: pre-synapsis, synapsis and post-synapsis. During pre-synapsis, DNA is resected to generate ssDNA 3' overhangs and recombinases are loaded. In synapsis, recombinases polymerise onto the ssDNA, which then undergoes a homology search to find an intact homologous DNA sequence. Once found, strand exchange is catalysed and the ssDNA molecule invades the dsDNA homologous sequence forming a D-loop. During the final stage of post-synapsis, the DNA complex is resolved allowing the repair of DNA or restart of the stalled replication fork. The resulting products of HR can either be crossover, where an exchange of genetic material occurs between the two DNA molecules, or non-crossover where no genetic exchange occurs, Figure 1.12. A detailed account of each stage of homologous recombination will be discussed in Sections 1.4.3.1 to 1.4.3.3.

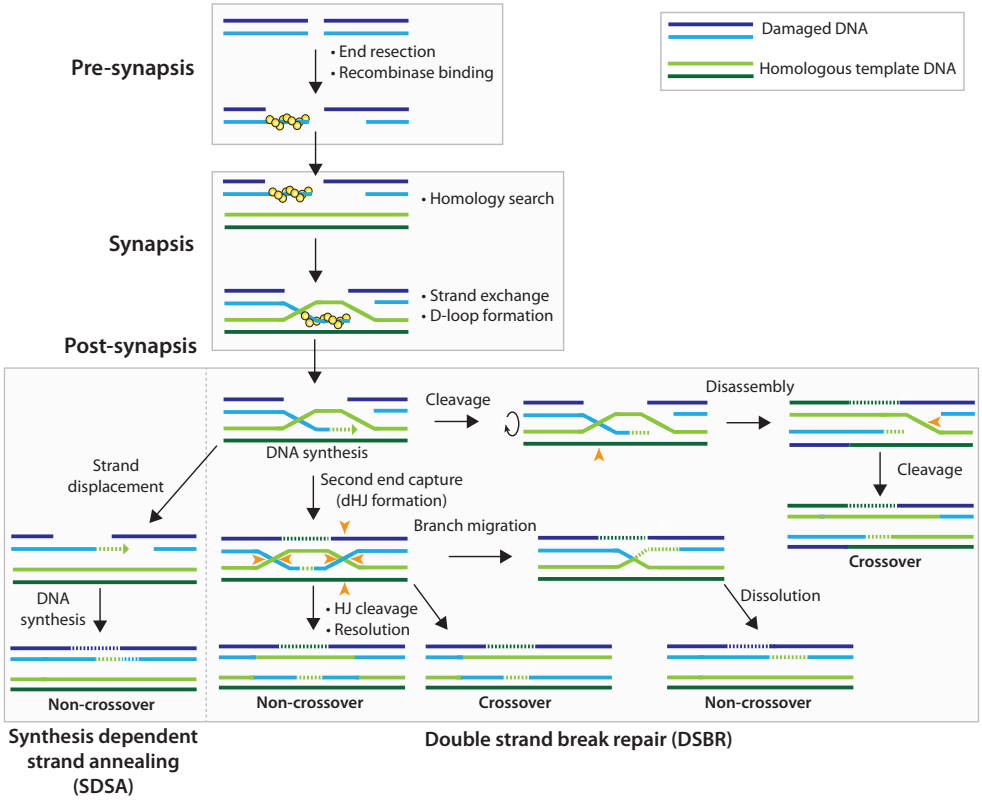


Figure 1.12: **Homologous recombination pathways.** A schematic of homologous recombination using the repair of double strand breaks as an example. Recombination occurs in three stages: pre-synapsis, synapsis and post-synapsis. Yellow spheres indicate recombinases, dashed lines indicate newly synthesised DNA, Orange arrowheads indicate cleavage events.

1.5.1 Pre-synapsis

During pre-synapsis, dsDNA is resected in a 5' to 3' direction to generate ssDNA 3' overhangs. Recombinases are loaded onto the 3' ssDNA tails displacing the single stranded binding proteins; the recombinase in bacteria is RecA, in eukaryotes it is Rad51 (DMC1 during meiosis) and in archaea it is RadA, Figure 1.13.

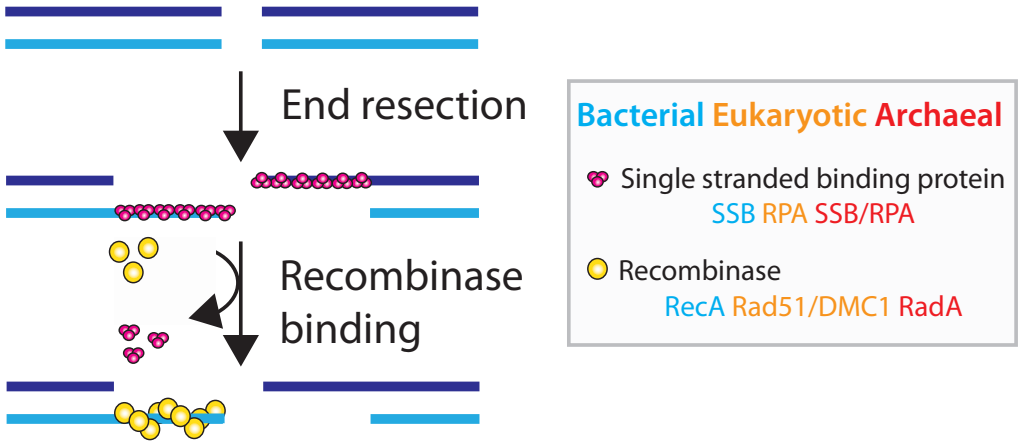


Figure 1.13 **Homologous recombination pre-synapsis.** Displacement of single stranded binding protein and recombinase nucleoprotein filament formation upon 3' ssDNA tails following end resection of DSBs. Recombinase filament formation is a dynamic and reversible process.

Bacterial pre-synapsis

HR in *E. coli* can either be initiated by the RecBCD or by the RecFOR pathway. Both pathways generate ssDNA onto which the bacterial recombinase RecA is loaded prior to invasion of a homologous DNA molecule (Rocha et al 2005). RecBCD binds dsDNA ends and acts to repair DSBs, the RecBCD holoenzyme contains helicase and nuclease functions that unwind and degrade dsDNA until it reaches a χ site. A χ site is a regulatory DNA sequence (5'-GCTGGTGG-3') that acts to attenuate the 3' nuclease activity of RecBCD (Dixon & Kowalczykowski 1993). At the χ site the helicase activity of RecBCD is promoted to generate 3' ssDNA onto which the recombinase RecA is loaded (Kowalczykowski 2000, Singleton et al 2004).

RecFOR binds to gapped ssDNA and displaces single strand binding protein (SSB) to allow RecA binding (Morimatsu & Kowalczykowski 2003). If necessary, the endonuclease RecJ acts alongside RecFOR to enlarge the ssDNA region. In both the RecBCD and RecFOR pathways, RecA forms a filament with right-handed helical geometry, with six RecA molecules and 18 nucleotides per turn on the ssDNA that is able to slide along dsDNA molecule in search for homology during synapsis (Ragunathan et al 2012).

Eukaryotic pre-synapsis

In yeast, 3' ssDNA tails are generated via end resection by the Mre11/Rad50/Xrs2 (MRX) complex along with Sae2 and by the Mre11/Rad50/Nbs1 (MRN) complex along with CtIP in mammals (Jasin & Rothstein 2013). BRCA1 interacts with MRN and CtIP and promotes HR and SSA, suggesting that BRCA1 plays a role during end resection (Stark et al 2004). If extensive resection is required, this is performed by the exonuclease Exo1 or by the helicase/nuclease Sgs1/Dna2 (Mimitou & Symington 2008).

The eukaryotic homologue of bacterial RecA is Rad51, which is loaded on to ssDNA that has been generated from resecting 5' strands at DSBs or has arisen from perturbations in DNA replication. Rad51 displaces the already bound RPA with assistance from the recombination mediators Rad52, Rad54 and Swi5-Sfr2 (Benson et al 1998, Kurokawa et al 2008, Mazin et al 2010). In yeast the recombination mediator is the Rad55/57 heterodimer, which promotes the stability of the Rad51 presynaptic filaments (Krejci et al 2002, Sung 1997). BRCA2 is also suggested to be a recombinase mediator in mammals by promoting the assembly of Rad51 onto ssDNA in preference over dsDNA, and also aiding Rad51 in displacing RPA (Jensen et al 2010). Rad51 is loaded onto the ssDNA in a right-handed helical geometry, with six Rad51 molecules and 18 nucleotides per turn (Chen et al 2008). The ssDNA within the nucleofilament is stretched by 50%, which aids in efficient homology search during synapsis (Klapstein et al 2004). In eukaryotes, Dmc1 is the meiosis-specific recombinase that is able to displaces RPA and form filaments

upon ssDNA. This action catalyses HR between homologous chromatids at programmed DSBs during meiosis (Sehorn et al 2004).

Archaeal pre-synapsis

The eukaryotic HR initiating proteins Mre11 and Rad50 are conserved in archaea. In thermophilic archaea, Mre11-Rad50 is commonly found within an operon containing the hexameric helicase HerA and the 5' to 3' nuclease NurA (Constantinesco et al 2002, Constantinesco et al 2004, White 2011). In *Pyrococcus furiosus* Mre11-Rad50 was shown to generate short 3' overhangs on which the HerA-NurA complex initiates end resection from to generate 3' ssDNA tails (Hartman et al 2010). However in *H. volcanii* Mre11-Rad50 appears to delay the repair of DSBs by HR, suggesting that Mre11-Rad50 acts to control the entry into the HR pathway following DNA damage (Delmas et al 2009). Additionally, HerA and NurA are not present in *H. volcanii*. Once end resection has occurred, the SSB/RPA homologue binds to the ssDNA but must be displaced to allow the loading of the archaeal recombinase RadA. The deletion of *radA* in *H. volcanii* leads to severe growth, DNA repair and recombination defects (Delmas et al 2009). Most archaeal species contain at least one copy of the RadA paralogue, RadB. RadB is not involved directly in strand exchange but does assist RadA in forming a nucleoprotein filament on ssDNA. Strains deleted for *radB* in *H. volcanii* show decreased levels of recombination and increased levels of sensitivity to DNA damaging agents, but to a lesser extent as a *radA* deletion (Guy et al 2006). Suppressor mutations S101P and A196V in RadA alleviate the phenotypic effect of a *radB* deletion, suggesting that RadB induces conformational changes in RadA. These changes promote RadA polymerisation on ssDNA during pre-synapsis, Figure 1.14. Furthermore, protein;protein interaction studies in *Pyrococcus furiosus* and *H. volcanii* have shown that RadA and RadB interact *in vivo* (Komori et al 2000, Wardell 2013).

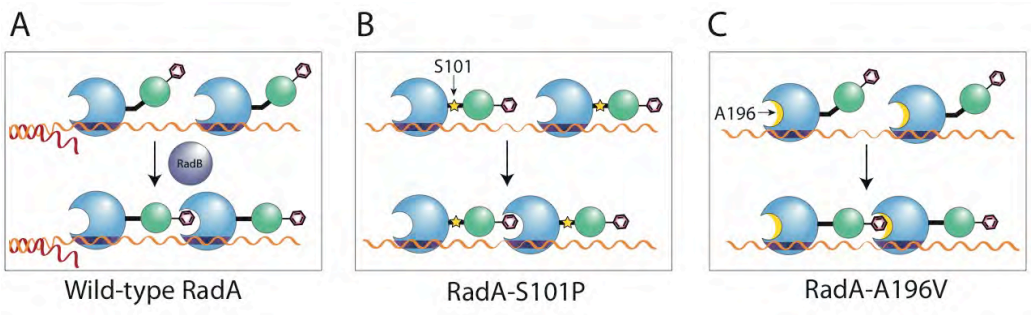


Figure 1.14: ***ΔradB* suppressors *radA-S101P* and *radA-A196V*. (A)** Wild-type RadA may not be in the correct conformation to form a nucleoprotein filament on ssDNA. RadB may be required for efficient polymerisation. **(B)** RadA-S101P may lock RadA into a polymerisation-competent position and therefore not require RadB. S101P is indicated by yellow star. **(C)** RadA-A196V increases the hydrophobicity of the RadA binding pocket, which could result in stronger hydrophobic interactions between RadA monomers, negating the need for RadB. A196V indicated by yellow arc. RadA core domain in blue and N-terminal domain in green. Adapted from (Li et al 2008, Wardell 2013).

1.5.2 Synapsis

The core reaction of strand exchange in homologous recombination occurs during synapsis. Strand exchange is catalysed by the recombinases, RecA in bacteria, Rad51/Dmc1 in eukaryotes and RadA in archaea. The recombinase facilitates a physical connection between the invading ssDNA and the homologous duplex DNA template, leading to the formation of a heteroduplex D-loop (McEntee et al 1979). In eukaryotes, the homology search catalysed by the Rad51 ssDNA filament is assisted by Rad54 and Rhd54, which facilitate the sliding of the ssDNA along homologous duplex DNA (Krejci et al 2012).

1.5.3 Post-synapsis

Recombination intermediates generated by strand exchange during synapsis are processed by several pathways, which will be described below. Resolution of homologous recombination leads to either a crossover product where genetic material is exchanged between two homologous DNA molecules or a non-crossover product, the latter is also known as a gene conversion.

1.5.3.1 Synthesis-dependent strand annealing

One way to resolve recombination intermediates following strand invasion is by synthesis-dependent strand invasion (SDSA), Figure 1.15, which can occur

during both mitotic and meiotic DSB repair in eukaryotes (San Filippo et al 2008). It is suggested that there are two waves of repair during HR; the first is SDSA that leads to only non- crossover products which reduces the likelihood of genomic rearrangements. The second wave follows the route of double strand break repair (DSBR) forming Holliday junctions and is mainly characterised by crossover products.

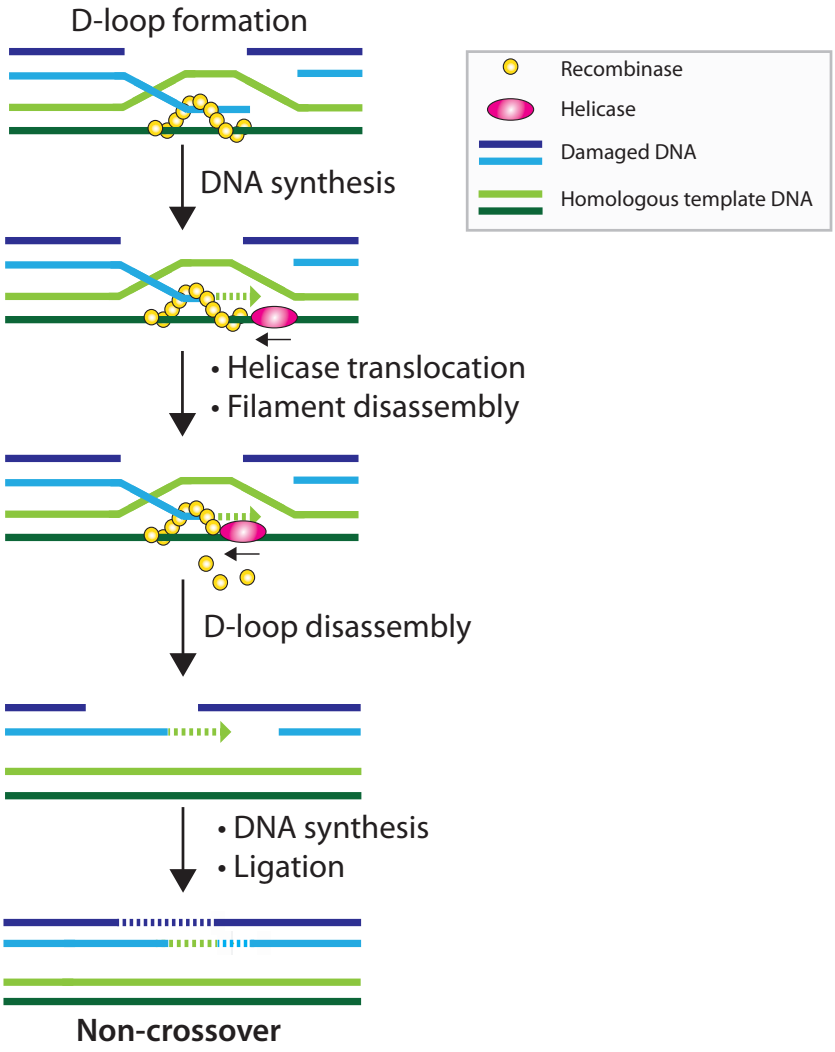


Figure 1.15: **Synthesis-dependent strand annealing (SDSA).** Following strand invasion, DNA is synthesised and helicases act to disrupt the recombinase nucleoprotein filament. Resulting in the disassembly of the D-loop. The gaps are filled and ligated resulting in a non-cross over product.

During SDSA the extended D-loop is reversed through the action of helicases leading to the annealing of the newly synthesised DNA strand with the resected strand of the second end of the DSB. In yeast, the Srs2 helicase dissociates bound Rad51 from D-loops and therefore is thought to promote SDSA. It is

suggested that Srs2 activity is directly stimulated by Rad51 bound to dsDNA (Ira et al 2003, Krejci et al 2003, Paques & Haber 1999). The yeast Mph1, a homologue of the human FANCM translocase, is able to disrupt Rad51 coated D-loops; Mph1 in particular is able to displace the extended primers during DNA synthesis within D-loops (Prakash et al 2009). The antirecombinase RecQ helicases RecQ5, BLM and FANCJ are also able to disrupt Rad51 filaments, suggesting that they could promote the resolution of recombination intermediates via the SDSA pathway (Bugreev et al 2007, Hu et al 2007, Sommers et al 2009)

1.5.3.2 Double-strand break repair

A second way to resolve recombination intermediates is by the double strand-break repair (DSBR) pathway. This mechanism involves the capture of the second end of the DSB, which in eukaryotes depends on the Rad52 mediator protein, and the formation of a double Holliday junction (dHJ) (McIlwraith & West 2008). A Holliday junction is a branched structure composed of two linked DNA duplexes in a cross-shaped configuration. To complete DNA repair, these two DNA molecules require separating. This occurs either by cleavage by endonucleases that resolve the HJ or by helicase unwinding which leads to dissolution of the HJ.

Holliday junction resolution

Holliday junctions are resolved by endonucleases (also known as resolvases) that are specialised to cleave branched DNA structures. If these resolvases cleave a dHJ in the same sense this results in non-crossover DNA products whereas if cleavage occurs in opposite sense strands of a dHJ then crossover products are formed, Figure 1.16.

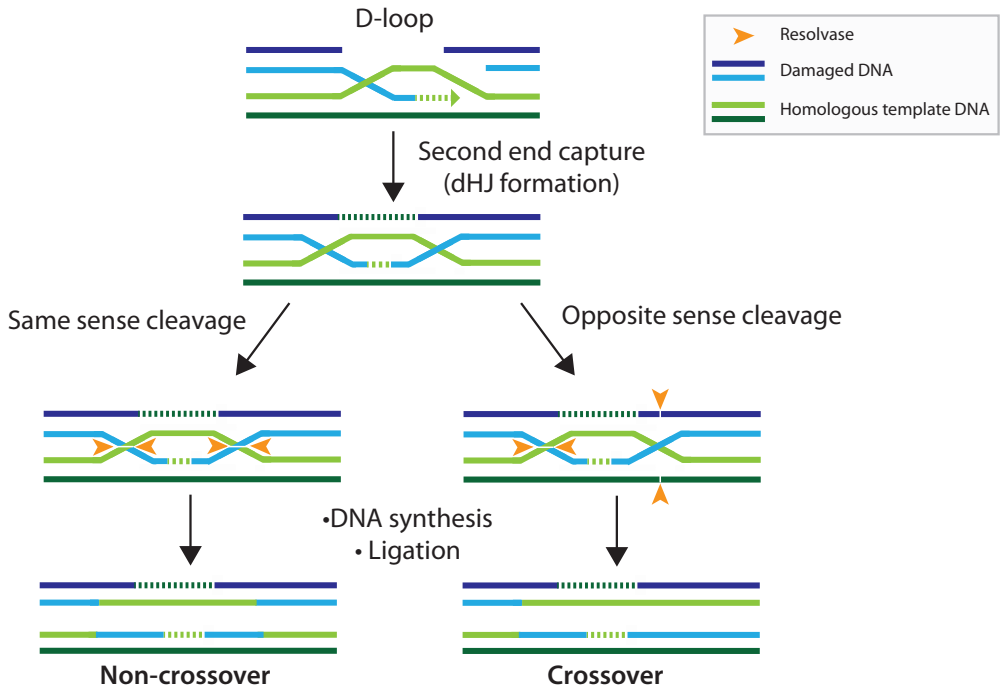


Figure 1.16: **Resolution of double-Holliday junctions.** Same sense cleavage by resolvases results in a non-crossover recombination event. Opposite sense cleavage results in the production of a crossover recombination event.

In *E. coli*, Holliday junctions are resolved by the RuvABC complex. The RuvA protein constrain the junction in a square planar configuration and the RuvB helicase catalyses branch migration, Figure 1.17. RuvC cleaves the Holliday junction intermediates into duplex products at specific target sequences by making a dual symmetric incision (Sharples et al 1999). Symmetric cleavage results in perfect nicked duplexes that can be directly ligated (West 1997). Homologs of RuvA and RuvB are highly conserved across bacteria, however RuvC is less conserved. It has been suggested that species lacking RuvC could resolve Holliday junctions using RusA. The helicase RecG can also perform branch migration, processing of stalled replication forks and plays a role in the directing of DNA synthesis during HR (Azeroglu et al 2016, Singleton et al 2001)

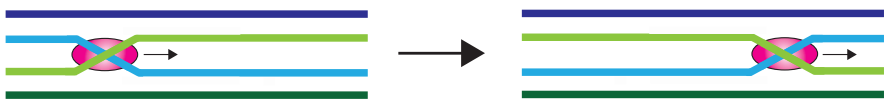


Figure 1.17: **Branch migration.** Helicases migrate the Holliday-junctions along the duplex.

HJ resolution in eukaryotes is not as straightforward as that catalyzed by the RuvABC system in bacteria. Multiple Holliday junction resolution pathways are present in eukaryotes and their use depends on the species. A number of DNA endonucleases have been proposed to carry out Holliday junction resolvase activity in eukaryotes, such as: Mus81–Mms4 (Eme1), Slx1–Slx4 (BTBD12/Mus312), XPF– ERCC1, Yen1 (GEN1) and Sgs1/MutLγ/ExoI (Schwartz & Heyer 2011). The resolution of joint DNA complexes is thought to be a complex and multistep process where HR intermediates are nicked in sequential steps by endonucleases (Matos et al 2011).

Hjc is an archaeal HJ resolving endonuclease and has shown to have analogous resolving properties to the *E. coli* protein RuvC. Like to RuvC, Hjc cuts Holliday junctions symmetrically (Bolt et al 2001) and the Hje endonuclease in *Sulfolobus solfataricus* has shown to resolve four-way DNA junctions by introducing a pair of nicks in the stacked-X form of the Holliday junction. Hjr from *Pyrococcus furiosus* on the other hand is able to cleave all four arms of a Holliday junction (Kvaratskhelia et al 2001). In addition to Hjc, *H. volcanii* contains the second Holliday junction nuclease Hef. Neither Hjc or Hef is essential, and strains deleted for either nuclease shows no growth defects and only mild sensitivity to MMC. However either nuclease becomes essential for cell survival when the other is deleted, suggesting a redundancy in functions of Hjc and Hef. Hjc acts exclusively in the HR pathway whereas Hef does not (Lestini et al 2010).

Asymmetric cleavage

A second route of resolving Holliday junctions involves the asymmetric cleavage of HR intermediates. This step occurs before a double HJ is ligated and results in crossover DNA products, Figure 1.18.

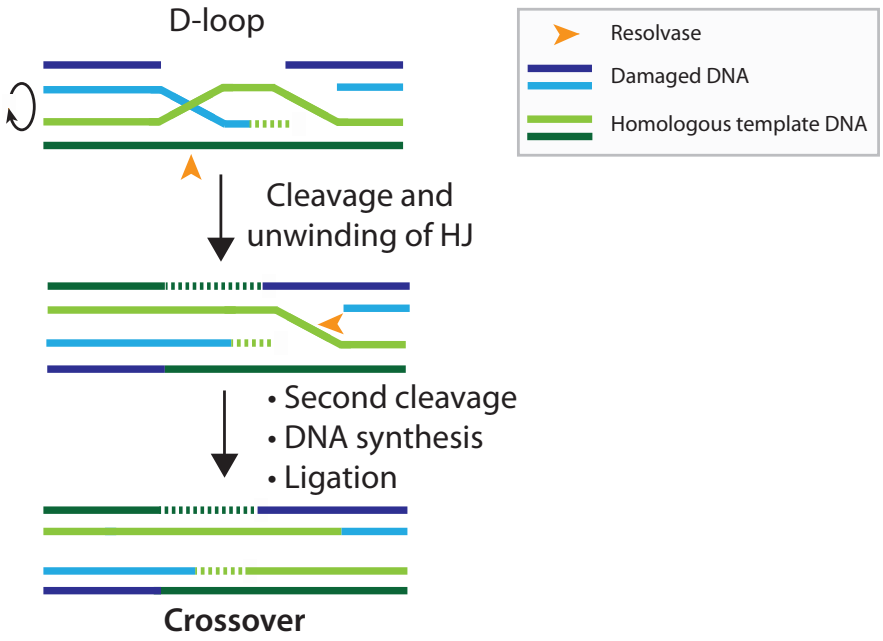


Figure 1.18: **Asymmetric cleavage of D-loops.** A single Holliday junction is cleaved asymmetrically and the D-loop is then unwound. A second asymmetric cleavage event occurs followed by DNA synthesis and ligation, resulting in a crossover DNA product.

In eukaryotes, the endonuclease module Slx1–Slx4 generates a nicked Holliday junction intermediate that can then be targeted by a second endonuclease complex: Mus81-Eme in fission yeast or by Mus81-Mms4 in budding yeast (Nishino et al 2005, Schwartz & Heyer 2011, Svendsen et al 2009). This asymmetric cleavage results in gapped duplex DNA crossover products that are then filled in by DNA polymerases and ligated.

The Mus81 homologue in archaea is Hef, which is able to cleave flaps and forked structured DNA. Hef is homologous to FANCM and consists of an N-terminal helicase domain and a C-terminal nuclease domain linked by a helix-hairpin-helix DNA binding domain, and acts to resolve Holliday junctions by forming homodimers (Komori et al 2002, Meetei et al 2005). Hef in *H. volcanii* is shown to be essential in backgrounds where the Holliday junction resolvase Hjc is deleted (Lestini et al 2010). Hef was also shown to form increased amounts of localised foci under replication stress suggesting that Hef may enhance replication fork stability by directly interacting with collapsed replication forks (Lestini et al 2013).

Holliday junction dissolution

An alternative mechanism to process double Holliday junctions is by unwinding of the structure using helicases to generate DNA hemi-catenanes, which are then processed by topoisomerases. This pathway does not rely on endonuclease action like in Holliday junction resolution methods and generates exclusively non-crossover DNA products, Figure 1.19 (Schwartz & Heyer 2011).

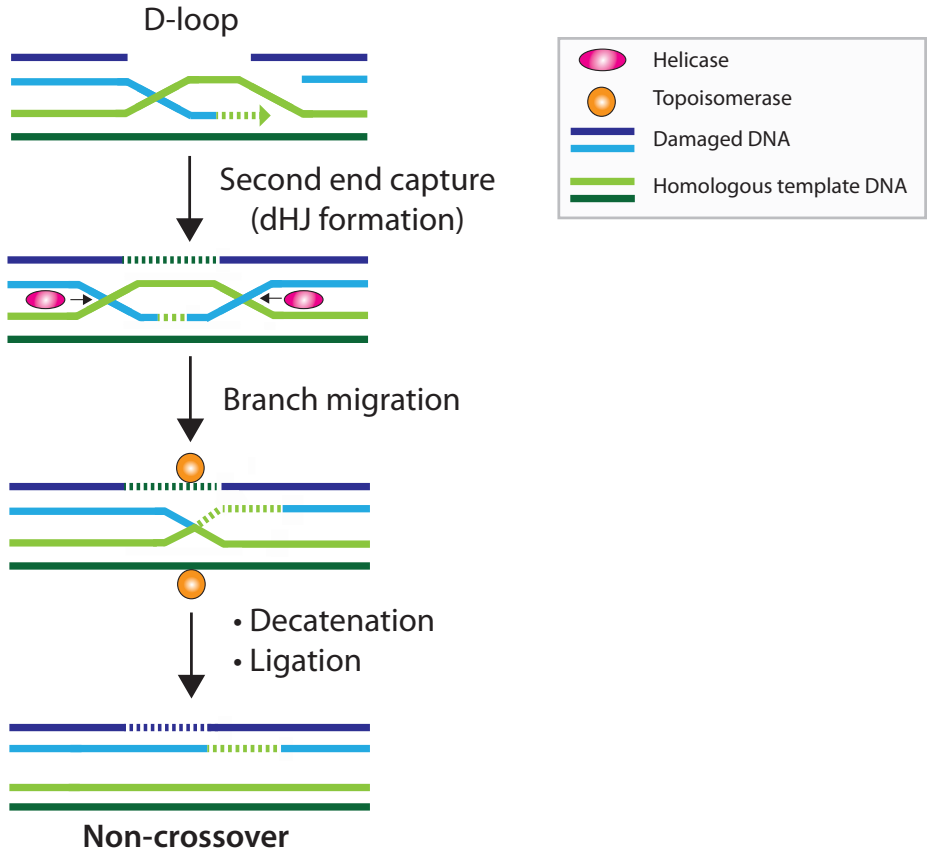


Figure 1.19: **Dissolution of a double Holliday junction.** Double holiday junctions can be migrated together by the action of two helicases to form a hemi-catenane. Once the two Holliday junctions have been merged the intermediate DNA complex is unwound by topoisomerases resulting in a non-crossover DNA product.

RecQ family 3' to 5' helicases are responsible for migrating dHJs to form DNA hemi-catenanes during Holliday junction dissolution. In *E. coli* this helicase is RecQ, in *Saccharomyces cerevisiae* it is Sgs1, in *Schizosaccharomyces pombe* it is Rqh1, and it is BLM in humans (Bizard & Hickson 2014). Defects in the BLM helicase cause Bloom's syndrome, a disorder that leads to cancer predisposition. In *Saccharomyces cerevisiae* and humans the hemi-catenane is

then processed by the topoisomerases Top3 and TOPOIIIα respectively (Fasching et al 2015). In humans, the OB-fold proteins Rmi1 and Rmi2 stabilise the topoisomerase in an open conformation that favours decatenation over relaxation of the DNA substrate (Bachrati & Hickson 2009, Bocquet et al 2014).

1.5.4 Regulation of Homologous Recombination

Tight regulation of HR is essential as uncontrolled recombination can lead to aberrant chromosomal rearrangements. Furthermore, recombination can be harmful in certain situations, for example stalled replication forks may be more safely restored using translesion DNA synthesis. In some circumstances the nucleoprotein intermediates generated by HR are toxic and can cause cell cycle arrest and cell death (Krejci et al 2012).

Negative regulators

Several prominent negative regulators of HR are either DNA helicases or DNA translocases that act to antagonise recombinases such as Rad51, these are known as anti-recombinases, some of which are discussed here.

In *Saccharomyces cerevisiae*, the 3' to 5' DNA helicase Srs2, which has homology to the *E. coli* helicase UvrD, is capable of dismantling Rad51 filaments by promoting the displacement of Rad51 by RPA (Veaute et al 2003). This mechanism prevents untimely or unwanted recombination. Srs2 generally functions as a antirecombinase but also plays a role in promoting SDSA. This mechanism is not well understood, it is suggested that Srs2 could remove Rad51 filaments from D-loops, prevent second end capture (by Rad52) or collaborate with nucleases to cleave DNA tails and other intermediate structures (Krejci et al 2012). Additionally, Rad51 and Rad55/57 have been shown to antagonise Srs2 activity (Liu et al 2011). No mammalian homologue of the anti-recombinase Srs2 has been identified, however, other helicases have been shown to have similar functions such RecQ5, BLM and FANCJ. All of

these helicases have shown to be able to disrupt Rad51-ssDNA filaments (Bugreev et al 2007, Hu et al 2007, Sommers et al 2009).

Sgs1 in yeast is a RecQ helicase, which is thought to directly dismantle pre-synaptic filaments, inhibit aberrant invasion events and resolve recombination intermediates (Mankouri et al 2002, Oh et al 2007). The human orthologues of Sgs1 include the cancer-associated helicases BLM, WRN and RTS, which have demonstrated similar activities. The BLM homologue Mus309 in *Drosophila melanogaster* is able to free invading ssDNA from D-loops and channel it towards the strand annealing step in SDSA (Adams et al 2003, Bugreev et al 2007).

The Hel308 homologue HELQ-1 from *Caenorhabditis elegans*, along with RFS-1, have also been shown to promote the disassembly of Rad51 filaments from strand invasion intermediates (Ward et al 2010). Hel308 helicase and its homologs will be discussed in further detail in Section 1.7.

Positive regulators

In mammals, the recombinase mediators BRCA2, Rad52, Rad54 and Swi5-Sfr2 promotes the assembly of Rad51 onto ssDNA and aids Rad51 in displacing RPA (Jensen et al 2010). The archaeal recombinase mediator RadB facilitates the loading of RadA onto ssDNA to promote strand invasion and D-loop formation.

The Shu complex found in *Schizosaccharomyces pombe* promotes Rad51 function during replication-associated repair but may also function to antagonise Srs2 (Krejci et al 2012, Mankouri et al 2007).

The DNA-dependent ATPases, Rad54 and Rdh54/Tid1 regulate Rad51 by stabilising pre-synaptic filaments, stimulating Rad51 mediated strand invasion and promoting D-loop and HJ migration (Mazin et al 2010, Sung et al 2003).

1.6 Replication Fork Restart

Faithful replication of the genome is crucial for accurate transmission of genetic information from generation to generation. DNA lesions caused by endogenous or exogenous agents, or nicks in the leading strand, can cause replication forks to stall or collapse. The progression of a replication fork can also be hindered if it collides with a DNA-bound protein, for example stalled RNA polymerase. Survival of the organism is now dependant on removal of the obstruction and re-establishment of the replication fork. Several pathways are available to restart a stalled replication fork, which include the recombination dependent mechanisms of break-induced repair (BIR) and Holliday junction-mediated restart. Mechanisms of repair by homologous recombination are conservative but are complex as they involve the removal and subsequent reloading of the replication machinery. Replication forks can also be repaired using fork remodelling and fork resetting mechanisms that are independent of homologous recombination, Figure 1.20.

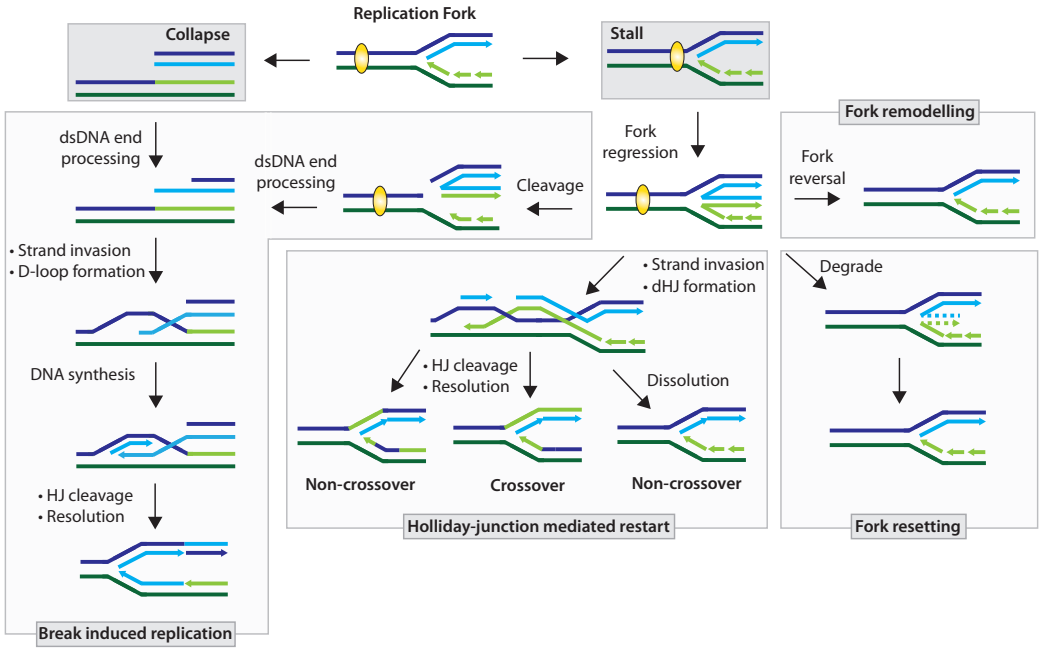


Figure 2.20: **Replication fork restart pathways.** Stalled replication forks can be restarted via several pathways such as the recombination-dependent Holliday junction-mediated restart and break induced replication (BIR) mechanisms. Stalled forks can be restarted by more direct recombination independent methods such as fork remodelling and fork resetting. Collapsed forks are repaired only via BIR.

1.6.1 Stalled fork processing

Stalled forks can be processed by several different pathways to reset the replication fork. Stalled forks can be repaired by either homologous recombination dependent or independent methods.

Fork regression

Regression of a fork into a Holliday junction is one method of stabilising stalled replication forks, Figure 1.20. The four-way Holliday junction like intermediate formed by fork regression can be reset into a replication fork via recombination dependent and recombination independent mechanisms.

In bacteria, the RecG helicase facilitates the regression/reversal of the replication fork to generate a Holliday junction intermediate termed a 'chicken foot'. Regression of the replication fork allows access for endonucleases such as RuvC to cleave the Holliday junction. RuvAB has also been shown to regress replication forks and to completely unwind the Holliday junction intermediate (Gupta et al 2014). The paired nascent leading and lagging strand substrates generated by the action of either RuvC or RuvAB are then used to reform the replication fork via the break-induced replication (BIR) pathway, which will be discussed in further detail below.

In eukaryotes the Bloom's syndrome and Werner's syndrome helicases BLM and WRN, respectively, as well as FancM are capable of stabilising stalled replication forks by facilitating fork regression into a Holliday junction (Gari et al 2008, Machwe et al 2006). BLM could also act to repress the formation of aberrant recombination intermediates at stalled forks (Ralf et al 2006). The ssDNA annealing helicase SMARCAL1 assists in stabilising stalled replication forks by re-annealing long stretches of ssDNA generated during the stalling process (Petermann & Helleday 2010). The Rad5 DNA helicase in *Saccharomyces cerevisiae* has shown fork regression activity by unwinding and annealing the nascent and parental strands of replication fork structures (Blastyak et al 2007).

The Hel308 homolog Hjm from *Sulfolobus tokodaii*, an archaeon belonging to the Crenarchaeota, displays fork regression activity *in vitro*. Furthermore, in *Pyrococcus furiosus* Hjm shows *in vitro* Holliday junction migration activity suggesting that Hjm may have a similar function to the bacterial RecG at targeting replication forks and forming Holliday junctions (Fujikane et al 2006, Li et al 2008). The branch migration of activity of Hjm is poor compared to its activity in the unwinding of the lagging strand in replication forks or D-loops. The archaeal FancM homologue Hef contains an N-terminal helicase, which has shown to be able to regress the branch point of a replication fork stalled on the leading strand (Komori et al 2004).

Fork remodelling

Following regression, the fork can be reversed by helicases to re-establish replication; this is a direct method of replication fork restart that is recombination-independent, Figure 1.20. In bacteria, the helicases RecG and RecQ are able to reverse replication fork regression *in vitro,* suggesting that they are capable of resetting replication forks (Manosas et al 2013, McGlynn & Lloyd 2000, McGlynn & Lloyd 2002a). In addition to generating regressed replication forks, the eukaryotic helicases BLM and WRN may also reverse the formation of Holliday junction-like structures via their HJ migration activities, thus promoting replication fork restart (Constantinou et al 2000, Karow et al 2000). FancM in eukaryotes and the archaeal FancM homologue Hef are capable of migrating Holliday junctions and may act in the reversal of regressed forks during fork resetting (Atkinson & McGlynn 2009, Lestini et al 2010, Whitby 2010).

Fork resetting

Fork resetting is a mechanism involving the degradation of the regressed DNA strands to restore the replication fork, Figure 1.20. In *E. coli*, nascent DNA strands at replication forks have seen to be partially degraded by the nuclease RecJ and this is aided by the helicase RecQ (Chow & Courcelle 2007). In humans, the Dna2 nuclease/helicase and the WRN helicase have been shown to interact at reversed replication forks to degrade dsDNA with a 5' to 3' polarity.

The human RecQ1 helicase inhibits Dna2 to prevent excessive nascent strand degradation and is has been suggested that Dna2 plays a partially redundant role with the exonuclease Exo1 in the repair of replication forks (Karanja et al 2012, Thangavel et al 2015).

Holliday junction-mediated resetting

Holliday junction like structures generated by fork regression can be reassembled into replication forks via two recombination dependent mechanisms. One mechanism is called Holliday junction-mediated resetting where the newly regressed DNA strand invades and recombines with the parental duplex DNA. Strand invasion is catalysed by recombinases and this process forms a double-Holliday junction. The replication fork is restored when the Holliday junction intermediates are resolved or dissolved as described in Section 1.4.3.5.

Break-induced replication

Break-induced replication (BIR) is a homologous recombination-dependant method of restarting stalled replication forks, Figure 1.20. DNA replication during BIR proceeds via a bubble-like replication fork that results in conservative inheritance of the new genetic material. The Pif1 helicase promotes DNA synthesis during BIR by stimulating DNA polymerase Polδ at D-loops (Saini et al 2013, Wilson et al 2013). After fork stalling, the replication fork or a regressed replication fork is cleaved via endonucleases to generate a one-ended DSB. In both instances, the DSB undergoes 5' end resection near the fork junction to generate 3' ssDNA end that is capable of invading the homologous template. Homologous recombination proteins process the double strand break and promote strand exchange. The enzymology of homologous recombination across all three domains of life is detailed in Section 1.4.3. The resolution of the Holliday junction intermediates by the endonuclease Mus81 restores the replication fork (Mayle et al 2015). Mutation rates as a result of BIR are around 1000 times higher than those seen during normal DNA replication (Sakofsky et al 2012).

In bacteria the cleavage of regressed replication forks is carried out by the RuvC complex following fork regression by RuvAB or RecG (Gupta et al 2014, Seigneur et al 1998). In eukaryotes the exonuclease/endonuclease protein EEPD1 along with the end resection nuclease Exo1 and BLM enhances 5' end resection at stalled replication forks (Wu et al 2015). 5' end resection prevents the competing NHEJ pathway from repairing the double strand break and therefore channels the repair of a replication fork down a homologous recombination-dependant pathway. This is advantageous as NHEJ during replication fork repair could lead to aberrant fusing of dsDNA ends that are generated as intermediates during replication fork repair. In archaea, one-sided DNA breaks are thought to be generated by the nuclease Hef and the Holliday junction resolvase Hjc during the restart of replication forks (Komori et al 2002, Komori et al 2004, Lestini et al 2010).

1.6.2 Replisome reassembly

In *E. coli*, origin-independent reassembly of the replisome is catalysed by the DNA structure-specific factors PriA and PriC. PriC facilitates the loading of the replicative helicase DnaB onto the lagging strand DNA of fork structures with large gaps on the leading strand; around 20 bp of ssDNA is required for efficient DnaB loading (Heller & Marians 2005a). If the initiation of the last Okazaki fragment on the lagging strand occurs close to the replication fork junction, the 3'-5' helicases Rep or PriA act to unwind the nascent lagging stand. This exposes enough single stranded DNA for PriC-mediated DnaB loading (Heller & Marians 2005b). If the blocked leading strand is located close to the fork junction, then replication can be initiated by a PriA-dependent mechanism (Gabbai & Marians 2010, Heller & Marians 2005a). PriA preferentially acts at fork structures with no gaps on the leading strand such as the junction of a D-loop; PriA, with PriB and DnaT, acts to load the replicative helicase DnaB to the displaced strand of the D-loop to reform the replisome. Following replisome reassembly by PriA, a leading strand priming event would allow replication to proceed downstream from the original blockage, similar to the PriC-dependent pathway. These methods of replication restart result in

ssDNA gaps on the leading strand of DNA, which can be repaired by homologous recombination.

Unlike bacteria, eukaryotes and archaea contain multiple origins of replication; therefore other replication forks on the chromosome can act as a back up to rescue terminally-stalled replication forks (Ge et al 2007, Newman et al 2013). During replicative stress in eukaryotes, excess amounts of the replicative helicase MCM have been seen to activate 'dormant' origins of replication to alleviate replication pressures such as stalled replication forks (Ibarra et al 2008).

1.7 Helicases

Helicases are ubiquitous enzymes that play a fundamental role in nearly all DNA and RNA processes including replication, recombination, DNA repair, transcription, translation and RNA splicing (Hall & Matson 1999, Matson et al 1994).

Helicases are molecular motors that catalyse the separation of two complementary strands of a nucleic acid duplex, a reaction dependent on energy derived from nucleoside 5' triphosphate (NTP) hydrolysis, Figure 1.21 (Hall & Matson 1999). Once loaded onto single stranded DNA or RNA, a helicase will translocate unidirectionally; based on the directionality of this translocation, helicases can be classified as 5'-3' or 3'-5' (Brosh 2013).

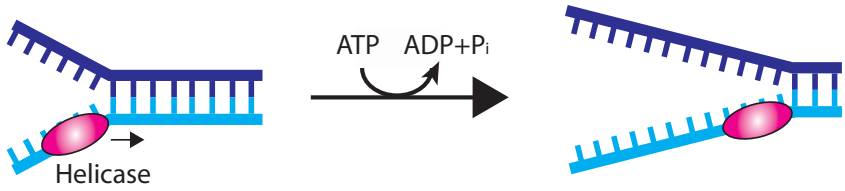


Figure 1.21: **Helicase DNA/RNA unwinding.** Strand separation of complementary DNA or RNA by a helicase (pink oval) is ATP dependant.

Due to their essential roles in nucleic acid metabolism, helicases have been implicated in a range of human genetic disorders. The helicases XPB and XPD

are part of the TFIIH transcription factor and play a role in nucleotide excision repair (NER). Mutations within these subunits are associated with the genetic disorders xeroderma pigmentosum (XP), which is typified by an extreme predisposition to skin cancer, in addition to Cockayne syndrome (CS) and trichothiodystrophy (TTD), which cause premature aging with profound neurological defects (Coin et al 1999, Fuss & Tainer 2011). BLM and WRN are 3' to 5' RecQ-family helicases that when mutated lead to the genetic disorders Bloom's syndrome (BS) and Werner's syndrome (WS), respectively (Mohaghegh et al 2001). BS is associated with phenotypes such as immunodeficiency, facial erythema, small body size and sub-fertility whereas in WS loss of skin elasticity, development of cataracts and loss of subcutaneous fat is prevalent. In both syndromes, individuals are susceptible to cancers. At the cellular level, a high degree of genetic instability is observed in BS and WS patients, for BS a high frequency of homologous recombination events occur and in WS illegitimate recombination and high frequency of chromosomal deletions are seen (German 1993, Shen & Loeb 2000). The FANCM protein contains an N-terminal helicase domain and a C-terminal nuclease domain and is orthologous to archaeal Hef (Meetei et al 2005). FANCM promotes branch migration of Holliday junctions and replication forks, and mutations in this helicase lead to Fanconi anemia (FA), which is characterized by cancer predisposition developmental defects, bone marrow failure and sensitivity to DNA crosslinking agents (Gari et al 2008).

1.7.1.1 Helicase structure

Helicases are ordered into six superfamilies (SF1-6), the classification of helicases into these superfamilies is based on sequence and organisation of nine signature motifs designated Q, I, Ia, Ib, II, III, IV, V and VI. Motifs I and II are more commonly known as the Walker A and Walker B ATPase motifs, respectively (Gorbalenya & Koonin 1993, Singleton et al 2007, Tuteja & Tuteja 2004). The majority of helicases belong to SF1 and SF2; these helicases are typically monomeric and mostly have a 3′ to 5′ directionality. In these superfamilies, the Walker A and B motifs are highly conserved and the other motifs are less so. Helicases from SF3-6 usually form hexameric rings. SF3

enzymes are commonly found in DNA or RNA viruses and contain only Walker A and B motifs and a SF3 specific motif III (Singleton & Wigley 2002). SF4 have a 5′ to 3′ polarity and act as replicative helicases, in bacteria SF4 helicases associate with primases (Singleton et al 2007). The Rho helicase is the SF5 helicase and is responsible for the termination of transcription in bacteria by binding to a specific sequence on the nascent RNA and then unwinding the DNA/RNA hybrid (Kaplan & O'Donnell 2003). Finally, the eukaryotic replicative helicase MCM complex and the bacterial Holliday junction migration helicase RuvB are examples of SF6 helicases (Enemark & Joshua-Tor 2008). SF6 along with SF3 helicases also belong to the AAA+ family of ATPases (Neuwald et al 1999). The AAA+ superfamily is a large and functionally diverse superfamily of NTPases that are characterized by a conserved nucleotide-binding and a catalytic module; this AAA+ module contains an αβα core domain where the Walker A and B motifs are found (Snider et al 2008).

The Q, I, Ia, Ib, II, III, IV, V and VI conserved motifs are usually clustered in a region of 200-700 amino acids termed the 'helicase core', which consists of two RecA-like folds (Tuteja & Tuteja 2004). The RecA folds contain the highly-conserved Walker A and Walker B boxes that are responsible for coupling the energy from ATP hydrolysis to conformational changes within the helicase that, in turn, drive DNA translocation and unwinding (Patel & Picha 2000).

It should be noted that the presence of these helicase motifs does not necessarily mean that the protein will be capable of unwinding DNA, the motifs are just a representative of NTP dependant translocases (Singleton & Wigley 2002). A summary of the helicase motifs can be seen in Table 1.3.

Table 1.3: **A summary of the nine helicase signature motifs; Q, I, Ia, Ib, II, III, IV, V and VI.**

Motif	Information	Consensus sequence	Reference
Q	Contains highly conserved glutamine. Regulates ATP binding and hydrolysis and affinity for RNA	GAxxPoxxG	(Cordin et al 2004)
I (Walker A)	Contains invariant lysine that contacts the β-phosphate of the ATP upon binding and hydrolysis. Couples hydrolysis to DNA unwinding	AxxGxGKT (SF1), GxxxxGKT/S (SF2)	(Hall & Matson 1999, Pause & Sonenberg 1992).
Ia	Involved in ssDNA binding	FTNKAA (SF1/SF2)	(Caruthers & McKay 2002, Marintcheva & Weller 2003)
Ib	Involved in RNA binding	TPGR	(Rocak & Linder 2004)
II (Walker B)	Contains glutamic acid which is important for ATP hydrolysis	DEAD (SF1/SF2) QxxR (RNA helicases)	(Caruthers & McKay 2002)
III	Coupling of ATP to unwinding by hydrogen bond and stacking interactions with the DNA bases.	GDADQSIYRWR (SF1/SF2)	(Caruthers & McKay 2002)
IV	Makes direct contacts with ADP in the enzyme-ADP binary complex	AVLYRTNAQSR (SF1/SF2)	(Caruthers & McKay 2002)
V	Contains invariant glycine involved in affinity for ssDNA	HAAKGLE (SF1/SF2)	(Caruthers & McKay 2002)
VI	Mediates conformational changes associated with nucleotide binding	VGITRAEE (SF1/SF2)	(Gorbalenya & Koonin 1993)

1.7.1.2 Mechanism of Helicase Unwinding

The mechanism by which nucleotide strand displacement occurs can be classified as either active or passive. This classification depends on whether the helicase participates directly in the nucleotide-unwinding event or if it only acts to stabilize the resulting ssDNA. In the passive mechanism, DNA unwinding is achieved by binding of the helicase to ssDNA that has formed through fraying at an ssDNA/dsDNA junction. At the junction, the helicase traps the ssDNA and unwinding occurs as the helicase translocates (Lohman & Bjornson 1996). Active mechanisms of strand separation include the 'inchworm' and 'active rolling' models.

Inchworm Model

Most superfamily 1 and 2 helicases unwind according to the 'inchworm' model (Mackintosh & Raney 2006). The inchworm model requires coordinated alternate binding of nucleic acid at two different sites within the helicase. One site in the helicase is needed to bind ssDNA and the other binds both ssDNA and dsDNA. Figure 1.22. In this model, translocation and unwinding is coupled to ATP binding and hydrolysis, respectively. ATP coupled conformational changes within the helicase destabilise and 'flip out' bases of the nucleic acid duplex in a manner that has been proposed to resemble a 'Mexican wave'.

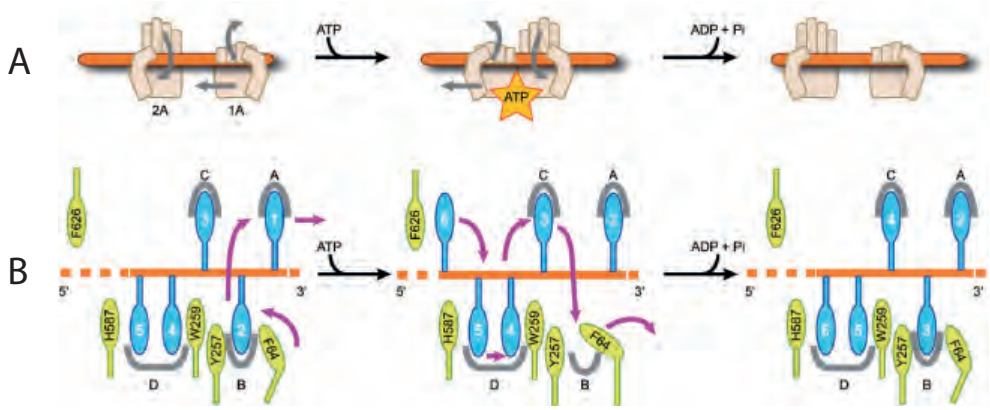


Figure 1.22 **Schematic of the inchworm model of helicase translocation and unwinding.** The helicase is bound to an ssDNA tail; upon binding of ATP there is a conformational change within the helicase, which leads to unwinding of DNA at an ss/dsDNA junction. Following the hydrolysis of ATP the helicase returns to its starting conformation. **(A)** Schematic showing the relative movement of two RecA-like domains illustrated by hands (1A and 2A). The open hand represents a loose grip on the DNA and a closed hand represents a tighter grip. **(B)** Illustration of the ssDNA-binding region within the helicase. During DNA unwinding and translocation the conformational changes in the helicase causes nucleotide bases to flip between binding pockets of the helicase. The bases are numbered arbitrarily in the 3′ to 5′ direction. Adapted from (Singleton et al 2007, Velankar et al 1999).

Co-operative Inchworm Model

Some monomeric helicases show enhanced unwinding when multiple monomers function cooperatively, this is known as the 'cooperative inchworm' model. This can occur when the monomers encounter a protein block or duplex DNA, the cooperation between monomers increases unwinding and therefore increases the chance that the obstacle will be overcome The superfamily 1 T4 phage helicase Dda is functional as a monomer but was seen to be more processive when acting co-operatively with other monomers to clear

streptavidin from biotin-labeled oligonucleotides (Mackintosh & Raney 2006), Figure 1.23.

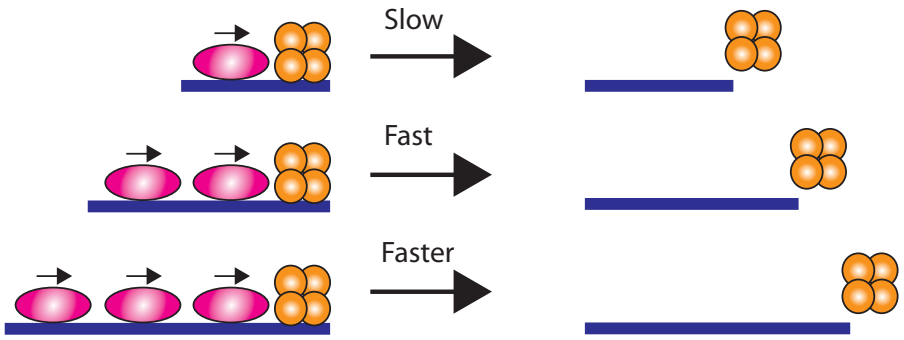


Figure 1.23: **Schematic of the co-operative inchworm model of helicase translocation and unwinding.** Multiple monomeric helicases assemble along the ssDNA function together to increase the rate of blockage clearing and therefore an increase in helicase unwinding and translocation. Helicase (pink oval), blockage (orange circles) Adapted from (Mackintosh & Raney 2006)

Active rolling Model

The active rolling model was proposed to explain the mechanism by which *E. coli* Rep helicase unwinds duplex DNA. This model requires an oligomeric helicase with each monomer having identical nucleic acid binding sites that can bind both ssDNA and dsDNA (Soultanas & Wigley 2001). In this model, at least one subunit is bound to ssDNA ahead of the ss/dsDNA junction at any given time and the second binding site is bound either to the same ssDNA region or to dsDNA ahead of the replication fork Figure 1.24. The *E.coli* the two Rep subunits alternate their affinity for DNA in an ATP dependent fashion so that DNA translocation occurs in a hand over hand mechanism and multiple base pairs are unwound by each Rep subunit in each binding event (Bjornson et al 1996).

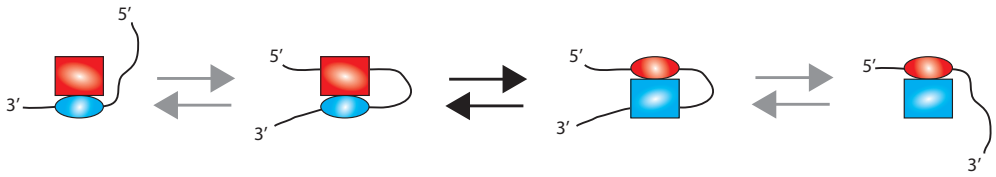


Figure 1.24 **Schematic of the active rolling mechanism of helicase translocation and unwinding.** The dimeric *E. coli* Rep helicase binds to ssDNA with the high affinity subunit (blue oval) and then the low affinity subunit (red rectangle) where one subunit is bound to ssDNA ahead of the ss/dsDNA junction. An ATP dependant isomerisation step can switch the affinities of the two Rep subunits. Translocation occurs by rolling of the Rep dimer on the DNA lattice. Adapted from (Bjornson et al 1996).

1.8 Hel308

Hel308 is a monomeric superfamily 2 helicase that is conserved across metazoans and archaea but is absent from bacteria and fungi (Richards et al 2008b, Woodman & Bolt 2009). Hel308 was first identified after the isolation of mutants in *Drosophila melanogaster* that are hypersensitive to DNA crosslinking agents (Boyd et al 1981), this led to the identification of the human Hel308 homologue. Confusingly, Hel308 is known by several different names depending on the organism of study. In higher eukaryotes, Hel308 is known as HELQ, in *C. elegans* it is known as HELQ-1 and in some archaea such as *Pyrococcus furiosus* and *Sulfolobus tokodaii* it is known as Hjm. Furthermore, in higher eukaryotes, the DNA polymerase Polθ also known as POLQ, since it contains a C-terminal DNA polymerase domain and an N-terminal helicase-like domain that shows similarity to Hel308.

1.8.1.1 Structure of Hel308

Crystal structures of three Hel308 homologues are available, one from the crenarchaeon *Sulfolobus solfataricus*, an apo and a DNA-Hel308 co-crystal from the euryarchaeon *Archaeoglobus fulgidus*, and the Hel308 homologue Hjm from the hyperthermophile *Pyrococcus furiosus* (Buttner et al 2007, Oyama et al 2009, Richards et al 2008b). All three crystal structures show structural conservation with each other even though their amino acid sequence is poorly conserved. *P. furiosus* Hjm shares only 30% and 37% amino acid identity with the *A. fulgidus* and *S. solfataricus* Hel308 helicases, respectively, but the overall protein folding is very similar (Oyama et al 2009). These crystal structures show that Hel308 consists of 5 structural domains that form a central pore lined with essential DNA binding residues. All five domains surround the DNA and drive a β-hairpin plough found in domain 2 through the duplex to separate the DNA strands, Figure 1.25.

Domains 1 and 2 contain the RecA folds that house the ATP-binding site motifs I, II, V and VI. Domain 3 contains a non-canonical winged-helix (WH) fold with four α-helices and two parallel β- strands. WH domains are common

nucleic acid binding domains and in Hel308 the WH in domain 3 is thought to be important in maintaining the structural integrity of the RecA folds found in domains 1 and 2 (Woodman & Bolt 2011); it is noteable that domain 3 is tightly packed against domain 1. Domain 4 contains a seven-helix bundle whose central helix is thought to act as a 'ratchet' to unwind DNA (Woodman et al 2007). Along with domains 1 and 3, domain 4 forms a ring around the 3′ tail of ssDNA. Domain 5 consists of a helix-loop-helix (HLH) structure that binds ssDNA as it extrudes from the central pore of the helicase. Domain 5 is involved in coupling ATP hydrolysis to helicase unwinding and is thought to act as an auto-inhibitory domain or 'molecular brake'(Richards et al 2008b, Woodman et al 2007). The domains of Hel308 will be discussed in further detail in Chapter 5: Genetic Characterisation of *hel038* Point Mutants.

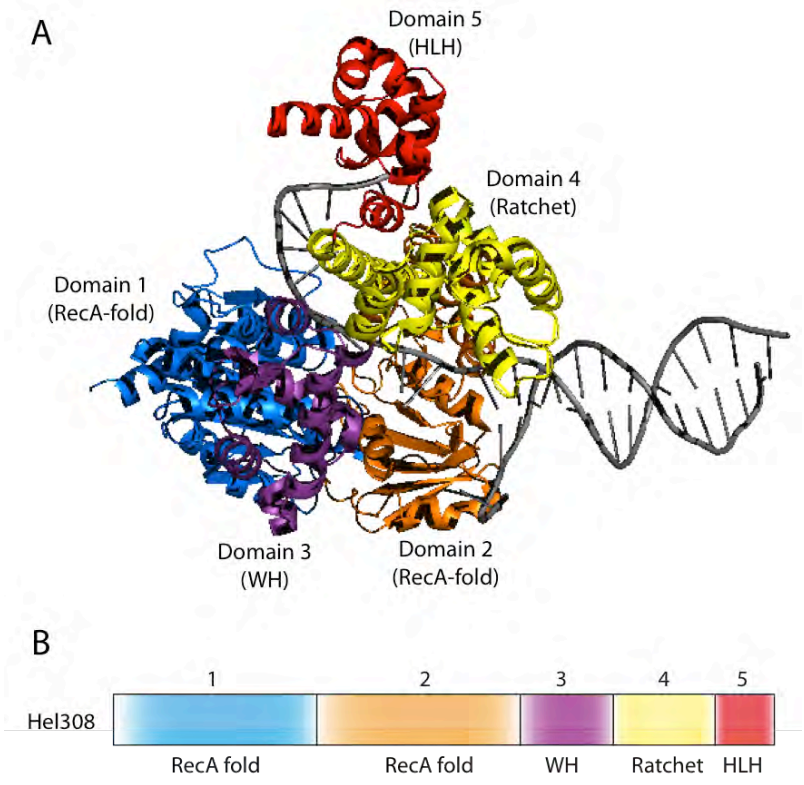


Figure 1.25: **Hel308-DNA co-crystal from *Archaeoglobus fulgidus*. (A)** Hel308 in complex with the 15 base pair DNA duplex and 10 base single stranded 3′ tail. Hel308 consists of 5 domains that form a central pore around ssDNA**. (B)** Linear schematic of the structural domains of Hel308. Structure taken from Protein Data Bank 2P6R (Buttner et al 2007)

1.8.1.2 Biochemical analysis of Hel308

Both human and archaeal Hel308 helicases have shown to translocate 3′ to 5′ *in vitro* and have ATPase activity that is stimulated by ssDNA but not dsDNA (Guy & Bolt 2005, Marini & Wood 2002). Hel308 has shown to be able to unwind and migrate Holliday junctions in *P. furiosus;* the Hel308 homologue is named Hjm for Holliday junction migration after this observation (Fujikane et al 2005). The authors suggested a role for Hel308 in the later stages of homologous recombination.

In *Methanothermobacter thermautotrophicus*, Hel308 has been shown to unwind replication fork structures with preference for displacing lagging strand structures (over leading strands) from a branch-point, however Hel308 has difficulty in unwinding parental duplex DNA (Guy & Bolt 2005). Hel308 targets DNA fork substrates most efficiently by engaging the fork branch point, which requires parental duplex, and either a strand nick or perhaps a ss–dsDNA junction. The Hel308 homologue Hjm from the hyperthermophilic archaeon *Sulfolobus tokodaii* also showed a preference for binding ss/dsDNA junctions and unwinding of nascent strands in replication fork structures. However, Hjm in *S. tokodaii* is able to unwind both leading and lagging strand structures with 5′ and 3′ overhangs, respectively; this is in contrast to Hel308 from *M. thermautotrophicus* which can only unwind 3′ overhangs (Fujii et al 2002, Li et al 2008). The difference between Hjm and Hel308 unwinding may be due to the evolutional divergence between Euryarchaeota and Crenarchaeota, and may indicate divergent cellular functions of Hjm and Hel308 helicases in these two major archaeal subdomains.

Hel308 is also able to dissociate the ‘invading’ strand in D-loop structures (Guy & Bolt 2005). A minimal length of between 7–15 nt of 3′ ssDNA tail is required for Hel308 to unwind a replication fork-like structure (Tafel et al 2011). The ability of Hel308 to unwind forks with a lagging strand suggests that Hel308 might act at damaged replication forks in which DNA replication on the leading strand template has stalled. It has also been proposed that Hjm from *S. tokodaii* can form “chicken foot” structures from the replication forks,

suggesting that Hjm/Hel308 homologues could be involved in remodeling of DNA replication forks after stalling events (Li et al 2008).

GFP tagged human Hel308 has been shown to localise to stalled replication forks after treatment with camptothecin (an agent that induces replication fork stalling and collapse). The Hel308 foci also co-localised with RPA, FANCD2 and Rad51 foci (Tafel et al 2011); FANCD2 is associated with the promotion of fork repair by homologous recombination and Rad51 is a recombinase (Taniguchi et al 2002). This indicates that Hel308 is recruited to sites of stalled replication forks following damage and is involved in the processing of stalled forks that require recombination-mediated processes for their restart.

1.8.1.3 Hel308 protein:protein interactions

In *M. thermautotrophicus*, Immobilised Hel308 was shown to interact *in vitro* with purified replication protein A (RPA). Mutational analysis revealed that it is the C-terminus of Hel308 that interacts with RPA (Woodman et al 2011). RPA was also was also seen to interact with mouse HELQ in immunoprecipitation assays (Adelman et al 2013). RPA is the eukaryotic and archaeal single stranded binding (SSB) protein that binds single stranded DNA to protect it from degradation and secondary structure formation. Additionally, RPA has been seen to directly stimulate the helicase activity of human Hel308 (Tafel et al 2011). It is proposed that in archaea, RPA could act as a platform to recruit Hel308 to stalled replication forks.

In *S. tokodaii*, gel filtration, affinity pulldown, and yeast two-hybrid analyses revealed that the Hel308 homologue Hjm physically interacts with Hjc *in vitro* (Li et al 2008). Hjc is a Holliday junction resolvase that binds specifically to Holliday junctions and cleaves two opposing strands symmetrically to generate two recombinant duplexes, and as been suggested to restart stalled replication forks (Lestini et al 2010). Hong and colleagues found that Hjm prevents the formation of Hjc-Holliday junction complexes, suggesting that Hjm may regulate the activity of the Hjc endonuclease (Hong et al 2012). This interaction with Hjc suggests a possible role for Hel308 in homologous

recombination, however neither the mechanism of this interaction or its function is known.

The human Hel308 homologue HELQ was shown to associate with the RAD51 paralogues RAD51B/C/D and XRCC2, which are collectively known as the BCDX2 complex, Figure 1.26. The BCDX2 complex is required for homologous recombination. HELQ was also shown to associate *in vivo* with the DNA damage-responsive kinase ATR, which is activated in response to persistent single-stranded DNA to cause a signaling pathway which can lead to the promotion of homologous recombination (Takata et al 2013). The BCDX2 complex and ATR were also seen to interact with mouse HELQ in reciprocal immunoprecipitation assays (Adelman et al 2013). These results suggest that HELQ may play a role in the restart of stalled replication forks, since the BCDX2 complex functions to promote replication-coupled homologous recombination. The Rad51 paralog *rfs-1* in *C. elegans* shows synthetic lethality when deleted in combination with *helq-1*, which may be caused by a mitotic defect (Taylor et al 2015). Observations of a *helq-1 rfs-1* double mutant strain suggest that this mitotic defect is due to a failure to dissasemble RAD51 from strand invasion intermediates. This was corroborated using purified HELQ-1 and RFS-1, which were found to independantly bind to and promote the disassembly of RAD51 from double stranded, but not single stranded DNA filaments *in vitro* (McClendon et al 2016, Ward et al 2010). This indicates that HELQ-1 and RFS-1 have overlapping but distinct roles in homologus recombination.

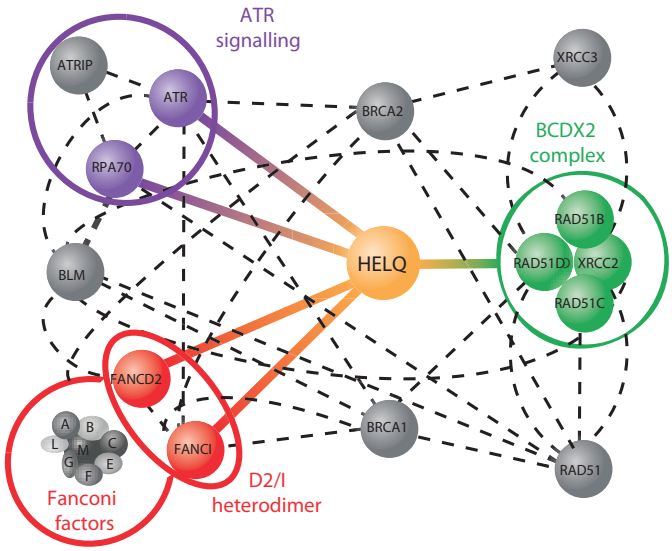


Figure 1.26: **HELQ interaction network.** Coloured lines depict interactions based on protein:protein pull downs and mass spectrometry analysis for the human Hel308 homologue HelQ. Dashed lines indicate reported interactions from BIOGRID, STRING and MINT databases Adapted from (Adelman et al 2013)

In vivo immunoprecipitation analyses in *P. furiosus* have found that the proliferating cell nuclear antigen (PCNA) co-precipitates with Hel308 homologue Hjm via an interaction with the C-terminal domain of Hjm. PCNA is the DNA sliding clamp that anchors DNA polymerase to the template DNA to prevent dissociation. Furthermore, PCNA was shown to stimulate the helicase activity of Hjm at fork structured DNA (Fujikane et al 2006).

1.8.1.4 Genetic analysis of Hel308

In almost all organisms that contain Hel308, deletions are viable meaning that Hel308 is not essential. Hel308 appears to be essential in *Sulfolobus islandicus* but not in other related *Sulfolobus* species (Hong et al 2012, Zhang et al 2013). In *H. volcanii hel308* deletions are slow growing and show sensitivity to DNA crosslinking agents such as MMC.

In *Drosophila melanogaster*, Hel308 deletion mutants are sensitive to cisplatin, a DNA crosslinking agent and potent inhibitor of DNA replication (Boyd et al 1990). Deletion of the Hel308 homologue HELQ in human cells also shows this same phenotype after treatment with the DNA crosslinking agent MMC

(Takata et al 2013). This suggests a specific function for Hel308 in the processing and repair of crosslinked DNA.

HELQ helicase deficient mice exhibit subfertility, germ cell attrition, DNA cross-linking sensitivity and a predisposition to tumours (Adelman et al 2013). These phenotypes are similar (albeit milder) to those observed in mouse models of Fanconi anaemia, therefore Adelman and collegues generated *Helq Fancd2* double mutant mice to ivestigate these smilarities (Parmar et al 2009). Double mutant cells exhibited a greater sensitivity than either single mutant to the replication blocking agents MMC and camptothecin. Furthermore, spontaneous and MMC-induced chromosomal aberrations were significantly increased over the *Helq* single mutant, suggesting that HELQ and FANCD2 act in parallel pathways of DNA interstrand cross link repair. This was corroborated by similar findings in *C. elegans*, FANCD2 RNAi studies in human cells, and with FANCC deleted mice models (Luebben et al 2013, Muzzini et al 2008, Takata et al 2013). Although HELQ was shown to act in a distinct pathway to the FA complex, HELQ was found to physically interact with the FANCD2–FANCI heterodimer by immunoprecipitation assays in mouse, Figure 1.26; this suggests that regulation of the two pathways may be modulated by a physical interaction (Adelman et al 2013).

Certain Hel308 single nucleotide polymorphisms (SNPs) have seen to be more prevelant in several types of head and neck cancers such as esophageal squamous cell carcinoma and gastric adenocarcinoma (Babron et al 2014, Li et al 2013, Liang et al 2012). These findings suggest a link between Hel308 and genome stability.

Summary

Various studies have shown Hel308 to interact with replication fork-like structures *in vitro*, as well as interacting with proteins commonly associated with replication fork maintenance such as RPA and PCNA. Additionaly, Hel308 has been seen to unwind D-loop structures and associate with proteins involved in the early stages of homologous recombination such as RAD51.

Hel308 was also shown to associate with Hjc a Holliday junction resolvase that acts in the latter stages of homologous recombination, and the Hel308 homologue Hjm has seen to migrate Holliday junctions, which again hints at a role in the latter stages of homologous recombination. Hel308 clearly is implicated in the restart of stalled replication forks and homologous recombination, but the exact function and mechanism of the role played by Hel308 is unknown.

1.8.2 Hel308b

In *H. volcanii* a second *hel308* gene is present, in addition to the canonical *hel308;* this second Hel308 helicase is termed Hel308b. Hel308b is 639 amino acids in length whereas the canonical Hel308 is 827 amino acids in length. The *hel308b* gene (HVO_0971) is found at bp 880821-882740 on the *H. volcanii* chromosome and is likely to have arisen from a gene duplication event. Similar hel308b genes have also been found in four other closely related haloarchaeal species. *hel308b* is a previously unstudied gene and will be described in further detail in Chapter 7: Phylogenetic and Genetic Characterisation of *hel308b*.

1.9 DNA Sequencing Using Nanopores

One aim of this study was to isolate novel DNA processing enzymes from halophilic archaeal viruses for the development of a nanopore based DNA sequencing technology, this work was in collaboration with Oxford Nanopore Technologies. The basic principles of DNA sequencing using nanopores and will be described here. An overview of halophilic archaeal viruses and the potential of halophilic viral proteins to be used in nanopore sequencing technologies will be described in detail in Chapter 8: *Novel Haloviral DNA Processing Enzymes For The Use In Nanopore DNA Sequencing Technologies,* Sections 8.1.1 to 8.1.3.

1.9.1 Basic Principles of Nanopore sequencing

When inserted into a lipid bilayer or a synthetic membrane, biological nanopores such as alpha hemolysin (α-HL) and *Mycobacterium smegmatis* porin A (MspA) can be utilised in the sequencing of ssDNA and ssRNA molecules (Braha et al 1997, Derrington et al 2010). If a potential is applied across the membrane, single molecules such as ssDNA and ssRNA are driven through the central aperture of the nanopore and cause perturbations of the current within the pore. Each base disrupts the ionic current in a characteristic way, therefore enabling the sequence of the ssDNA or ssRNA to be determined, Figure 1.27 (Deamer 2010). Several nanopore-based sequencing technologies are currently available or under development from Oxford Nanopore Technologies.

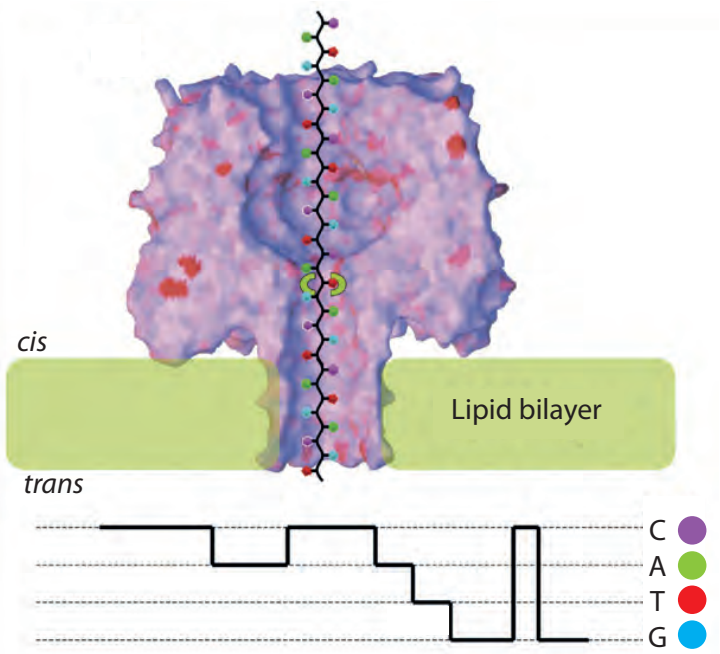


Figure 1.27: **DNA and RNA sequencing using a Nanopore.** Upon the application of a potential, DNA/RNA enters a membrane bound nanopore and disrupts the internal ionic current. Each DNA/RNA base perturbs the current in a characteristic by allowing for its sequence to be determined. Adapted from (Bayley 2006).

There are several advantages of nanopore-based sequencing platforms developed by Oxford Nanopore Technologies over conventional next generation sequencing methods. Firstly, sample preparation time is minimal compared to conventional sequencing methods, since the DNA does not

require shearing, cloning or enzymatic amplification prior to sequencing. Long read lengths of up to tens of kb can be achieved by nanopore sequencing compared to the 300-700 nt achieved by conventional methods (Bayley 2006, Shendure & Ji 2008). As well as providing greater sequence information, longer read lengths reduces the computational assembly time of sequence data. Finally, data can be acquired from nanopore sequencing in real time, allowing for faster analysis of results (Benner et al 2007).

1.9.2 Improvement of Nanopore Sequencing Technologies

The trans-membrane potential required to drive DNA or RNA through the central aperture of a pore is around 120 mV. However, at this voltage a single base takes around 2 μs to translocate the pore giving only around 2 pA of difference in ionic current between a purine and pyrimidine base. This small current does not provide a large enough signal-to-noise ratio to accurately discriminate between bases. One method to overcome this problem is to slow the translocation rate of DNA/RNA bases through the aperture of the pore from a microsecond timescale to a millisecond timescale. This would allow for better signal averaging and resolution of ionic current differences of a few picoamps (Deamer 2010).

Translocation of DNA/RNA through the pore may be slowed by the addition of a nucleic acid processing enzyme such as a helicase, polymerase or a nuclease to the *cis* face of the pore. The nucleic acid processing enzyme will ratchet the DNA/RNA through the central cavity of the pore in a controlled fashion, which will include pausing as the enzyme cycles. This regulated motion of the nucleic acid will allow for improved signal averaging of the ionic current within the pore, therefore giving a more reliable signal-to-noise ratio.

Many DNA processing enzymes have been trialled for their suitability in the nanopore sequencing apparatus. The addition of *E. coli* exonuclease I to a nanopore system was partially successful and the dwell time of translocating DNA within the pore was increased 10 fold. However, under the forces

produced by the applied voltage, the exonuclease was seen to dissociate from the pore within milliseconds after capture (Hornblower et al 2007)

The addition of the phi29 viral DNA polymerase to both a α-HL and an MspA nanopore has been shown to be successful in slowing the translocation of DNA through the central cavity of the pore, Figure 1.28 (Craig et al 2015). The phi29 DNA polymerase acted as a motor to pull ssDNA through the aperture of the pore, giving dwell times of bases within the pore of around 28 milliseconds and ionic current differences of up to 40 pA, thereby achieving single nucleotide resolution of signal (Maitra et al 2012, Manrao et al 2012).

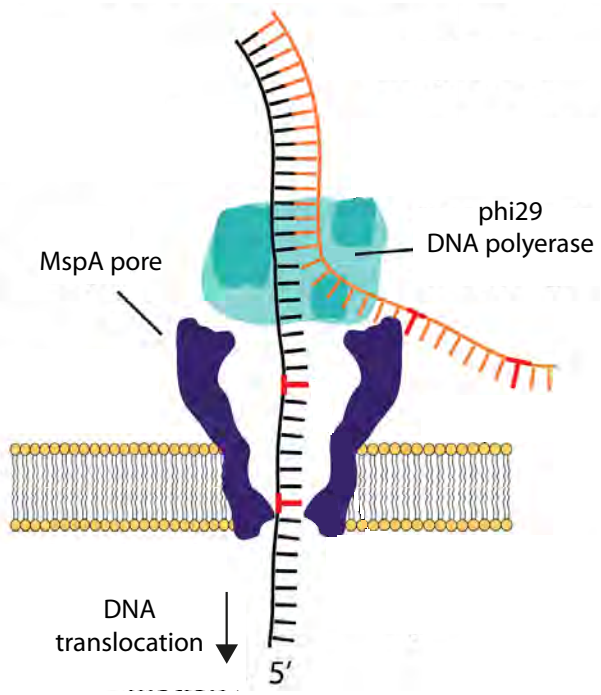


Figure 1.28: **phi29 DNA polymerase in nanopore DNA sequencing.** phi29 DNA polymerase slows the translocation of DNA through the MspA nanopore to enable better signal resolution and base calling. Adapted from (Craig et al 2015)

Developments in nanopore sequencing using DNA processing enzymes are promising, but there are still some challenges that require addressing. For example, the MinION nanopore sequencing technology developed by Oxford Nanopore currently has an error rate in base calling of around 12% to 40% (Goodwin et al 2015, Ip et al 2015, Laver et al 2015). Nanopores used in sequencing can accommodate up to 5 nucleotides at one time, meaning that slippages will lead to miscalling of bases. This is particular problem for

DNA/RNA tracts of the same base, which will cause a prolonged perturbation in current that cannot be easily resolved or differentiated from a stalled complex. Additionally, errors in the DNA processing enzyme could lead to backtracking of the DNA, or motor steps that are faster than can be resolved with existing technology (Maitra et al 2012).

The challenges faced during sequencing could be addressed by the discovery and development of DNA processing enzymes that can modulate the translocation of nucleotides through the nanopore in a tighter fashion, thus reducing errors in base calling. Furthermore, the development of RNA processing enzymes would be advantageous for the sequencing of RNA substrates for transcriptional studies.

Candidate nucleotide processing enzymes must be hardy enough to withstand the voltages applied across the nanopore, as well as the salinity of the bathing solution within the nanopore flow cell. Enzymes from halophilic organisms are adapted to high salt conditions and therefore could function in the buffer conditions found in a nanopore sequencing flow cell. Furthermore, viral proteins tend to be more robust and do not require numerous co-factors, compared to their cellular counterparts (Choi 2012, Kukkaro & Bamford 2009, Pauling 1982). Therefore, nucleotide processing enzymes from halophilic viruses could be key in the development of improved nanopore sequencing technologies in the future. The potential use of proteins from halophilic viruses will be discussed and investigated further in Chapter 8: *Novel Haloviral DNA Processing Enzymes For The Use In Nanopore DNA Sequencing Technologies,* Sections 8.1.1 to 8.1.3.

1.10 Aims

The aims of this study is to further understand the role of Hel308 in *H. volcanii,* this will be achieved by several approaches. The first is to delete *hel308* in combination with other DNA repair proteins, in particular ones involved in homologous recombination. The characterisation of these deletion mutants could give insight into the role of Hel308. Point mutations of Hel308 will also be studied which could reveal details about the mechanistic action of

Hel308. The role of Hel308 will also be studied by analysing its *in vivo* protein:protein interactions. This approach has the potential to reveal novel proteins that interact with Hel308 and could give insight into the pathways in which Hel308 acts. The previously unstudied Hel308b protein will also be phylogenetically and genetically characterised which could elucidate its role in *H. volcanii*. Lastly, in collaboration with Oxford Nanopore Technologies, this study also aims to isolate novel DNA processing enzymes from halophilic archaeal viruses for the development of a nanopore based DNA sequencing technology.

Chapter 2: Materials and Methods

2.1 Materials

2.1.1 Strains

Strain construction is detailed in Chapter 3.

Table 2.1: **Strains used in this study.**

Parent: H26 (*ΔpyrE2*)			
Strain	**Parent**	**Genotype**	**Notes**
H358	H26	*Δhef*	Constructed by ZD (Lestini et al 2010)
H1049	H26	*Δhjc*	Constructed by RL (Lestini et al 2010)
H1391	H26	*Δhel308*	Constructed by TA
H1392	H26	*Δhel308*	Constructed by TA
H1467	H1049	*Δhjc, Δhel308::trpA+*	Constructed by TA
H1468	H358	*Δhef, Δhel308::trpA+*	Constructed by TA
H1843	H26	*Δhel308b*	Chapter 3.2.2.2
H1844	H1391	*Δhel308, ΔradB*	Chapter 3.2.2.1
H1845	H26	*ΔradB*	Constructed by TA
H2076	H1392	*hel308-E422G*	Chapter 3.2.3.1
H2077	H1392	*hel308-D420A*	Chapter 3.2.3.1
H2078	H1392	*hel308-E330G*	Chapter 3.2.3.1
H2097	H1392	*hel308-H317G*	Chapter 3.2.3.1
H2396	H1392	*hel308-F316A*	Chapter 3.2.3.1
H2426	H1844	*Δhel308, ΔradB {hel308+ pyrE2+}*	Chapter 3.2.1
H2427	H1844	*Δhel308, ΔradB {pyrE2+}*	Chapter 3.2.1
H2488	H1391	*Δhel308, Δhel308b*	Chapter 3.2.2.2
H2572	H1391	*Δhel308 {hel308+ pyrE2+}*	Chapter 3.2.1
H2573	H1391	*Δhel308 {pyrE2+}*	Chapter 3.2.1

Parent: H53 (*ΔpyrE2, ΔtrpA*)			
Strain	**Parent**	**Genotype**	**Notes**
H1554	(H1393)	*hel308-K53R*	Constructed by TA
H1555	(H1393)	*hel308-D145N*	Constructed by TA
H1804	(H53)	*ΔoriC1, ΔoriC2, ΔoriC3, Δori-pHV4-2*	Constructed by KP
H1953	H1804	*ΔoriC1, ΔoriC2, ΔoriC3, Δori-pHV4-2, Δhel308::trpA+*	Chapter 3.2.2.1

Parent: H164 (*ΔpyrE2, bgaHa-Bb, leuB-Ag1, ΔtrpA*)			
Strain	**Parent**	**Genotype**	**Notes**
H2007	H164	*Δhel308b*	Chapter 3.2.2.2
H2117	H164	*Δhel308::trpA+*	Chapter 3.2.2.1
H2417	H2117	*ΔradB, Δhel308::trpA+*	Chapter 3.2.2.1
H2257	H2117	*hel308-E422G*	Chapter 3.2.3.1
H2259	H2117	*hel308-D420A*	Chapter 3.2.3.1
H2261	H2117	*hel308-H317G*	Chapter 3.2.3.1
H2263	H2117	*hel308-E330G*	Chapter 3.2.3.1
H2397	H2117	*hel308-F316A*	Chapter 3.2.3.1

H2398	H2117	*hel308-R743A*	Chapter 3.2.3.1
H2400	H2117	*hel308-D145N*	Chapter 3.2.3.1
H2641	H164	*ΔradB*	Chapter 3.2.2.1
H2643	H2007	*Δhel308b, Δhel308::trpA+*	Chapter 3.2.2.2

Parent: H2047 (*ΔpyrE2, Nph-pitA, Δmrr, cdc48d-ct, ΔtrpA)*			
Strain	**Parent**	**Genotype**	**Notes**
H2131	H2047	*Δhel308::trpA+*	Chapter 3.2.3.2
H2418	H2131	*his6 tag-hel308-strepII tag*	Chapter 3.2.3.1

Parent: H1424 (*ΔpyrE2, Nph-pitA, Δmrr, cdc48d-ct, ΔhdrB)*			
Strain	**Parent**	**Genotype**	**Notes**
H1737	H1424	*{p.tnaA:: his6 tag-hel308+strepIItag pyrE2+ hdrB+}*	Chapter 3.2.1
H1738	H1424	*{p.tnaA:: his6 tag-hel308b+strepIItag pyrE2+ hdrB+}*	Chapter 3.2.1
H1739	H1424	*{p.tnaA:: his6 tag-RadA+strepIItag pyrE2+ hdrB+}*	Chapter 3.2.1
H1740	H1424	*{p.tnaA::hel308+strepIItag pyrE2+ hdrB+}*	Chapter 3.2.1
H1741	H1424	*{p.tnaA::hel308b+strepIItag pyrE2+ hdrB+}*	Chapter 3.2.1
H1742	H1424	*{p.tnaA::RadA+strepIItag pyrE2+ hdrB+}*	Chapter 3.2.1
H1743	H1424	*{p.tnaA::strepII tag-hel308+ his6 tag pyrE2+ hdrB+}*	Chapter 3.2.1
H1744	H1424	*{p.tnaA::strepII tag-hel308b+ his6 tag pyrE2+ hdrB+}*	Chapter 3.2.1
H1745	H1424	*{p.tnaA::strepII tag-RadA+ his6 tag pyrE2+ hdrB+}*	Chapter 3.2.1
H1746	H1424	*{p.tnaA::strepII tag-hel308+ pyrE2+ hdrB+}*	Chapter 3.2.1
H1747	H1424	*{p.tnaA::strepII tag-hel308b+ pyrE2+ hdrB+}*	Chapter 3.2.1
H1748	H1424	*{p.tnaA::strepII tag-RadA+ pyrE2+ hdrB+}*	Chapter 3.2.1
H2167	H1424	*ΔtnaA::hdrB+*	Chapter 3.2.3.2
H2169	H1424	*ΔpilB3C3, ΔtnaA::hdrB+*	Chapter 3.2.3.2
H2682	H2169	*ΔpilB3C3, tnaAΔEco*NI	Chapter 3.2.3.2

TA = Thorsten Allers. SH = Sam Haldenby. ZD = Zhenhong Duan. RL= Roxane Lestini. KP= Katarzyna Ptasinska. Parental strains in brackets indicate intermediate strains were generated during construction, but these are not listed. [ ] indicates integrated plasmid, { } indicates episomal plasmid.

Table 2.2: ***E. coli* strains used in this study.**

Strain	Genotype	Notes
XL1-Blue MRF'	*endA1, gyrA96* (NalR), *lac* [F' proAB *lacIqZΔM15 Tn10* (TetR)], *Δ(mcrA)183, Δ(mcrCB-hsdSMR-mrr)173, recA1, relA1, supE44, thi-1*	Standard cloning strain for blue/white selection using pBluescript-based plasmids. Tetracycline resistant, Restriction endonuclease and recombination deficient, *dam+*. From Stratagene.
N2338 (GM121)	*F–, ara-14, dam-3, dcm-6, fhuA31, galK2, galT22, hsdR3, lacY1, leu-6, thi-1, thr-1, tsx-78*	*dam- dcm-* mutant for preparing DNA for *Haloferax volcanii* transformations. From (Allers et al 2004) M.G. Marinus via R.G. Lloyd.

2.1.2 Plasmids

Plasmid construction is detailed in Chapter 3.

Table 2.3: **Plasmids used in this study.**

Name	Use	Notes
pBluescript sII SK+	Standard *E. coli* vector with functional blue/white screening capability	From Stratagene
pTA83	*radA* deletion construct	Constructed by TA
pTA131	For making deletions in Δ*pyrE2* background	Constructed by TA (Allers et al 2004)
pTA163	Integrative vector containing *leuB-Ag1* and *pyrE2*. For use in recombination assays.	Constructed by GN
pTA236	*hjc* deletion construct	Constructed by GN
pTA312	*radB* deletion construct	Constructed by SH
pTA325	*radA::trpA+* deletion construct	Constructed by TA
pTA354	Episomal plasmid with *oripHV1/4*	Constructed by SH (Allers et al 2010)
pTA414	Episomal plasmid for *in trans* expression of *radA*	Constructed by TA
pTA415	*hel308* genomic clone	Constructed by ZD
pTA637	Episomal plasmid for *in trans* expression of *radA*	Constructed by TA
pTA1276	*hel308* deletion construct	Constructed by TA
pTA1277	*hel308::trpA+* deletion construct	Constructed by TA
pTA1316	*hel308* genomic clone	Constructed by TA
pTA1334	*hel308-K53R* gene replacement construct	Constructed by TA
pTA1335	*hel308-D154N* gene replacement construct	Constructed by TA
pTA1364	*hel308b* genomic clone	Chapter 3.1.1
pTA1368	*hel308b* deletion construct	Chapter 3.1.2.2
pTA1392	Vector for tryptophan inducible gene expression under the control of *p.tnaA* with an N-terminal *his$_6$*-tag and a C-terminal *StrepII* tag	Chapter 3.1.3.1
pTA1403	Vector for tryptophan inducible gene expression under the control of *p.tnaA* with an N-terminal *StrepII* tag and a C-terminal *his$_6$*-tag.	Chapter 3.1.3.1
pTA1419	For the expression of N-terminally *his$_6$*-tagged and a C-terminally *StrepII*-tagged Hel308 under control of *p.tnaA*	Chapter 3.1.3.1
pTA1422	For the expression of C-terminally *StrepII*-tagged Hel308 under control of *p.tnaA*	Chapter 3.1.3.1
pTA1425	For the expression of N-terminally *StrepII*-tagged and a C-terminally *his$_6$*-tagged Hel308 under control of *p.tnaA*	Chapter 3.1.3.1
pTA1428	For the expression of N-terminally *StrepII*-tagged Hel308 under control of *p.tnaA*	Chapter 3.1.3.1
pTA1503	Intermediate construct for creating pTA1508	Chapter 3.1.2.3
pTA1508	*tnaA* deletion construct	Chapter 3.1.2.3
pTA1545	*hel308-E422G* gene replacement construct	Chapter 3.1.2.4
pTA1546	*hel308-D420A* gene replacement construct	Chapter 3.1.2.4
pTA1548	*radB::trpA+* deletion construct	Constructed by TA
pTA1567	For the expression of N-terminally *his$_6$*-tagged and a C-terminally StrepII-tagged *radB* under control of *p.tnaA*	Chapter 3.1.3.1
pTA1575	*hel308-H317G* gene replacement construct	Chapter 3.1.2.4
pTA1576	*hel308-E330G* gene replacement construct	Chapter 3.1.2.4
pTA1615	*tnaA::hdrb+* deletion construct	Chapter 3.1.2.3
pTA1647	*hel308-F316A* gene replacement construct	Chapter 3.1.2.4
pTA1648	*hel308-R743A* gene replacement construct	Chapter 3.1.2.4
pTA1661	Intermediate construct for creating pTA1662	Chapter 3.1.3.2
pTA1662	N-terminally *his$_6$*-tagged and a C-terminally StrepII-	Chapter 3.1.3.2

	tagged *hel308* gene replacement construct	
pTA1669	Episomal plasmid for *in trans* expression of *hel308*	Chapter 3.1.3.2

TA = Thorsten Allers. GN= Greg Ngo. SH = Sam Haldenby. ZD = Zhenhong Duan.

2.1.3 Oligonucleotides

Table 2.4: **PCR Primers used in this study.**

Name	Sequence (5'-3')	Description	Used to construct
CgiBglR	CGAGTGGAGaTCTT GAGTACGGATTCGC	Amplification of *hel308*	pTA1648
dHel308b_D_BamHI_F	GAGTTCCGCGgGaTC CGGGCGC	Amplification of downstream flanking region of *hel308b* and introduction of an *Bam*HI site	pTA1368
dHel308b_D_XbaI_R	ACCCGCATCtAGAC GCCGTGG	Amplification of downstream flanking region of *hel308b* and introduction of an *Xba*I site	pTA1368
dHel308b_U_BamHI_R	CTCACATAAGgATCc CGCCGGC	Amplification of upstream flanking region of *hel308b* and introduction of an *Bam*HI site	pTA1368
dHel308b_U_KpnI_F	CTACGCCgGTAcCTC CTACGCC	Amplification of upstream flanking region of *hel308b* and introduction of an *Kpn*I site	pTA1368
Hel308_BamHI_F	GTCTGGaTccTTTCG AATGAGGCTCCTCG	Amplification of *hel308* and introduction of an *Bam*HI site	pTA1661
hel308-D420A-F	GTACATCTGGGCCG cCGCCGAGGACGTG C	Generation of a *hel308-D420A* mutation.	pTA1546
hel308-D420A-R	GCACGTCCTCGGCG gCGGCCCAGATGTA C	Generation of a *hel308-D420A* mutation.	pTA1546
hel308-E422G-F	CTGGGCCGACGCCG gcGACGTGCGGTCG A	Generation of a *hel308-E422G* mutation	pTA1545
hel308-E422G-R	TCGACCGCACGTCgc CGGCGTCGGCCCAG	Generation of a *hel308-E422G* mutation	pTA1545
Hel308b_d_BamHI_R	GCTTAAAACGGGAT cCCGAAGCG	Amplification of *hel308b* and introduction of an *Bam*HI site	pTA1429
hel308b_U_SphI_F	CCTCTGTGACgcATG CGCGTGCG	Amplification of *hel308b* and introduction of an *Sph*I site	pTA1420, pTA1426
Hel308bF	ACCAGTTCGGCTTT CGTGTCCG	Forward primer for *hel308b* probe	-
Hel308bF_NdeI	GTGACAtATGCGCG TGCGTGACCTCCCG C	Amplification of *hel308b* and introduction of an *Nde*I site	pTA1423
Hel308bR	GCTGTACATCGTCC CGCTTCGC	For *hel308b* probe and for RT-PCR to check *hel308b* transcript	-
Hel308bR_NheI	CCGATgctagcGGGCG ACAGCAGCGCCCG	Amplification of *hel308b* and introduction of an *Nhe*I site	pTA1420, pTA1423, pTA1426
Hel308bRTR	TCCGGGACGCGACG CTGGAGGG	For RT-PCR to check *hel308b* expression	
Hel308E330GF	CTCGTGGgcGACGC CTTCCGCGACAGAC	Generation of a *hel308-E330G* mutation	pTA1576
Hel308E330GR	AGGCGTCgcCCACG AGCGTTCGGTGTTC C	Generation of a *hel308-E330G* mutation	pTA1576

hel308EcoR	AGGTAGTCGAGCAC GCGGTCC	Amplification of internal region of *hel308*	-
Hel308F_NdeI	AGATcatATGCGAAC TGCGGACCTGACGG	Amplification of *hel308* and introduction of an *Nde*I site	pTA1422
Hel308F316AF	CGCGGCGgcCCACC ACGCGGGACTCGCC G	Generation of a *hel308-F316A* mutation	pTA1647
Hel308F316AR	CGTGGTGGgcCGCC GCGCCTTTGGCGAC C	Generation of a *hel308-F316A* mutation	pTA1647
Hel308FInt	AGCGCTGGGAGGA GTACGGC	Amplification of internal region of *hel308*	-
Hel308H317GF	GGCGTTCggCCACG CGGGACTCGCCGCG G	Generation of a *hel308-H317G* mutation	pTA1575
Hel308H317GR	CCGCGTGGccGAAC GCCGCGCCTTTGGC G	Generation of a *hel308-H317G* mutation	pTA1575
hel308Nde5R	CAGTTCGCATatgAT CTCCCTTGG	Amplification of upstream flanking region of *hel308* and introduction of an *Nde*I site	pTA1276, pTA1661
hel308NsiF	GAACGCTCGTGGAA GACGCC	Amplification of internal region of *hel308*	-
Hel308R_NheI	GCCgctagcTTCGAAA TCACCCAGACTGG	Amplification of *hel308,* removal of stop codon and introduction of an *Nhe*I site	pTA1419, pTA1422, pTA1425
Hel308R743AF	CCGAAAGgcCGCCC GCCGGCTGTTCGAG G	Generation of a *hel308-R743A* mutation	pTA1643
Hel308R743AR	GGCGGGCGgcCTTTC GGCCGACGCCGCGG	Generation of a *hel308-R743A* mutation	pTA1643
Hel308RTR	GTGTTGACGCCGGC GGCGAGCG	For RT-PCR to check *hel308* transcript	-
Hel308SphF	GGAGATGgcATGCG AACTGCGG	Amplification of *hel308* and introduction of an *Sph*I site	pTA1419, pTA1425, pTA1428, pTA1429
helQ(R)BamHI02	CGAGGAtCCTCATTC GAAATCACCCAGAC TGG	Amplification of *hel308* and introduction of an *Bam*HI site	pTA1428
RadA6HF	ACCTATTGCGCATA TGCACCACCACCAC CACCACATGGCAGA AGACGACCTCG	Amplification of *radA* incorporating a *his6* tag and introduction of an *Nde*I site	pTA1421
RadABamR	CCGACgGAtCcACGG CTTACTCGG	Amplification of *radA* and introduction of an *Bam*HI site.	pTA1430
RadANdeF	GAACGACTGcaTAT GGCAGAAGACG	Amplification of *radA* and introduction of an *Nde*I site	pTA1424
RadAR_NheI	CGGCgctagcCTCGGG CTTGAGACCGGCG	Amplification of *radA,* removal of stop codon and introduction of an *Nhe*I site	pTA1421, pTA1424, pTA1427
RadAStrepIIF	ACCTATTGCGCATA TGTGGTCGCACCCG CAGTTCGAGAAGAA CATGGCAGAAGAC GACCTCG	Amplification of *radA* incorporating a StrepII tag	pTA1427, pTA1430
radBBsF	CCTCCTGtCaTGACA GAGTCAGTCTCC	Amplification of *radB* and introduction of an *Bsp*HI site.	pTA1567
radBNheR	CGACCCCTgcTagcCA CGTCAGTCGCGGAG AGCCC	Amplification of *radB,* removal of stop codon and introduction of an *Nhe*I site	pTA1567
rpoARTF	CGGCGAGCACCTGA TTGAC	For RT-PCR to check *rpoA*	-

		transcript	
rpoARTR	ACGGACGAGGAAG CAGACG	For RT-PCR to check *rpoA* transcript	-
Ski2F	CCTCGCTCGTCTTC GTGAACTC	Amplification of *hel308*	-

Lower case letters indicate mismatches with the wild-type gene. Novel restriction sites are underlined.

Table 2.5: **Oligonucleotides for plasmid construction used in this study.**

Name	Sequence (5'-3')	Description	Used to construct
StrepII_N_F	TATGTGGTCGCACCC GCAGTTCGAGAAGA ACATGTGGCGCCCA GCTAGCCACCACCAC CACCACCACTGAGAT ATCG	Incorporating in-frame N-terminal StrepII tag and a C-terminal his_6 tag. With 5' *Nde*I and 3' *Eco*RI compatible ends. Introduction of *Pci*I/*Nsp*I, *Nar*I, *Nhe*I and *Eco*RV sites.	pTA1403
StrepII_N_R	AATTCGATATCTCAG TGGTGGTGGTGGTGG TGGCTAGCTGGGGCG CCACATGTTCTTCTC GAACTGCGGGTGCGA CCACA	Incorporating in-frame N-terminal StrepII tag and a C-terminal his_6 tag. With 5' *Nde*I and 3' *Eco*RI compatible ends. Introduction of *Pci*I/*Nsp*I, *Nar*I, *Nhe*I and *Eco*RV sites.	pTA1403
StrepIIF	CATGTGGCGCCCCAG CTAGCTGGTCGCACC CGCAGTTCGAGAAGT GAGATATCG	Incorporating in-frame StrepII tag. With 5' *Pci*I/*Nsp*I and 3' *Eco*RI compatible ends. Introduction of *NarI*, *Bst*XI, *Bmt*I, *Nhe*I and *Eco*RV sites.	pTA1392
StrepIIR	AATTCGATATCTCAC TTCTCGAACTGCGGG TGCGACCAGCTAGCT GGGGCGCCA	Incorporating in-frame StrepII tag. With 5' *Pci*I/*Nsp*I and 3' *Eco*RI compatible ends. Introduction of *NarI*, *Bst*XI, *Bmt*I, *Nhe*I and *Eco*RV sites.	pTA1392

Novel restriction sites are underlined.

2.1.4 Chemicals and Enzymes

All chemicals were purchased from Sigma and all enzymes from New England Biolabs (NEB) unless otherwise stated. Specific buffers and solutions are detailed with the appropriate method.

2.1.4.1 Media

Haloferax volcanii Media

All media is sterilised by autoclaving (1 minute 121ºC) and stored in the dark at room temperature (unless otherwise stated). Plates are stored in sealed bags

in the dark at 4ºC to prevent desiccation and dried for at least 30 minutes before use.

30% salt water (SW): 4 M NaCl, 148 mM $MgCl_2.6H_2O$, 122 mM $MgSO_4.7H_2O$, 94 mM KCl, 20 mM Tris.HCl pH 7.5.

18% salt water (SW): Made up with 30% SW, 3 mM $CaCl_2$. $CaCl_2$ added after autoclaving.

Trace elements: 1.82 mM $MnCl_2.4H_2O$, 1.53 mM $ZnSO_4.7H2O$, 8.3 mM $FeSO_4.7H_2O$, 200 µM $CuSO_4.5H_2O$. Filter sterilised and stored at 4ºC.

Hv-Min salts: 0.4 M NH_4Cl, 0.25 M $CaCl_2$, 8% v/v of trace element solution. Stored at 4ºC.

*Hv-Min carbon source***:** 10% DL-lactic acid Na_2 salt, 8% succinic acid Na_2 salt·$6H_2O$, 2% glycerol, pH to 7.0 with NaOH. Filter sterilised.

10 × YPC: 5% yeast extract (Difco), 1% peptone (Oxoid), 1% casamino acids, 17.6 mM KOH. Not autoclaved, used immediately.

10 × Ca: 5% Casamino acids, 17.6 mM KOH. Not autoclaved, used immediately.

Hv-Ca Salts: 362 mM $CaCl_2$, 8.3% v/v of trace elements, 615 µg/ml thiamine, 77 µg/ml biotin.

KPO_4 Buffer: 308 mM K_2HPO_4, 192 mM KH_2PO_4 pH 7.0

Hv-YPC agar: 1.6% Agar (Bacto), 18% SW, 1× YPC, 3 mM $CaCl_2$. Microwaved without 10× YPC to dissolve agar. 10× YPC added, then autoclaved. $CaCl_2$ added prior to pouring.

Hv-Ca agar: 1.6% Agar (Bacto), 18% SW, 1× Ca, 0.84% v/v of Hv-Ca Salts, 0.002% v/v of KPO_4 buffer (pH 7.0). Microwaved without 10× Ca to dissolve agar. 10× Ca added, then autoclaved. Hv-Ca Salts and KPO_4 Buffer added prior to pouring.

Hv-Min agar: 1.6% Agar (Bacto), 18% SW, 30 mM Tris·HCl pH 7.5, 2.5% Hv-Min carbon source, 1.2% Hv-Min Salts, 0.002% v/v of KPO_4 buffer (pH 7.0), 444 nM biotin, 2.5 μM thiamine. Microwaved to dissolve agar. Tris·HCl pH 7.5 added, then autoclaved. Hv-Min carbon source, Hv-Min Salts, KPO_4 buffer, biotin and thiamine added prior to pouring.

Hv-YPC broth: 18% SW, 1× YPC, 3 mM $CaCl_2$. $CaCl_2$ added after autoclaving, when cool.

Hv-Ca+ broth: 18% SW, 30 mM Tris.HCl pH 7.0, 1×Ca, 2.5% v/v of Hv-Min carbon source, 1.2% v/v of Hv-Min Salts, 0.002% v/v of KPO_4 buffer (pH 7.0), 444 nM biotin, 2.5 μM thiamine. 30% SW, dH_2O and Tris-HCl pH 7.0 autoclaved. All other components added when cool.

Haloferax volcanii media supplements

All solutions sterilised by filtration through a 0.2 μm filter.

Table 2.6: **Media supplements used with *H. volcanii***

Supplement	Abbreviation	Final Concentration
Leucine	Leu	50 μg/ml
Uracil	Ura	50 μg/ml
Thymidine	Thy	50 μg/ml (+50 μg/ml hypoxanthine in Hv-Ca and Hv-Min)
Tryptophan	Trp	50 μg/ml
5-Fluoroorotic acid	5-FOA	50 μg/ml (+ 10 μg/ml uracil)
Mevinolin	Mev	4 μg/ml

Growth of *H. volcanii* auxotrophic mutants on different media is shown below. Where a supplement is named, this supplement requires addition to the growth media.

Table 2.7: ***H. volcanii* mutant growth in different media.**

Genotype	Hv-YPC	Hv-Ca	Hv-Min
ΔpyrE2	+	Ura-	Ura-
ΔleuB	+	+	Leu-
ΔtrpA	+	Trp-	Trp-
ΔhdrB	Thy-	Thy-*	Thy-*

*In addition to thymidine, *ΔhdrB* strain cultures are supplemented with hypoxanthine in Hv-Ca and with hypoxanthine, methionine, glycine, and pantothenic acid in Hv-Min

Escherichia coli Media

Sterilised by autoclaving and stored at room temperature.

LB (Lysogeny Broth): 1% tryptone (Bacto), 0.5% yeast extract (Difco), 170 mM NaCl, 2 mM NaOH, pH 7.0.

LB agar: 300 ml of LB broth, plus 1.5% agar.

Escherichia coli media supplements

Table 2.8: **Media supplements used with *E. coli*.**

Supplement	Abbreviation	Final concentration
Ampicillin	Amp	50 μg/ml
Tetracycline	Tet	3.5 μg/ml
5-Bromo-4-chloro-3-indolyl β-D-galactopyranoside	X-Gal	66.67 μg/ml

2.1.4.2 Other Solutions

TE: 10 mM Tris.HCl pH 8.0, 1 mM EDTA.

2.2 Methods

2.2.1 General *Escherichia coli* Microbiology

Growth and Storage

Small liquid cultures (1-10 ml) were grown overnight in a static incubator (LEEC) at 37ºC with 8 rpm rotation. Larger cultures (300 ml) were grown overnight in an Innova 4330 floor-standing shaking incubator (New Brunswick Scientific) at 37ºC with 110 rpm shaking. Cultures on solid media were grown at 37ºC overnight in a static incubator. For short-term storage, solid and liquid cultures were stored at 4ºC. For long-term storage, glycerol was added to 20% (v/v) to cultures, snap frozen on dry ice and stored at -80ºC.

Preparation of Electrocompetent Cells

Two strains were used to prepare electrocompetent *E. coli* cells: XL-1 Blue (*dam*+, resistant to tetracycline) or N2338 (*dam*-).

A 5 ml overnight culture was grown at 37ºC with 8 rpm rotation with appropriate antibiotic selection. Cells were diluted 1/100 in LB broth supplemented with appropriate antibiotics and grown at 37ºC to A_{650} = 0.5-0.8. Cells were pelleted at 6000 × *g* for 12 minutes at 4ºC and the supernatant removed. The pellet was resuspended in an equal volume of ice-cold sterile 1 mM HEPES (pH 7.5). This process was repeated using 0.5 volumes 1 mM HEPES, 0.25 volumes 1 mM HEPES + 10% glycerol, 0.1 volumes 1mM HEPES + 10% glycerol, and 0.001 volumes 1mM HEPES + 10% glycerol. Cells were snap frozen on dry ice and stored in 100 µl aliquots at -80ºC.

Transformation of *Escherichia coli* by Electroporation

SOC broth: 2% tryptone (Bacto), 0.5% yeast extract (Difco), 10 mM NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, 10 mM $MgSO_4$, 20 mM glucose.

1 µg of DNA in 4 µl sterile dH_2O was added to 40 µl of electrocompetent cells, on ice and transferred to a pre-chilled electroporation cuvette (1 mm electrode gap, GENEFLOW). The cuvette was pulsed at 1.8 kV in an *E. coli* gene pulser (BioRad) and 1 ml of SOC was immediately added. After 1 hour incubation at 37ºC with 8 rpm rotation, cells were plated onto LB+2×Amp plates and incubated at 37ºC overnight.

2.2.2 General *Haloferax volcanii* Microbiology

Growth and Storage

Small liquid cultures (1-10 ml) were grown overnight in a static incubator (LEEC) at 45ºC with 8 rpm rotation. Larger cultures (50-333 ml) were grown overnight in an Innova 4330 floor-standing shaking incubator (New Brunswick Scientific) at 45ºC with 110 rpm shaking. Cultures on solid media were grown for 5-10 days at 45ºC in a static incubator (LEEC) in a plastic bag to prevent drying. For short-term storage, solid and liquid cultures were stored at room temperature. For long-term storage, glycerol (80% glycerol in 6% SW) was

added to 20% (v/v) to cultures, snap frozen on dry ice and stored at -80ºC.

Transformation of *Haloferax volcanii* Using PEG600

H. volcanii can be transformed by using PEG600 (Cline et al 1989). *H. volcanii* encodes a restriction endonuclease (Mrr) that targets methylated DNA; therefore, plasmid DNA requires passage through a *dam- E. coli* host prior to transformation (Holmes et al 1991). *Δmrr* strains of *H. volcanii* can be transformed directly with *dam+* plasmid DNA.

Buffers and solutions all sterilised by filtration through a 0.2 μM filter.

Buffered Spheroplasting Solution: 1 M NaCl, 27 mM KCl, 50 mM Tris.HCl pH 8.5, 15% sucrose.
Unbuffered Spheroplasting Solution: 1 M NaCl, 27 mM KCl, 15% sucrose, pH 7.5.
Transforming DNA: 5 μl 0.5 M EDTA, pH 8.0, 15 μl unbuffered spheroplasting solution, 10 μl DNA (~1-2 μg).
60% Polyethylene Glycol 600 (PEG 600): 150 μl PEG 600 and 100 μl unbuffered spheroplasting solution.
Spheroplast Dilution Solution: 23% SW, 15% sucrose, 37.5 mM $CaCl_2$.
Regeneration Solution: 18% SW, 1×YPC, 15% sucrose, 30 mM $CaCl_2$.
Transformation Dilution Solution: 18% SW, 15% sucrose, 30 mM $CaCl_2$.

A 10 ml culture of *H. volcanii* was grown in Hv-YPC (+ additives) overnight at 45ºC until an A_{650} of 0.6-0.8. Cells were centrifuged at 3300 × *g* for 8 minutes in a 10ml round-bottomed tube, the supernatant removed and cells gently resuspended in 2 ml of buffered spheroplasting solution. Cells were transferred to a fresh 2 ml round-bottom tube, pelleted again, supernatant removed then resuspended gently in 600 μl buffered spheroplasting solution. A 200 μl aliquot per transformation was transferred to a fresh 2 ml round-bottom tube and a 20 μl drop of 0.5 M EDTA pH 8.0 was pipetted onto the side of the tube, gently inverted and incubated for 10 minutes at room temperature. Transforming DNA was added in the same manner as EDTA and incubated at room temperature for 5 minutes. 250 μl of 60% PEG 600 was added and mixed by gentle rocking followed by incubation at room temperature for 30 minutes. 1.5 ml of spheroplast dilution solution was then added and mixed by inversion and after 2 minutes incubation at room temperature was centrifuged at 3300 × *g* for 8 minutes at 25ºC. The cell pellet was transferred whole to a sterile 4 ml tube containing 1 ml regeneration solution (+ 40 μg/ml thymidine for *ΔhdrB* strains). After an undisturbed recovery at 45ºC for 90 minutes, the pellet was

resuspended by tapping the tube and incubated for a further 3 hours at 45ºC rotating at 8 rpm. Cells were transferred to a fresh 2 ml round-bottom tube and centrifuged at 3300 × *g* for 8 minutes. The cell pellet was resuspended gently in 1 ml of transformation dilution solution. Appropriate dilutions were made and 100 µl of each dilution plated on appropriate media. Plates were incubated for at least 5 days at 45ºC.

2.2.3 DNA Extraction From Cells

Plasmid Extraction from *Escherichia coli*

Plasmid DNA extraction from *E. coli* was performed using Macherey-Nagel Nucleospin (Mini) and Nucleobond AX (Midi) kits. Protocol was as described in the manufacturer's guidelines. 1-2 ml and 300ml of *E. coli* cell culture (LB broth+Amp) was used for minipreps and midipreps respectively. Plasmid DNA from minipreps was eluted in 30 µl elution buffer, and for midipreps the plasmid DNA was isopropanol and ethanol precipitated then resuspended in 200 µl of TE. Plasmid DNA samples were stored at -20ºC.

Plasmid Extraction from *Haloferax volcanii*

ST buffer: 1 M NaCl, 20 mM Tris.HCl pH 7.5.

Macherey-Nagel Nucleospin (Mini)/Nucleobond AX (Midi) kits were used to obtain circular plasmid DNA from *H. volcanii*. Due to *H. volcanii* having a low plasmid copy number and a large amount of cellular debris, the following amendments to the manufacturers guidelines are required:

Starter culture: 10 ml of cell culture was used for minipreps, and 300 ml for midipreps.

Resuspension step: Cell pellets were resuspended in the same total volume as *E. coli* minipreps or maxipreps. However, the pellet was initially resuspended in 1⁄2 volume of ST buffer supplemented with 50 mM of EDTA, then made up to total volume with standard resuspension buffer (Macherey-Nagel).

Chloroform extraction: To avoid column blockage, a chloroform extraction step was required to separate DNA from cellular debris. Sample was mixed

with an equal volume of chloroform, vortexed to mix and centrifuged for 5 minutes at 3300 × *g*. The aqueous top layer was then loaded onto the column.

Genomic DNA Extraction from *Haloferax volcanii*

ST buffer: 1 M NaCl, 20 mM Tris.HCl pH 7.5.
Lysis Solution: 100 mM EDTA pH 8.0, 0.2% SDS.

A 10 ml culture of *H. volcanii* was grown overnight at 45ºC until at an O.D. 650 = 0.6-0.8, then centrifuged at 3300 x *g* at 25ºC for 5 minutes. The supernatant was removed, the cells resuspended in 200 µl of ST buffer followed by addition of 200 µl of lysis solution. The tube was mixed by inversion and the cell lysate overlaid with 1 ml of 100% EtOH. DNA was spooled at the interface onto a capillary tip until the liquid was homogenous and clear. The spool of DNA was washed twice in 1 ml of 100% EtOH, and excess EtOH allowed to drain from the DNA. The DNA was air-dried, resuspended in 450 µl of TE and precipitated with ethanol. The pellet was thoroughly resuspended 100 µl of TE and stored at 4ºC.

Crude Genomic DNA Extraction from *Haloferax volcanii* for PCR

When high quality genomic DNA from *H. volcanii* was not required, a crude genomic DNA extraction was carried out. A liquid culture was grown at 45ºC to an A_{650} = 0.6-0.8. Then 1 ml was pelleted at 10,000 × *g* for 1 minute. The supernatant removed and the cell pellet resuspended in 400 µl of dH_2O. This was incubated at 94ºC to lyse the cells, followed by quenching on ice for 10 minutes. DNA was stored at 4ºC.

2.2.4 Nucleic Acid Manipulation

PCR Amplification

Amplification of DNA was carried out using Phusion, DyNAzyme EXT (Finnzymes) or Q5 Hotstart. These enzymes are suitable for amplifying templates with a high GC content. Phusion was used for amplifications that required high fidelity, DyNAzyme EXT was used for diagnostic amplifications and Q5 Hotstart was used for long or difficult amplicons. Reaction conditions are shown below. All reactions were carried out using a Techne TC-512

thermocycler.

Table 2.9: **PCR reaction components**

DyNAzyme EXT	Phusion	Q5 Hotstart
200 µM of each dNTP	200 µM of each dNTP	200 µM of each dNTP
0.5 µM of each primer	0.5 µM of each primer	0.5 µM of each primer
10-50 ng of template DNA	10 ng of template DNA	1 ng - 1 µg genomic DNA template or 1 pg – 1 ng plasmid DNA template
1× Optimised DyNAzyme Buffer	1× Phusion GC Buffer	1× Q5 Reaction Buffer
5% DMSO	3% DMSO	1x Q5 High GC Enhancer
1 U of Dynazyme EXT	1 U of Phusion	0.02 U/µl Q5 Hotstart

Table 2.10: **PCR reaction conditions**

Step	DyNAzyme EXT	Phusion	Q5 Hotstart	
Initial Denaturation	94ºC, 120 seconds	98ºC, 30 seconds	98ºC, 30 seconds	
Denaturation	94ºC, 15 seconds	98ºC, 10 seconds	98ºC, 5-10 seconds	30 cycles
Annealing	TmºC, 20 seconds	TmºC, 15 seconds	TmºC, 10-30 seconds	
Extension	72ºC, 40 seconds/kb	72ºC, 20 seconds/kb	72ºC, 20-30 seconds/kb	
Final Extension	72ºC, 10 seconds	72ºC, 10 seconds	72ºC, 120 seconds	

Annealing temperatures for primers (TmºC) were calculated using the following equation (Howley et al 1979) (Equation 2.1).

$$81.5 + (16.6 \times \log_{10}[Na^+]) + (0.41 \times \%GC) - (100 - \%homology) - \left(\frac{600}{length}\right)$$

Equation 2.1: **Calculating annealing temperature of primers.** %GC: percentage guanine and cytosine in the primer. Homology: percentage homology shared between primer and template. Length: length of primer in bases.

Touchdown PCR

When primers were not 100% homologous to the template DNA (e.g. when introducing restriction sites or mutations), two annealing temperatures were calculated. The first was based on the original percentage of homology (Tm_S),

and the second on 100% homology (Tm_e). The reaction annealing temperature started at Tm_s and increased to Tm_e across 10 cycles. The remaining 20 cycles used Tm_e as the annealing temperature.

Colony PCR

In order to screen large numbers of colonies for a desired plasmid or chromosomal gene, colony PCRs were used. *H. volcanii* or *E. coli* colonies growing on solid media were touched gently with a sterile yellow tip, ensuring only a small number of cells were picked up and the colony was not disturbed. The yellow tip was used to pipette up and down in 100 µl of dH_2O. This was boiled at 100ºC to lyse cells, then cooled on ice. 1 µl of this was then used in a PCR reaction with DyNAzyme EXT polymerase.

Overlap extension PCR

In order to introduce point mutations within a gene, complementary forward and reverse primers were designed containing the desired mutation. Two PCRs were carried out, one using each of the primers and an appropriate external primer to amplify the region either side of the desired mutation. A third PCR was then carried out using the external primers only, which amplified the mutated region of interest.

Restriction Digests

Digests were carried out as by the manufacturers instructions (NEB). All digests were supplemented with 200 ng/µl BSA (NEB). For double digests, NEB buffers were selected such that each enzyme had at least 75% activity. Plasmid DNA was digested for at least one hour, and genomic DNA for 16 hours.

Blunt-end Filling with Klenow enzyme

Overhangs generated by restriction digests were filled-in to produce a blunt ends using Klenow (NEB). Samples were incubated with 1 units of Klenow per µg of DNA, 1 x NEB Buffer 4 and 1 mM dNTPs for 30 minutes at 25ºC. The reaction was stopped by heat inactivation at 75ºC for 20 minutes.

Dephosphorylation of Vector DNA

To prevent self-ligation of vector DNA, Shrimp alkaline phosphatase was used to remove 5′ phosphate groups. Samples were incubated with 5 units of Shrimp alkaline phosphatase/μg of DNA and 1× Antarctic phosphatase buffer for 30 minutes at 37ºC. Phosphatase was heat inactivated at 65ºC for 10 minutes.

Ligation of DNA

Ligations were performed using T4 DNA ligase. For each μg of DNA, 5 units of ligase were used in a reaction with 1× T4 ligase buffer. For vector:insert ligations, reactions contained a molar ratio of ~3:1 insert to vector DNA. Ligations were carried out at 15ºC overnight or 4ºC for 36 hours, ethanol precipitated, resuspended in 4 μl dH_2O and transformed into *E. coli* or *H. volcanii* (*Δmrr* strains only).

Ethanol Precipitation of DNA

1/10 volume of 3 M sodium acetate (pH 5.3) and 2 volumes of 100% ethanol were added to DNA samples and incubated at -20ºC for 1 hour. Samples were centrifuged at 4ºC, 20,000 × *g* for 30 minutes and the supernatant removed. Pellets were washed in 400 μl of 70% ethanol followed by centrifugation at 4ºC, 20,000 × *g* for 10 minutes. The supernatant was removed and pellets air-dried before resuspension in sterile dH_2O.

Nucleic Acid Purification

PCR products, ligations, restriction digests and dephosphorylated DNA products were purified using Macherey-Nagel DNA purification kits. Protocol was as described in the manufacturer's guidelines. In these kits, DNA is bound pH dependently to a silica membrane and is separated from contaminants by ethanolic washing. DNA was eluted in 30 μl of the provided elution buffer.

DNA Sequencing

All DNA sequencing reactions and analysis were performed by the Biopolymer Synthesis and Analysis Unit, University of Nottingham. Sequencing was carried out using the dideoxy chain termination method (Sanger et al 1977).

Oligonucleotide Synthesis

Oligonucleotides were synthesised by Eurofins MWG, Germany.

Nucleic Acid Quantification

To determine the concentration and purity of plasmid preparations the absorbance at 260 nm and the 260:280 nm absorbance ratio, respectively, were measured by spectrometer (Beckman Coulter DU 530).

Agarose Gel Electrophoresis

TBE (Tris/Borate/EDTA): 89 mM Tris.HCl, 89 mM boric acid, 2 mM EDTA.
TAE (Tris/Acetic acid/EDTA): 40 mM Tris.HCl, 20mM acetic acid, 1 mM EDTA.
Gel Loading Dye (5×): 50 mM Tris·HCl, 100 mM EDTA, 15% Ficoll (w/v), 0.25% Bromophenol Blue (w/v), 0.25% Xylene Cyanol FF (w/v).

As standard practice TBE buffer was used to cast and run agarose gels. When high quality resolution and/or Southern blotting was required, TAE buffer was used.

Agarose gels were cast using agarose powder (Sigma) and TBE or TAE buffer. 1/5 by final volume of 5 x gel loading dye was added to the DNA samples and loaded alongside molecular markers, either a 1 kb ladder (NEB) or a 100 bp ladder (NEB). TBE gels (10 cm) were run at 100 V for 1 hour. TAE gels (25 cm) were run overnight (16 hours) with buffer circulation at 50 V. For visualization of bands, gels were stained with ethidium bromide at a final concentration of 0.5 μg/ml for 30 minutes. Gels used for DNA extraction were stained with ethidium bromide after removal of sample lanes.

Agarose Gel Extraction and Purification of DNA

To purify DNA from agarose gels without UV exposure, sample lanes were protected using foil while the appropriate band was excised. DNA was visualised by UV light using a UV transilluminator (UVP inc.). DNA was purified using the Macherey-Nagel DNA purification kit as described in the manufacturer's guidelines.

2.2.5 Genetic Manipulation of *Haloferax volcanii*

Generating a Genomic Clone

A genomic clone of a gene was obtained by restriction digest of high quality genomic DNA, restriction sites chosen were at least 500 bp upstream and downstream of the gene of interest. The digested genomic DNA was run on 0.75% TAE agarose gel for 16 hours at 50 V and the band of expected size (+/- 0.5 Kb) was extracted. DNA was purified and ligated to compatible sites in pBluescript II SK+. The resulting plasmid library was transformed into XL1-Blue *E. coli*. Transformants were selected on LB +Amp +X-gal and white colonies patched out and screened by colony hybridization using a gene specific probe.

Generating a Deletion Construct

A deletion construct was generated from a genomic clone by inserting the upstream and downstream flanking regions of the gene of interest into pTA131; this can be achieved by either restriction digest or PCR.

Generating a deletion construct by digest

A deletion construct could be made by digest if the gene of interest has restriction sites available within the coding sequence at both the 5' and 3' ends of the gene. Schematic shown in Figure 2.1 Blunt-ending with Klenow was required if these enzymes do not produce compatible ends. The genomic clone containing the gene of interest was digested with the relevant enzymes to remove the coding sequence, and the plasmid re-ligated. The upstream and downstream regions are then inserted into pTA131 by digest if appropriate sites are available ~800 bp upstream and downstream from the gene. Plasmids were transformed into XL-1 blue and plated on LB +Amp +X-gal, White colonies were selected and the plasmid confirmed by restriction digest and sequencing if necessary.

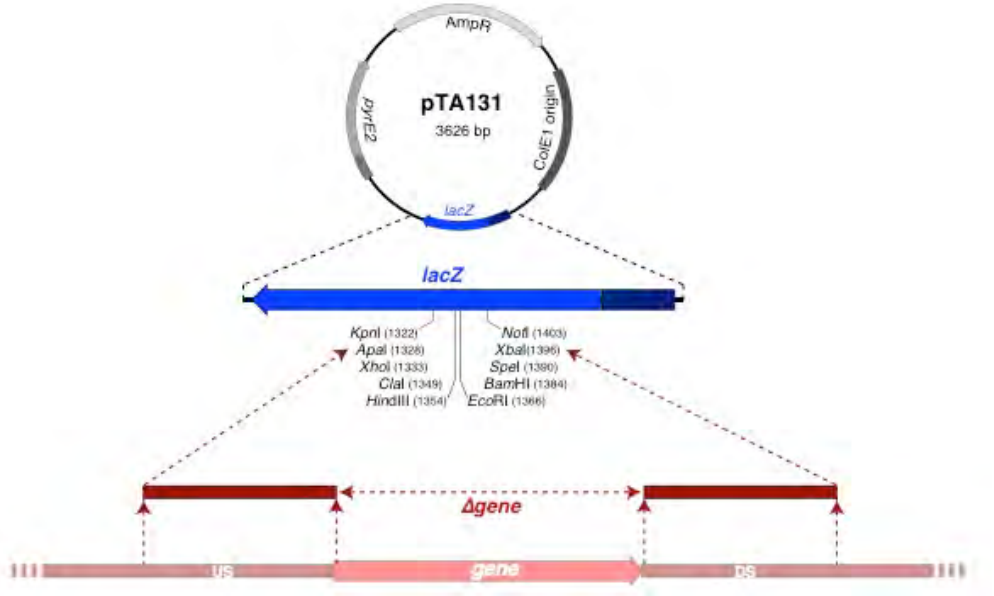


Figure 2.1: **Schematic for generating deletion constructs by digest.** Δ*gene* construct is generated by digesting appropriate restriction sites to remove the coding sequence and inserting the upstream (US) and downstream (DS) flanking regions into pTA131. *AmpR* (ampicillin resistance gene *E. coli*), *colE1 origin* (*E. coli* origin of replication), *lacZ* (*β-galactosidase* gene used for blue/white selection in *E. coli*), *pyrE2* (uracil biosynthesis in *H. volcanii*)

Generating a deletion construct by PCR

If no appropriate sites for restriction digest were present, then a deletion construct was obtained by PCR. Schematic is shown in Figure 2.2. A PCR was carried out to amplify the upstream region (US) of the gene of interest and another PCR to amplify the downstream (DS). External primers were designed to contain different restriction sites compatible with the pTA131 multiple cloning site. Internal primers were designed to make a 'clean' deletion of the gene of interest and introduce a *Bam*HI site at the site of deletion. The PCR products are both digested with *Bam*HI and ligated. This product was then digested with the external enzymes and inserted into pTA131. Plasmids were transformed into XL-1 blue and plated into LB +Amp +X-gal, White colonies were selected and the plasmid confirmed by restriction digest, and sequencing if necessary The presence of a *Bam*HI site marking the deletion allows for insertion of markers (e.g. *trpA*) into this region of the deletion construct.

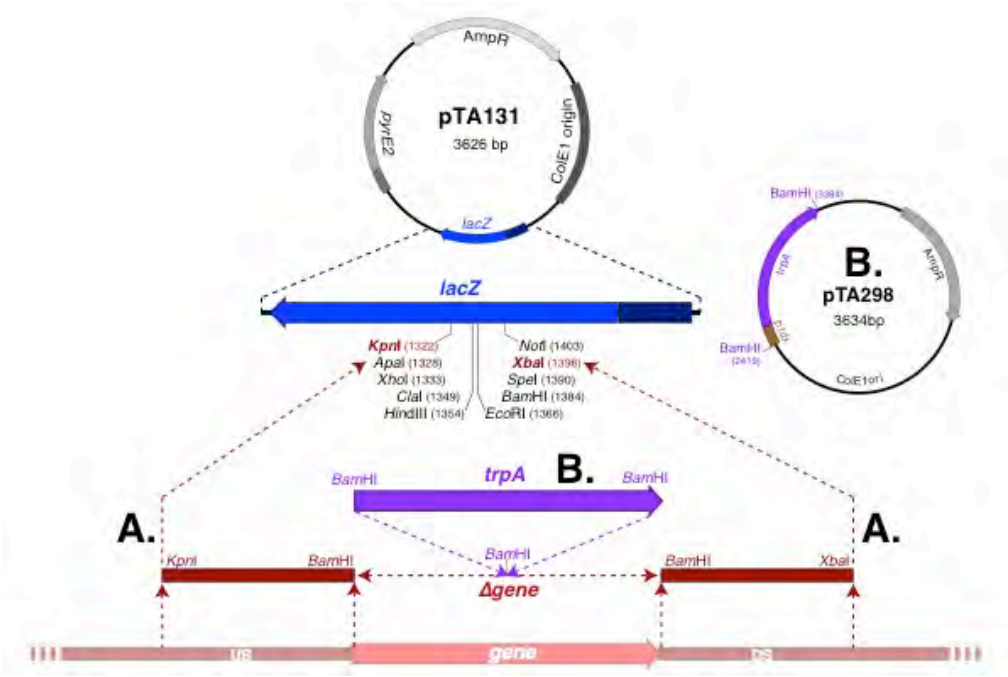


Figure 2.2: **Schematic for generating deletion constructs by PCR. (A)** Δ*gene* construct is generated by amplifying upstream (US) and downstream (DS) flanking regions of the gene of interest by PCR. This example shows primers with *Kpn*I/*Bam*HI US sites and *Bam*HI/*Xba*I DS sites. US and DS regions ligated at internal *Bam*HI sites and inserted into pTA131 at *Kpn*I and *Xba*I sites. **(B)** Optional. For making *Δgene::trpA+* construct. *trpA* gene taken from pTA298 using *Bam*HI and inserted between the US and DS regions at the internal *Bam*HI site. *AmpR* (ampicillin resistance gene *E. coli*), *colE1 origin (E. coli* origin of replication), *lacZ* (*β-galactosidase* gene used for blue/white selection in *E. coli*), *pyrE2* (uracil biosynthesis in *H. volcanii*)

Generating a gene replacement construct

Gene replacement constructs were made by inserting the gene of interest (usually containing a point mutation) along with its flanking regions into pTA131. The protocol for doing this varied depending on the mutation. For details on individual gene replacement plasmids constructed for this study see Chapter 3: *Plasmid and Strain Construction*. Section 3.1.2: *Gene Deletion and Replacement Plasmids*

Gene Deletion / Replacement

Deletion constructs were integrated at the gene locus by transformation of *H. volcanii ΔpyrE2* strains. The strains were plated on Hv-Ca (+ necessary additives) to select for the integrated *pyrE2*-marked plasmid (pop-in). A pop-in colony was picked and grown (non-selectively) overnight in a 5 ml Hv-YPC (+Thy) culture until A_{650} = 1.0. This culture was diluted 1/500 into a fresh

Hv-YPC (+Thy) culture and the growth and dilution repeated. By relieving selection for uracil, the integrated plasmid and native gene are lost from the chromosome. To select for *pyrE2*- pop-outs, the final culture was plated into Hv-Ca +5-FOA (+ necessary additives). Depending upon the location of the recombination event either wild-type or deletion mutants can be recovered. Replacing the gene with a selectable marker such as *trpA*+ enables direct selection for deletion mutants. Colonies were then restreaked onto selective media and tested for the desired genotype. A schematic is shown in Figure 2.3

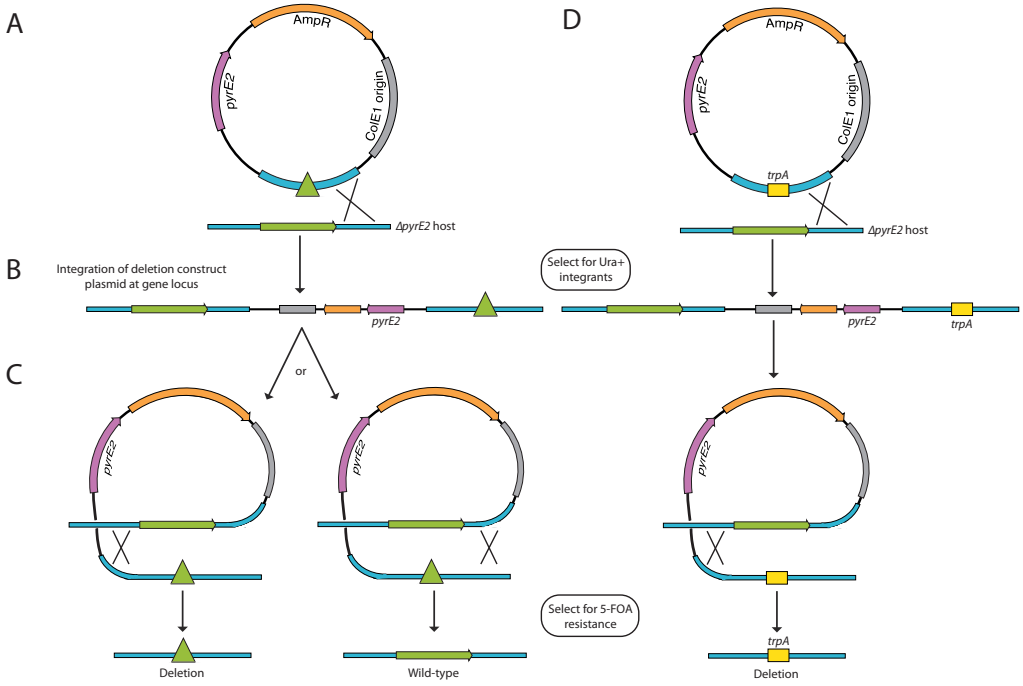


Figure 2.3: **Gene deletion by pop-in/pop-out. (A)** *ΔpyrE2* strains are transformed with a *pyrE2*+ deletion construct. **(B)** Pop-ins are selected for by their ability to grow on media lacking uracil **(C)** Relieving uracil selection allows for pop-out to occur, recombination between the regions of homology can be either upstream (left) or downstream (right). Pop-out is selected for by plating on 5-FOA The resulting gene locus will either be a deletion of wild-type **(D)** Replacing the gene with *trpA*+ marker enables direct selection for deletion mutants.

Deletion of *radA*

Recombination is essential for the pop-in and pop-out steps of the gene deletion and gene replacement protocol. RadA is a recombinase that facilitates the process of recombination; therefore, for a *radA* deletion to occur the strain must be complemented with an episomal copy of the *radA* gene during the pop-out step. *ΔpyrE2 ΔtrpA* strains were first transformed with the *ΔradA*

deletion construct (pTA83) and plated on selective media. Integrants were then transformed with a *radA*+ expressing episomal plasmid with Mevinolin resistance (pTA637), for *in trans* complementation of *radA*. To generate a pop-out, a colony was picked and grown in YPC+Mev to an A650 of 1.0, then diluted 1/100 in fresh YPC+Mev. When this culture was grown to an A650 of 1.0, it was diluted 1/100 in fresh YPC to allow loss of the episomal *radA*+ *MevR* plasmid from the strain. Cultures were then plated onto Hv-Ca +5-FOA to select for pop-outs that had also lost the episomal plasmid (*pyrE2*-). Colonies were restreaked onto selective media and tested for the desired genotype. This is shown in Figure 2.4.

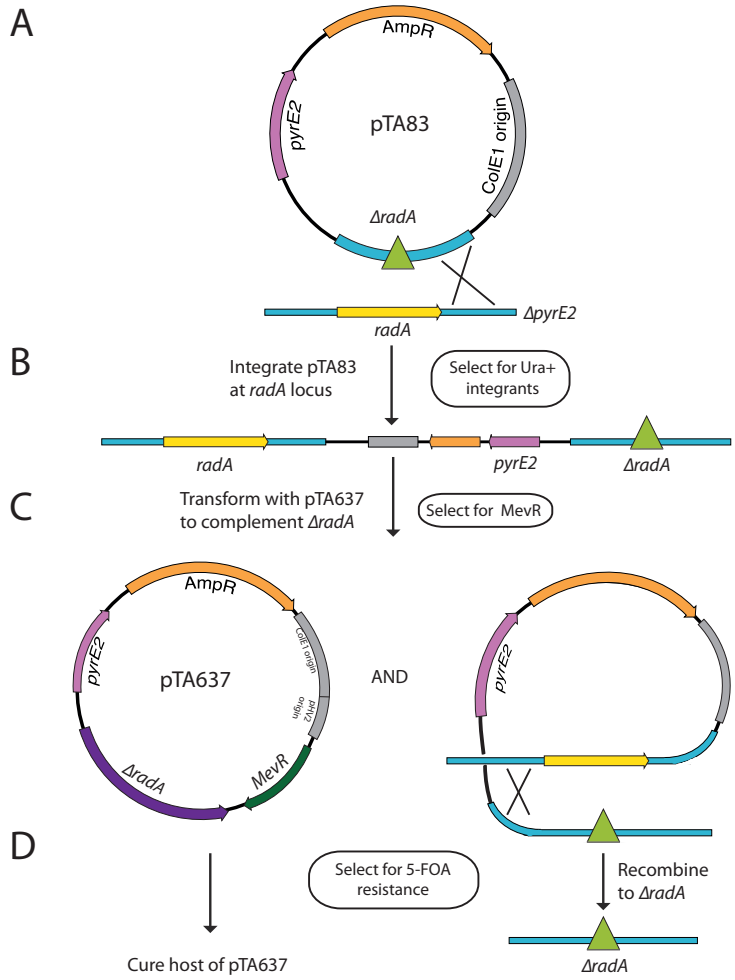


Figure 2.4: ***radA* gene deletion. (A)** A *ΔpyrE2, MevR*- strain is transformed with pTA83, a deletion construct containing a *radA* deletion and *pyrE2* marker (pop-in). **(B)** The strain is then transformed with pTA637 an episomal plasmid containing *radA* and *pyrE2* and *MevR* markers. **(C)** Pop-out is able to occur due to the presence of *radA*+ on pTA637. **(D)** Strains are selected for both the pop-out of pTA83 and the loss of pTA637 by plating out on 5-FOA (Delmas et al 2009).

2.2.6 Genotype Screening

A number of methods are used to screen for the presence or absence of a specific gene in pop-out strains. If the strain has a selectable phenotype (e.g. *ΔtrpA)* then colonies can be plated on selective media. However, *H. volcanii* is polyploid and so strains could be merodiploid, meaning that pop-outs could have a mixture of mutant and wild-type alleles on different chromosome copies. To select for strains that are homozygous, colony hybridization and Southern blotting must be carried out. For both of these methods, DNA is denatured and transferred to a positively charged membrane by either colony lift or vacuum transfer.

Colony Lift

20×SSPE: 3 M NaCl, 230 mM NaH_2PO_4, 32 mM EDTA, pH 7.4.
Denaturing Solution: 1.5 M NaCl, 0.5 M NaOH.
Neutralising Buffer: 1.5 M NaCl, 0.5 M Tris·HCl, 1 mM EDTA.

Candidate colonies and controls were patched on to Hv-YPC (+Thy if required) plates using sterile wooden toothpicks and incubated at 45ºC for ~3 days. The patched colonies were lifted from the plate by placing a circle of Bio-Rad Zeta-Probe GT positively charged membrane onto the surface for 1 minute. The membrane was transferred colony side up to Whatman paper soaked in 10% SDS for ≥5 minutes to lyse cells. The membrane was then transferred to Whatman paper soaked in denaturing solution for ≥5 minutes to denature proteins and DNA, then transferred to Whatman paper soaked in neutralising solution for ≥5 minutes. Neutralisation was repeated and the filter was washed briefly for 30 seconds in 2 × SSPE before being air-dried. DNA was crosslinked to the membrane with 120 mJ/cm^2 UV.

Southern Blot Vacuum Transfer

20× SSPE: 3 M NaCl, 230 mM NaH_2PO_4, 32 mM EDTA, pH 7.4.
Denaturing Solution: 1.5 M NaCl, 0.5 M NaOH.

Purified *H. volcanii* genomic DNA was digested with enzymes that cut ~500-1000 base pairs upstream and downstream of the gene of interest. The digested DNA was run on a 200 ml 0.75% TAE gel for 16 hours at 50 V with buffer circulation and stained with 0.5 μg/ml of ethidium bromide for 30 minutes.

Gel-embedded DNA was acid-nicked for 15 minutes in 0.25 M HCl, washed for 10 minutes in dH_2O and denatured in denaturing solution for 45 minutes. A 15 × 25 cm Zeta-Probe GT membrane was soaked in dH_2O before equilibrating in denaturing solution. Vacuum transfer was carried out using a Vacugene XL gel blotter and Vacugene Pump (Pharmacia Biotech) for 1 hour at 40 mBar. Following transfer, the membrane was washed briefly in 2 × SSPE and air-dried before the DNA was crosslinked with 120 mJ/cm^2 UV.

Hybridisation

100× Denhardt's Solution: 2% Ficoll 400, 2% PVP (polyvinyl pyrrolidone) 360, 2% BSA (bovine serum albumin, Fraction V).
20×SSPE*:* 3 M NaCl, 230 mM NaH_2PO4, 32 mM EDTA, pH 7.4.
Prehybridisation Solution: 6 × SSPE, 1% SDS, 5 x Denhardts, 200 µg/ml salmon sperm DNA (Roche, boiled for 5 minutes prior to addition).
Hybridisation Solution: 6 × SSPE, 1% SDS, 5% dextran sulphate.
Low Stringency Wash Solution: 2 × SSPE, 0.5% SDS.
High Stringency Wash Solution: 0.2 × SSPE, 0.5% SDS.

Membranes from colony lifts or vacuum transfer of agarose gels were prehybridised for ≥3 hours at 65ºC in 40 ml prehybridisation solution containing denatured salmon sperm DNA (10 mg/ml). Radiolabelled DNA probes were made with 50 ng of DNA and 0.74 Mbq of [α-^{32}P] dCTP (Perkin Elmer). DNA was denatured at 100ºC for 5 minutes then incubated with the radioisotope and HiPrime random priming mix (Roche) for 15 minutes at 37ºC. The radiolabelled probe was then purified on a BioRad P-30 column and mixed with 10 mg/ml of salmon sperm DNA, followed by denaturing at 100ºC for 5 minutes and quenching on ice. For Southern blots, 1 µl of 1 µg/ml 1 kb ladder was also included in the radiolabelling reaction. The prehybridiation solution was replaced with 40 ml of hybridisation solution and the probe DNA added. Membranes were incubated overnight at 65ºC. The membranes were washed twice with 50 ml of low stringency wash solution, once for 10 minutes and then 30 minutes, followed by another two washes with high stringency wash solution, both for 30 minutes. Membranes were allowed to dry before being wrapped in Saran wrap and exposed to a phosphorimager screen (Fujifilm BAS Cassette 2325) for ≥24 hours. The screen was scanned using a Molecular Dynamics STORM 840 scanner.

2.2.7 Phenotyping of Haloferax volcanii

Standard Growth Assay

In order to test growth rate in liquid media, cultures were set up in 5 ml Hv-YPC broth and grown at 45ºC to an A_{650} = 0.2-0.8 (exponential phase). Cultures were diluted and grown again to an A_{650} = 0.2-0.8, then diluted to an A_{650} = 0.05 (early-log phase) and incubated for 1 hour at 45 ℃ with 8 rpm rotation. 250 µl of culture and appropriate blanks were added to the wells of a 48 well microtiter plate (Corning). The plate was sealed around the edges with microporous tape (Boots economy brand) and incubated at 45 ℃ with double orbital shaking at 425 rpm for 48 hours in a Epoch 2 Microplate Spectrophotometer (BioTek). Readings at A_{600} were taken every 15 minutes. The generation time was calculated by plotting the growth on a log2 scale and using the following equation 2.2.

$$G = \frac{t}{n}$$

$$n = \frac{\log b - \log B}{\log 2}$$

G = generation time
t = time
n = number of generations
b = end OD
B = start OD

Equation 2.2: **Equation to calculate the generation time of a liquid culture using the OD A_{600}.**

It was noticed that generation times of strains varied between experiments when measuring the growth using the Epoch 2 Microplate Spectrophotometer (BioTek). Therefore, in this study comparisons are ony made between sets of strains within the same experiment. i.e. strains that have been incubated on the same 48 well microtitre plate and with the A_{600} measured simultaneously during a single run on the the Epoch 2 Microplate Spectrophotometer (BioTek). Since generation times vary between experiments, the generation times stated are not absolute. However, the relationship between sets of strains has been observed to be consistent, meaning that comparisons of generation times between strains within a single experiment is reliable. Therefore, in this study growth curves generated by this method are used to illustrate the

differences in generation times between a set of strains rather than an exact determination of each generation time.

Ultraviolet (UV) Irradiation Sensitivity

1 colony was used to inoculate 5 ml of HV-YPC broth (+Thy if required) and grown at 45ºC overnight. The culture was diluted appropriately into 5 ml of fresh HV-YPC broth and grown until A650 ~0.35-0.4 was reached. Cells were diluted in 18% SW (10^{-1} – 10^{-6}), and duplicate 20 µl samples were spotted out on Hv-YPC agar (+Thy if required) and allowed to dry at room temperature. Plates were exposed to UV light (254 nm, 1 J/m^2/sec) and shielded from visible light to prevent DNA repair by photo-reactivation. Plates were incubated at 45ºC for 4-7 days and colonies counted. Survival fractions were calculated relative to an unirradiated control.

Mitomycin C (MMC) sensitivity

1 colony was used to inoculate 5 ml of HV-YPC broth (+Thy if required) and grown at 45ºC overnight. The culture was diluted into 5 ml of fresh HV-YPC broth and grown until OD of A650 ~ 0.35-0.4 was reached. Cells were diluted in 18% SW (10^{0} – 10^{-6}), and duplicate 20 µl samples were spotted out on Hv-YPC agar containing 0 – 0.02 µg/ml of MMC. Plates were allowed to dry before incubation, incubated at 45ºC for 4-7 days and colonies counted. Survival fractions were calculated relative to an untreated control. MMC plates were made fresh and used within 5 days.

Recombination Assay

The recombination frequency between a plasmid and the chromosome can be measured by using a pair of mutant *leuB* alleles (Lestini et al 2010). Strains derived from a H195 or a H164 background have a mutant *leuB* allele, *leuB-Ag1*, and therefore cannot grow on media lacking leucine. Recombination between this allele and the *leuB-Aa2* allele present on pTA163 results in a wild-type *leuB* allele, this strain can now grow on media lacking leucine, Figure 2.5.

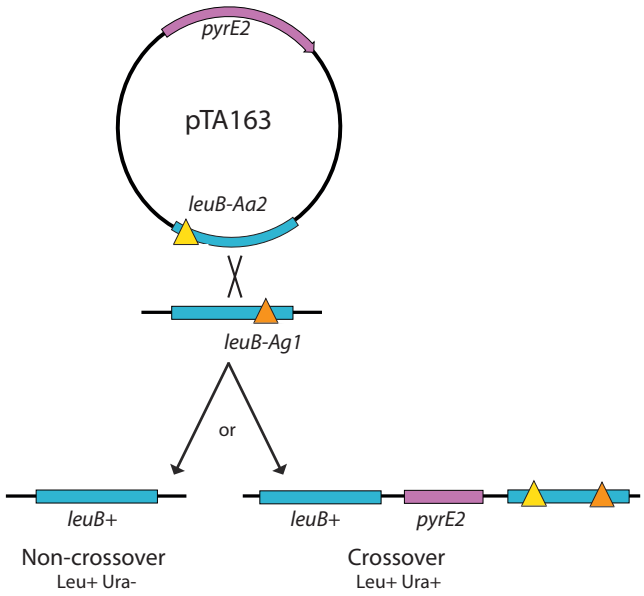


Figure 2.5: **Chromosome x plasmid recombination assays.** Δ*pyrE2* strains with a chromosomal *leuB-Ag1* allele (*leu*-) are transformed with pTA163, containing *pyrE2* and *leuB-Aa2*. A recombination event between the plasmid *leuB-Aa2* allele and chromosomal *leuB-Ag1* allele generates a wild-type *leu*+ allele and strains can grow on media lacking leucine. Crossover and non-crossover events are measured by studying the proportion of transformants that have retained or lost the *pyrE2* marker found on pTA163: crossover recombinants are *pyrE2*+ (*ura*+), and non-crossover recombinants Δ*pyrE2* (*ura*-).

Strains were transformed with 1 µg of pTA163 (*pyrE2*+, *leuB-Aa2*), transformants were allowed to recover for 1.5 hours with no disturbance followed by 3 hours with rotation. Transformants were plated on Hv-Min +Trp +Ura at 10^{0} – 10^{-3} dilutions to select for cells that had undergone a recombination event between the plasmid *leuB-Aa2* allele and the chromosomal *leuB-Ag1* allele, generating a wild-type *leu*+ allele. Previous work has shown that reversion of this allele is rare and does not affect the results of this assay (Haldenby 2007). To determine the total viable count, transformants were also plated onto (non-selective) Hv-YPC at 10^{-4} – 10^{-6} dilutions.

To determine the proportion of crossover (CO) vs non-crossover (NCO) recombination events, colonies were patched in duplicate on Hv-Min +Trp to select for CO recombination events in cells that had integrated the plasmid and become *pyrE2*+ *leu*+ and then on Hv-Min +Trp +Ura as a control to ensure that all colonies patched were *leu*+. The fraction of CO events (*leuB*+ *pyrE2*+)

is derived by comparison to the total recombination frequency. The remaining recombination events (*leuB+*, *pyrE2-*) were NCO.

The transformation efficiency of a strain was determined by transforming with 1 µg of pTA354 (*pyrE2+*), transformants were allowed to recover for 1.5 hours with no disturbance followed by 3 hours with rotation. Transformants were plated on Ca +Trp at 10^{0} – 10^{-3} dilutions to select for cells that had taken up the plasmid and on Hv-YPC at 10^{-4} – 10^{-6} dilutions to determine the total viable count. The recombination frequency was normalised to the transformation efficiency for each strain.

Flow Cytometry

Flow cytometry was used to determine the DNA content and cell size of *H. volcanii* cells (Delmas et al 2013). 5 ml Hv-YPC broth was inoculated and grown at 45ºC with 8 rpm rotation in two successive dilutions until an A_{650} was reached. 10 µl of acridine orange solution (0.1 mg/ml) was added to 1 ml of culture and incubated at room temperature for 1 minute. The sample run on an Apogee A40 flow cytometer with settings of: sample flow 0.7 µl /minute, volume 110 µl, sheath pressure = 150 psi. PMT voltages were 450 V for forward light scatter (LS1) and 500 V for emission at 510-580 nm (FL1). Data was analysed using Flow Jo (Tree Star Inc.). Debris was excluded and at least 50,000 cells counted for each time point. Gating on the peak/area plots for LS1 and FL1 was used to exclude cell doublets. Data was analysed using Flow Jo (Tree Star Inc.).

2.2.8 Gene Expression by RT-PCR

In order to determine the level of gene expression, RNA was extracted from cells and an RT-PCR (reverse transcriptase PCR) carried out. Strains were treated as follows prior to RNA extraction:

Expression level following UV-irradiation

5 ml cultures of *H. volcanii* were grown over two successive overnights in 10 ml Hv-YPC (+Thy if required) to an $A_{650} \sim 0.5$. 1 ml of culture was pelleted at 3300 × *g* for 8 minutes at 25ºC and then resuspended in 18% salt water. The

culture was UV irradiated at 20 J/m^2, then centrifuged at 3300 × *g* for 8 minutes at 25ºC and resuspended in 1 ml of Hv-YPC (+Thy if required). The cells were left to recover in the dark for 30 minutes at 45ºC with 8 rpm rotation before RNA was extracted. A control sample was treated exactly as above but without UV irradiation.

Expression level following Mitomycin C (MMC) treatment

5 ml cultures of *H. volcanii* were grown over two successive overnights in 10 ml Hv-YPC (+Thy if required) to an A_{650} ~ 0.5. 1 ml of the culture was then treated with 2 µg/ml MMC. The sample was incubated at 45ºC with 8 rpm rotation for 1 hour before RNA was extracted. A control sample was treated exactly as above but with addition of 18% salt water instead of MMC.

RNA Extraction

All equipment was wiped down with 0.1 M NaOH 1mM EDTA followed by 0.1% DEPC-treated water. Filter tips, individually wrapped tubes and RNase-free chemicals were used and gloves were worn at all times.

Unbuffered Spheroplasting Solution: 1 M NaCl, 27 mM KCl, 15% sucrose, pH 7.5.

Cultures were centrifuged at 3300 x *g* for 8 minutes at 25 ℃ in a round bottomed tube. The cells were resuspended in 250 µl of unbuffered spheroplasting solution and transferred to a fresh 1.5 ml tube. 500 µl of Trizol LS (Invitrogen) was added, homogenized by vortexing then followed by a 5 minute incubation at room temperature. 250 µl of chloroform was then added and samples were vortexed for 30 seconds before incubation for 3 minutes at room temperature. The samples were then centrifuged at 4ºC at 14,000 x *g* for 10 minutes and the top aqueous layer was transferred to a fresh 1.5 ml tube. 500 %l isopropanol was added and incubated for 10 minutes at room temperature. The samples were then centrifuged at 4 ℃, 14,000 x g for 10 minutes, the supernatant removed and the precipitated RNA washed in 1 ml of 75% ethanol and followed by centrifugation at 4ºC, 14,000 x *g* for 10 minutes. The supernatant was removed and the pellet left to air dry before resuspension in 45 µl of dH_2O. A 5 µl aliquot was used in a 1/10 dilution with TE to check

RNA concentration and purity by measuring 260:280 nm absorbance ratio of a spectrometer (Beckman Coulter DU 530). RNA samples were snap frozen using dry ice and stored at -80ºC.

DNase Treatment of RNA samples

5 µl of 10X TURBO DNase buffer and 1 µl of TURBO DNase (Ambion) were added to 45 µl RNA sample and incubated at 37ºC for 30 minutes. A further 1 µl of TURBO DNase was added and incubation repeated. 5 µl of DNase inactivation reagent (Ambion) was added, mixed by vortexing and incubated for 3 minutes at room temperature, vortexing occasionally. The samples were then centrifuged at 14,000 x *g* at room temperature and supernatant containing RNA transferred to a fresh tube. RNA samples were pipetted into 5 µl aliquots and snap frozen before being stored at -80ºC.

Reverse-Transcription PCRs (RT-PCRs)

To assess the level of gene expression, reverse-transcription PCRs (RT-PCRs) were carried out on RNA extracts using a OneStep RT-PCR kit (QIAGEN).

100 ng of template RNA was used in the following RT-PCR reaction (final volume 25 µl): 100 ng of template RNA, 1 × QIAGEN OneStep Buffer, 400 µM of each dNTP, 0.6 µM of each primer, 5-10 units RNase inhibitor (Super RNase.In), 1 × Q solution (aids transcription of GC rich templates), 1 µl enzyme mix (containing reverse transcriptase and DNA polymerase). Reaction conditions are shown in Table 2.11, all reactions were carried using a Techne TC-512 thermocycler and a no RT control was used for all RT-PCRs.

Table 2.11: **PCR conditions for RT-PCR**

Step	Conditions	
Reverse transcriptase	50ºC, 30 minutes	
Initial Denaturation	95ºC, 30 minutes	
Denaturation	94ºC, 30 seconds	26 cycles
Annealing	55ºC, 1 minute	
Extension	72ºC, 30 seconds/kb	
Final Extension	72ºC, 10 minutes	

2.2.9 Protein Overexpression and Purification

Protein Induction and Overexpression

Binding Buffer A: 20 mM HEPES pH 7.5, 2 M NaCl, 1 mM PMSF (phenylmethanesulphonyl fluoride), imidizole to desired concentration.
Binding Buffer B: 20 mM HEPES pH 7.5, 2 M NaCl, 1 mM PMSF

A starter culture was grown overnight at 45ºC in 5 ml Hv-Ca+ broth to A_{650} of ~1.0, then diluted 1/100 in to fresh 5 ml Hv-Ca+ broth and grown for ~8 hours at 45ºC until an A_{650} of ~ 1.0 was reached. The culture was diluted 1/200 in 50 ml Hv-Ca+ broth and grown for 24 hours until an A_{650} of ~ 0.5 was reached, then diluted 1/120 into 333 ml YPC broth and grown at 45°C with shaking (175 rpm) for 16 hours. When A_{650} = 0.5, 0.2 g (3 mM final concentration) of powdered tryptophan (Trp) was added to induce expression and the culture was incubated at 45ºC; after 1 hour a further 0.1 g Trp (~4.5 mM final concentration) was added and incubated for another hour. If native levels of protein expression were to be analysed then the culture was not induced by Trp. The culture was centrifuged at 3300 × *g* for 8 minutes at 4ºC. If proteins were *his6*-tagged and to be purified by a Ni^{2+} column then cells were resuspended in 7 ml ice-cold Binding Buffer A (20 mM imidazole); if proteins were StrepII tagged and to be purified by StrepII column then cells were resuspended in 7 ml ice-cold Binding Buffer B. Cells were lysed by sonication (4-5 x 30 seconds at ≤ 8 μm amplitude) on ice, cell lysate was then transferred to a 15 ml round bottomed flask and centrifuged at 20,000 × *g* for 15 minutes at 4ºC. The sample was filtered sequentially through 0.8 μm, 0.45 and 0.2 μm μm filters to remove DNA.

Charging of Ni^{2+} beads

To allow purification of *his6*-tagged proteins, IMAC Sepharose 6 Fast Flow beads (GE Healthcare) were charged with Ni^{2+}. Beads (0.5 ml per column) were washed twice with ≥2 volumes of dH_2O, and then equilibrated for 30 minutes with ≥0.2 volumes of 0.2 M NiSO4 with rotation at 4ºC. Two further washes with dH_2O were performed, the beads were then washed once with 5 ml of Buffer A (500 mM imidazole) then a further three times with Buffer A

(20 mM imidazole). Beads were resuspended in 0.5 ml Buffer A (20 mM imidazole) per sample.

Ni^{2+} gravity column

Cell lysates were incubated for 1 hour, rotating at 4ºC with 0.5 ml Ni^{2+} charged beads. The slurry was applied to a Poly-Prep column (Bio-Rad) and the flow-through collected and reloaded onto the column. The column was washed 3 x with 4 ml of ice-cold Binding Buffer A (20 mM imidazole) and bound protein was eluted with 2 column volumes of binding buffer containing 100, 200 and 500 mM imidazole. All protein purification steps were performed at 4°C.

Protein Concentration

Protein samples were concentrated using Vivaspin 20 ultrafiltration spin columns (GE healthcare). These columns could also be used for diafiltration, to remove imidazole after Ni^{2+} protein purification methods, in preparation for further protein purification by a Strep-Tactin column. A column with molecular weight cut off (MWCO) appropriate to the protein was selected. Protein samples were added to a Vivaspin column that has been pre-equilibrated with dH_2O and the appropriate protein buffer, then spun at 8000 × *g* in a swing bucket centrifuge at 4ºC until the desired final volume was reached. For application onto a StrepII column after Ni^{2+} protein purification, 1-2 ml of protein sample is desired.

Strep-Tactin gravity column

Elution Buffer: 20 mM HEPES pH 7.5, 2 M NaCl, 1 mM PMSF, 5 mM D-desthiobiotin (IBA)

For protein samples that have had no prior Ni^{2+} purification, use double the amount of each bead, wash buffer and elution buffer volumes stated below. For protein samples that had already been subjected to Ni^{2+} purification prior to StrepII purification, the following amounts of beads, wash volumes and elution volumes were used. 0.5 ml *Strep*-tactin Sepharose (IBA) was applied to a Poly-Prep column (Bio-Rad) and 2 x 1 ml Buffer B was added to equilibrate beads. Protein sample was applied to the column and flow through was collected. Flow through was applied to the column and this step was repeated again. The

column was washed with 5 x 0.5 ml Buffer B. Bound protein was eluted by applying 0.4 ml, 0.7 ml and then 0.4 ml of Elution Buffer to the column. All protein purification steps were performed at 4°C.

Protein Storage

For the short term, elutions were stored at 4ºC. For long term, proteins were snap frozen and stored at -80ºC with 10% glycerol.

Protein Precipitation

If samples required concentration prior to loading on an SDS-PAGE gel, proteins were precipitated using TCA (trichloroacetic acid).

4× Resuspension Buffer: 4% SDS, 0.2 M Tris pH 7.4, 0.15 M NaOH.

$1/10^{th}$ volume of 0.15% deoxycholate was added to samples, mixed by vortexing and incubated at room temperature for 10 minutes. $1/10^{th}$ of the original volume of 72% TCA (trichloroacetic acid) solution was then added and incubated for 5 minutes at room temperature. Samples were pelleted at 16,000 × *g* for 8 minutes and the supernatant removed. The pellet was resuspended in a 3:1 solution of 1 × resuspension buffer (protein sample buffer). The protein was then heat denatured by boiling at 94ºC for 5 minutes in 1× Laemmli buffer and analysed by SDS-PAGE.

SDS-Polyacrylamide Gel Electrophoresis

12% SDS-PAGE gel (resolving): 12.5% acrylamide/bisacrylamide Protogel (National Diagnostics), 0.37 M Tris (pH 8.8), 0.1% SDS 0.05% AMPS (ammonium persulfate), 0.05% TEMED (tetramethyleethylenediamine).
3.0% SDS-PAGE gel (stacking): 3% acrylamide/bisacrylamide Protogel, 0.25 M Tris (pH 6.8), 0.2% SDS, 0.125% AMPS, 0.125% TEMED, 0.77 ml H_2O.
SDS-PAGE running buffer: 0.25 M Tris, 1.92 M glycine, 1% SDS.
Laemmli buffer (4×): 50 mM Tris pH 6.8, 100 mM DTT, 2% SDS, 0.1% bromophenol blue, 10% glycerol.

Protein samples analysed using SDS-PAGE (sodium dodecyl sulphate polyacrylamide gel electrophoresis). Gels were cast in Novex cassettes (Invitrogen). A 12.5% resolving gel was poured with a layer of isopropanol on top to leave a flat surface. Once set, the isopropanol was washed off and a 3% stacking gel poured and a comb inserted. Protein samples were mixed with a

1/4 by final volume of 4× Laemmli buffer, denatured by boiling at 94 ℃ for 5 minutes and run alongside a PageRuler size ladder (Fermentas). Gels were run for ~1 hour 20 minutes in 1× SDS-PAGE running buffer (200 V, 36 mA per gel) and stained with PageBlue Protein Staining Solution (Thermo Scientific) for visualization of proteins.

Mass Spectrometry and Protein Identification

Mass spectrometry was performed by Dr Susan Liddell (Proteomics Facility, University of Nottingham). Proteins in gel bands were reduced using DTT (dithiothreitol) 10 mM, carboxyamidomethylated using IAA (iodoacetamide) 55 mM, and digested with Trypsin Gold (Promega) on a robotic platform for protein digestion (MassPREP station; Waters, MA, USA). The resulting peptides were analysed by liquid chromatography-electrospray ionization-tandem mass spectrometry (LC-ESI-MS/MS).

Tryptic peptides were separated and delivered on-line to an LTQ FT Ultra MS (Thermo Fisher Scientific, MA, USA) fitted with a nanoelectrospray ionisation source, via an Ultimate 3000 Nano LC system (Dionex/Thermo Fisher Scientific, MA, USA) (access to LC-MSMS courtesy of Dr Neil Oldham, School of Chemistry, University of Nottingham) fitted with a C18 reverse phase, 75 μm I.D., 15 cm column (Jupiter 4 μm Proteo 90 Å, Phenomenex, column made in-house, courtesy of David Tooth).

The software package Xcalibur™ (Version 2.0 SR2, Thermo Fisher Scientific, MA, USA), was used to acquire and process the RAW data files. The MS/MS data were searched on the MASCOT search engine (Version: 2.3.01, Matrix Science Ltd, London, UK) using the MS/MS ions search tool. MASCOT data reported in this thesis represent output from searches against our custom *Haloferax volcanii* database which consists of the entire set of 4,111 predicted protein sequences, supplemented with tagged construct sequences. Carbamidomethylation of cysteine and oxidation of methionine were set as variable modifications. One missed cleavage by trypsin was accepted. Other

than file type (.mgl) and instrument type (ESI-FTICR), all remaining search values were the pre-set defaults.

2.2.10 Halophilic Virus Isolation

Halophilic Virus Sample collection

On the 13th October 2011, 5 litre samples were collected from evaporation pools at a salt works outside Eilat, Israel these samples were labelled Eilat 1, 2 and 3. On the 14th October 2011, four more 5 litre water samples were collected from the west shore of the Dead Sea, Israel at Ein Gedi, Kalya, Mitspe Shalem and a pool near to the shore at Mitspe Shalem. Table 2.12 for co-ordinates of sample locations. Samples were collected in a plastic beaker and crudely filtered at the location through fabric to remove brine shrimp and debris. For transportation purposes the Ein Gedi and Kalya samples were pooled, as were the Eilat 1 and 3 samples. The samples were then decanted into 5 or 10 litre jerry cans, tightly sealed with parafilm and transported back to England by surface mail.

On the 29th January 2013, 50 litre sample was collected from a pond at the Bras del Port salterns, Alicante, Spain with help from Rodriguez-Valera Lab at the University Miguel Hernandez, Alicante. The sample was collected in a plastic beaker and decanted into 2 x 25 litre jerry cans, tightly sealed with parafilm and transported back to England by airmail.

Table 2.11: **Halophilic virus sampling locations.**

Sample ID	Country of Origin	Latitude	Longitude	Sample Size	Date Collected
Eilat 1	Israel	29º 33' 33" N	34º 57' 41" E	5 L	13th Oct 2011
Eilat 2	Israel	29º 33' 39" N	34º 57' 45" E	5 L	13th Oct 2011
Eilat 3	Israel	29º 33' 33" N	34º 57' 41" E	5 L	13th Oct 2011
Ein Gedi	Israel	31º 27' 35" N	35º 24' 00" E	5 L	14th Oct 2011
Kalya	Israel	31º 45' 41" N	35º 30' 13" E	5 L	14th Oct 2011
Mitspe Shalem	Israel	31º 34' 54" N	35º 24' 42" E	5 L	14th Oct 2011
Mitspe Shalem Pool	Israel	31º 34' 54" N	35º 24' 44" E	5 L	14th Oct 2011
Bras del Port Saltern ID - #30	Alicante, Spain	38º 11' 47.4"N	0º 35' 0.8" W	50 L	29th Jan 2013

NaCl Concentration Determination of Water Samples

In order to determine the NaCl concentration of water samples, conductivity readings were taken using an AKTA Prime Plus (Amersham Biosciences). Firstly, 10 ml of dH_2O was washed through the AKTA to zero the conductivity. Once zeroed, 5 ml of NaCl solutions with known concentrations ranging from 0 - 2.5 M were injected into the AKTA and the conductivity (in mS/cm) was taken to generate a standard curve. 5 ml of the seawater samples were then injected into the AKTA and with the conductivity readings the salt concentration of each sample was calculated from the standard curve. All samples were pre-filtered through a 0.2 μm filter to remove debris.

Viral Enrichment

5 ml Hv-YPC Broth was inoculated with wildtype *Haloferax volcanii* and incubated at 45ºC with rotation. When the OD_{650} had reached 0.5, 1 ml of seawater sample that had been filtered through a series of 5 μm, 0.45 μm and 0.2 μm filters was added to the culture. The virus-inoculated culture was incubated overnight at 42ºC with rotation. Following incubation, the culture was centrifuged at 15,557 x *g* for 10 minutes at 25ºC to pellet the *Haloferax volcanii* cells. 10% PEG600 was added to the supernatant containing the enriched viruses and then incubated at 4ºC with rotation. This sample was then spun at 15,557 x *g* for 30 minutes to pellet the viruses, the pellet was then resuspended in 1 ml of 18% SW.

Viral Plaque Assay

5 ml of Hv-YPC was inoculated with wild type *Haloferax volcanii* and incubated at 45ºC with rotation until the OD_{650} had reached 0.5. 300 μl of the culture was added to 3 ml of Soft Hv-YPC agar (0.5% Agar) pre warmed to 52ºC, vortexed, quickly poured over an Hv-YPC agar plate and left to set for 15 minutes at room temperature. Once set, 20 μl of enriched virus sample was spotted onto the plate and left to soak in for 30 minutes at room temperature. The plates were then incubated upright in a plastic bag at 45ºC, after 24 to 48 hours the plates were inspected for viral plaques.

Pre-filtration of Sea Water

To remove large particles and cellular matter, the seawater samples were sequentially filtered through a fiberglass filter membrane (Millipore) and then either a 0.45 μm or 0.22 μm nitrocellulose membrane (Millipore). The membranes were held in a 142 mm stainless steel filter holder and the sea water was driven across the filter using a high performance peristaltic pump (Cole Parmer Masterflex I/P) at 0.5 litre/min. The filtrate was collected in a clean 5 litre jerry can and kept at 4ºC. Following each filtration the tubing and filter apparatus were flushed with 0.1 N of NaOH heated to 40ºC and 5 litre of dH_2O.

Tangential Flow Filtration of Sea Water

To concentrate and further purify the virus particles, the filtrate from the pre-filtration protocol was circulated through a tangential flow filtration cartridge (Millipore) with a 30 kDa molecular weight cut off. Virus particles are retained within the membrane, whereas small molecules and water cross the membrane and are collected in the permeate. Unlike normal flow filtration, in tangential flow filtration (TFF) the sea water is flowed in at a tangent to the membrane, this means that the virus particles do not build up on the membrane surface but are swept along and therefore can be collected in the retentate. The waste was collected in the permeate. The sea water was driven through the TFF using a high performance peristaltic pump (Masterflex I/P) at around 1.5 litre/min and filtration was carried out until the retentate was concentrated to around 300 ml. Prior to each filtration, the tubing and TFF were flushed through with 7 litres dH_2O and equilibrated with 1 litre of NaCL at the same concentration of the sample. Following each filtration the tubing and TFF were flushed with 0.1 N of NaOH heated to 40ºC, 5 litre of dH_2O and then stored under sterile dH_2O at 4ºC.

Polyethylene glycol (PEG) Precipitation of Viruses

Phage Buffer: 30 mM Tris-HCl pH 7.5, 2 mM $CaCl_2$, 1.25M NaCl, 88 mM $MgCl_2.6H_2O$, 85 mM $MgSO_4.7\ H_2O$ (filter sterilized)

Following concentration by TFF, the sea water samples collected from Israel

were precipitated by PEG. The sea water sample were adjusted to 1.35 M NaCl, and to precipitate the viruses PEG8000 (Fisher) was added at a final concentration of 10 % (w/v). The sea water samples were agitated until all of the PEG8000 had dissolved, then incubated overnight (for at least 10 hours) at 4ºC in the dark. When a white phase was visible the sample was centrifuged at 17,000 x *g* (Sorval SLA3000 rotor) for 30 minutes at 4ºC. The supernatant was discarded and the pellet allowed to air dry for 10 minutes. The PEG-virus pellet was resuspended in 5ml of Phage Buffer with gentle agitation for 1 hour. To remove PEG from the virus sample, 7 ml of chloroform was added and gently vortexed for 10 seconds and centrifuged at 3.5 x *g* for 10 minutes. This was repeated as many times as needed until the aqueous layer turned clear. The supernatant was transferred to a fresh tube and store at 4ºC.

Iron Chloride Precipitation of Viruses

10 g/L Fe Stock Solution: $FeCl_3$-$6H_2O$, 100 ml dH_2O (note: the amount of $FeCl_3$ to add is calculated based on the amount of Fe and not the amount of the salt counterpart)
Resuspension buffer: 0.1 M EDTA-Na_2, 0.2M $MgCl_2$, 0.125 M Tris, 0.125 Ascorbic acid, pH to 6

Following concentration by TFF, the sea water samples collected from Alicante, Spain were precipitated by iron chloride (John et al 2011). 1 ml of a 10 g/litre Fe stock solution was added to each 10 litres of filtered sea water sample. The samples was shaken vigorously and incubated for 1 hour at room temperature with occasional shaking. The Fe treated sea water was then filtered through a 1.0 μm polycarbonate (PC) membrane filter on top of a 0.8 μm Supor800 support filter. The membranes were held in a 142 mm stainless steel filter holder and the sea water was driven across the filter using a high performance peristaltic pump (Cole Parmer Masterflex I/P) up to 0.5 litre/minute. The $FeCl_3$ precipitated viruses is captured on the PC membrane filter. As filtration slowed down the polycarbonate membrane filter was replaced with a fresh one. PC filters were placed in a glass Hybaid tube, 10 ml of Resuspension Buffer was added and incubated at 4ºC for 24 hours in the dark with rolling. The Resuspension Buffer, containing the precipitated viruses was removed and stored at 4ºC.

Proteinase K and SDS Treatment of Viruses

To aid lysis of the virus particles, proteinase K and sodium dodecyl sulphate (SDS) was added to the sample at final concentrations of 50 μg/ml and 0.1% respectively and incubated at 37ºC for 1 hour.

Preparation of Virus Samples for RNA and DNA Extraction

TRIzol LS Reagent (Invitrogen) was added in a 3:1 ratio to the sample and gently pipetted up and down to homogenize the viruses. To permit complete dissociation of nucleoprotein complex, the homogenized sample was incubated at room temperature for 5 minutes. 0.2 ml chloroform was added per 0.75 ml TRIzol LS reagent used for homogenization, shaken vigorously for 15 seconds then incubated for 15 minutes at room temperature. The sample was centrifuged at 8,000 x *g* for 20 minutes to separate RNA, DNA and protein into three phases. The aqueous phase containing RNA and interphase/organic phenol phase containing DNA was kept and stored at 4ºC for further processing.

Viral RNA Isolation

For RNA work all equipment was wiped down with 0.1 M NaOH 1mM EDTA and then 0.1% DEPC-treated water. Filter tips, individually wrapped tubes and RNase-free chemicals were used and gloves were worn at all times. 0.5 ml of 100% isopropanol was added to the aqueous phase per 0.75 ml TRIzol LS Reagent used for homogenization and incubated at room temperature for 10 minutes. The sample was centrifuged at 8,000 x *g* for 15 minutes and the supernatant discarded. The RNA pellet was washed with 1 ml 75% ethanol per 0.75 ml TRIzol LS Reagent used for homogenization and then centrifuged at 7,500 x *g* for 5 minutes at 4ºC, the supernatant was discarded. The pellet was left to air dry for 10 minutes and resuspended in 50 μl of 0.1% DEPC- treated water.

Viral DNA Isolation

0.3 ml of 100% ethanol per 0.75 ml TRIzol LS Reagent used for homogenization was added to the interphase/organic phenol phase containing

DNA, inverted several times to mix and incubated at room temperature for 3 minutes. The sample was centrifuged at 2,000 x *g* for 5 minutes at 4ºC to pellet the DNA, and the supernatant was discarded. The DNA pellet was washed with 1 ml of 0.1 M sodium citrate in 10% ethanol per 0.75 ml TRIzol LS Reagent used for homogenization, incubated for 30 minutes at room temperature and centrifuged at 2000 x *g* for 5 minutes at 4ºC. The supernatant was discarded and the wash repeated once more. 2 ml 75% ethanol per 0.75 ml TRIzol LS Reagent used for homogenization was added, then incubated for 20 minutes at room temperature then centrifuged at 2000 x *g* for 5 minutes at 4ºC. The supernatant was discarded and the pellet was allowed to air dry. The DNA pellet was then resuspended in 50-300 µl of 8 mM NaOH, for long term storage at 4ºC the pH was adjusted with HEPES and EDTA was added.

2.2.11 Deep Sequencing of Viral Genomes

Library preparation and viral genome sequencing was carried out by Sunir Malla and Raymond Wilson, Bioinformatic analysis and support was carried out by Jo Moreton and Martin Blythe, all at the Deep Seq facility at the University of Nottingham.

Library preparation

From the isolated viral RNA, double stranded DNA was generated using the NEB kit and then fragmented. The adapters were then ligated (DNA adapters from Illumina True seq kit). The fragmented-ligated DNA was gel purified using E-gel. The purified DNA was PCR amplified, the average size of the library was 450bp and the total adapter length was around 100bp (around 50bp at each end).

Miseq sequencing

The library was sequenced by Illumina MiSeq**,** 2 x 250bp (paired-end).

Bioinformatics

The raw reads from MiSeq sequencing were in a fastq format. Adaptor sequences were trimmed from the reads using Illumina Experiment Manager software. The paired-end reads were analysed using FLASH (Fast Length

Adjustment of Short reads), this software merges over lapping paired-end reads (Magoc & Salzberg 2011). The reads were run though the short read error corrector Musket (Liu et al 2013). The reads where run through the *de novo* assembler CLC Assembly Cell and the contigs from this *de novo* assembly were then searched against the NCBI non-redundant (NR) database using BlastX. To remove any contaminating "host" sequences the contigs were screened against 67 known haloarchaeal genome sequences. The reads were aligned to *de novo* contigs using Bowtie 2 run through the de novo assembler CLC Assembly Cell again and aligned to known haloviral genome sequences.

Chapter 3: Plasmid and Strain Construction

3.1 Plasmid construction

This chapter details how plasmid vectors used in this study were constructed, all plasmids were confirmed by restriction digest and sequencing. In all figures, the following abbreviation apply: *AmpR* (ampicillin resistance gene, *E. coli*), *colE1 ori* (*E. coli* origin of replication), *f1(+) ori* (*E.coli* origin of replication), *hdrB* (thymidine biosynthesis, *H. volcanii*), *his6-tag* (hexahistidine tag), *lacZ (*β*-galactosidase* gene used for blue/white selection *in E.coli*), MCS (multiple cloning site), *p.fdx* (ferredoxin promoter), *pHV2* (*H. volcanii* origin of replication), *p.tnaA* (tryptophan inducible promoter), *p.lac* (promoter for *lacZ*), *pyrE2* (uracil biosynthesis, *H. volcanii*), *strepII-tag* (streptavidin tag), *t.L11e* (terminator), *t.Syn* (terminator).

3.1.1 Genomic Clones

Genomic clones were constructed using the standard *E. coli* cloning vector, pBluescript II SK+ (Figure 3.1). The multiple cloning site (MCS) is located within *lacZ* (ß-galactosidase gene). Expression of *lacZ* results in blue colonies when plated onto media containing X-gal, if *lacZ* is disrupted the colonies are white. Genomic clones are constructed by cloning a gene of interest along with upstream and downstream flanking regions of ~1kb each into the MCS of pBluescript II SK+. The protocol for this method is described in Chapter 2: *Materials and Methods*, Section 2.2.5: *Genetic Manipulation of H. volcanii*.

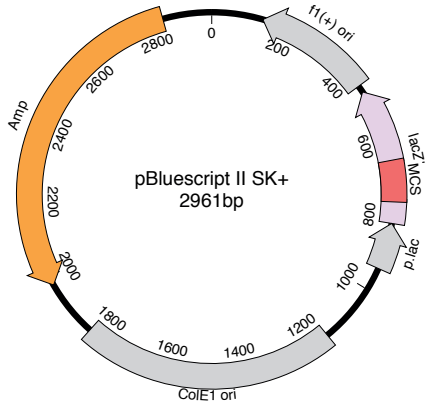


Figure 3.1: **pBluescript II SK+.** Standard *E. coli* cloning vector. The multiple cloning site (MCS) is located within *lacZ*, allowing for blue:white screening.

pTA415, *hel308* (HVO_0014) genomic clone

The *hel308* genomic clone, pTA415, was previously constructed by Zhenhong Duan in 2005.

pTA1364, *hel308b* (HVO_0971) genomic clone

Genomic DNA was extracted from wild-type *H. volcanii* (H26) and digested with *Xma*I and *Acl*I (Figure 3.2A). *Acl*I cuts 3572 bp upstream of *hel308b* and *Xma*I cuts 2937 bp downstream, producing an 8.4 kb fragment. Digested DNA was run on an agarose gel and bands between 8-9 kb extracted and ligated into pBluescript II SK+ at *Xma*I and *Cla*I sites, (Figure 3.2B). Transformants were patched onto LB+ Amp agar and transferred to a membrane via a colony lift. Colony lifts were probed with a 993 bp radiolabelled Hel308bF-Hel308bR PCR product amplified from genomic DNA. Patches that hybridised with the probe were restreaked and a colony PCR used to confirm the presence of the correct insert (data not shown).

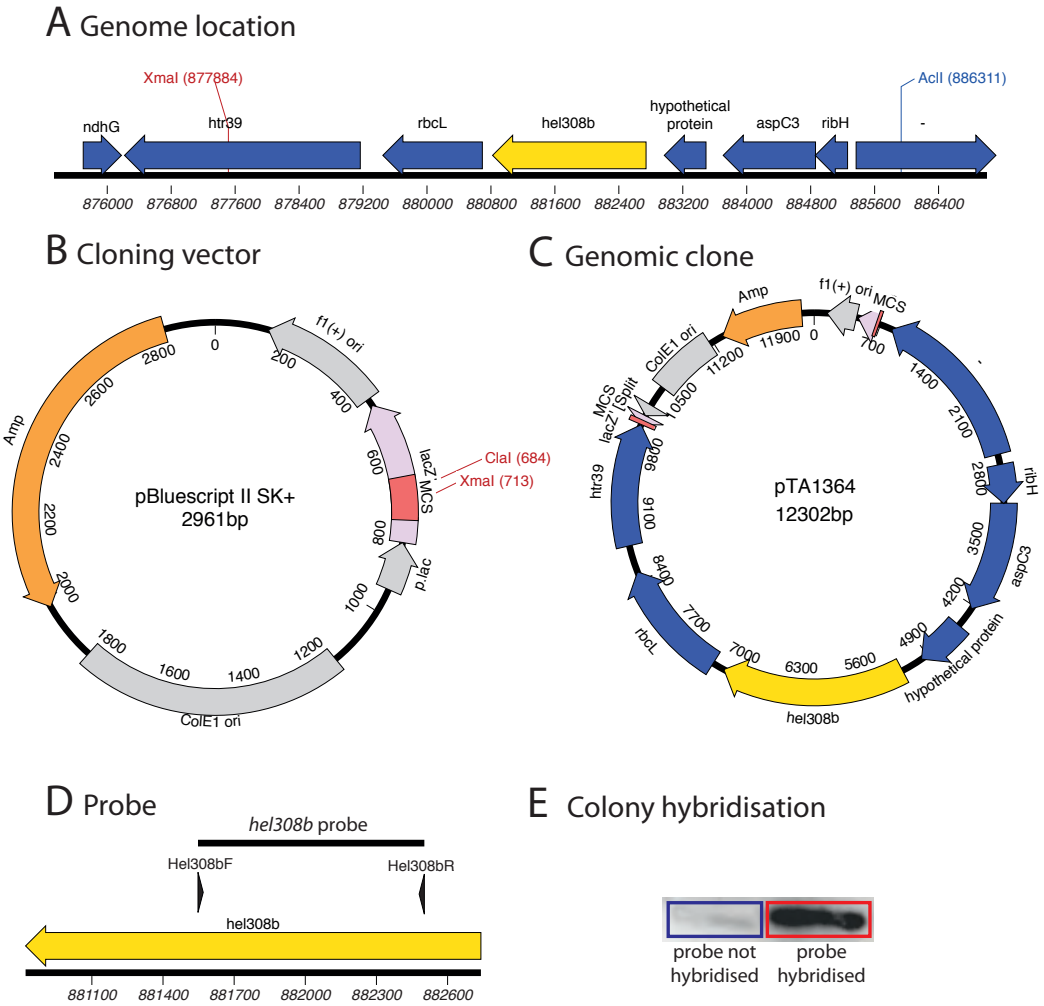


Figure 3.2: **Construction of *hel308b* genomic clone.** **(A)** The location on the *H. volcanii* genome containing *hel308b* (yellow), genomic DNA was digested with *Xma*I and *Acl*I. **(B)** pBluescript II SK+ plasmid where the *Xma*I to *Acl*I fragment was inserted at *Xma*I and *Cla*I sites. **(C)** Genomic clone pTA1364 constructed by inserting *Xma*I to *Acl*I fragment containing *hel308b* into pBluescript II SK+ **(D)** 993 bp *hel308b* probe (black) was constructed by PCR with primers Hel308bF and Hel308bR, was radiolabelled and used to probe colonies. **(E)** Colony hybridisation showing a patch to which probe hybridised to (containing *hel308b* genomic clone) in red and a patch to which the probe did not hybridise to (not containing *hel308b* genomic clone) in blue.

3.1.2 Gene Deletion / Replacement Plasmids

The pTA131 plasmid is utilised to generate gene deletion and gene replacement constructs (Figure 3.3). pTA131 is a derivative of pBluescript II SK+ that contains a *pyrE2* marker which encodes for orotate phosphoribosyl transferase, an enzyme involved in uracil biosynthesis (Allers et al 2004). Upstream and downstream flanking regions of the gene of interest are inserted into pTA131 and *pyrE2* is used to select for *H. volcanii* transformants that have

taken up pTA131 derived plasmid*s*, since Δ*pyrE2* strains cannot grow on Hv-Ca media lacking uracil.

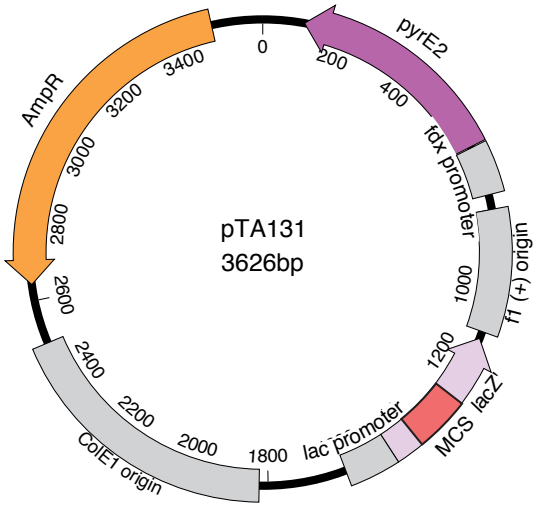


Figure 3.3: **pTA131**. Vector used for constructing gene deletion or replacement constructs for *H. volcanii* genetic manipulation. Based on pBluescript SK II+. The vector also contains *H. volcanii pyrE2* marker, which allows for selection in Δ*pyrE2* strains of *H. volcanii*.

3.1.2.1 Construction of plasmids for Chapter 4: Genetic Analysis of *hel308*

pTA1276, *hel308* deletion construct and pTA1277 *hel308::trpA+* deletion construct

The deletion constructs, pTA1276 (*hel308* deletion construct) and pTA1277 (*hel308::trpA+*) deletion construct, were previously constructed by Thorsten Allers in 2010.

3.1.2.2 Construction of plasmids for Chapter 7: Genetic and Phylogenetic Analysis of *hel308b*

pTA1368, *hel308b* deletion construct

From the genomic clone pTA1364, the upstream flanking region of *hel308b* was amplified using the primers dHel308b_U_KpnI_F and dHel308b_U_BamHI_R, which introduce *Kpn*I and *Bam*HI sites respectively.

The downstream flanking region was amplified using the primers dHel308b_D_BamHI_F and dHel308b_D_XbaI_R, which introduce *Bam*HI and *Xba*I sites respectively (PCRs 1 and 2 Figure 3.4A). The two PCR products were digested with *Bam*HI and ligated, the resulting product was ligated into pTA131 at *Kpn*I and *Xba*I sites to create the *hel308b* deletion construct, (Figure 3.4B).

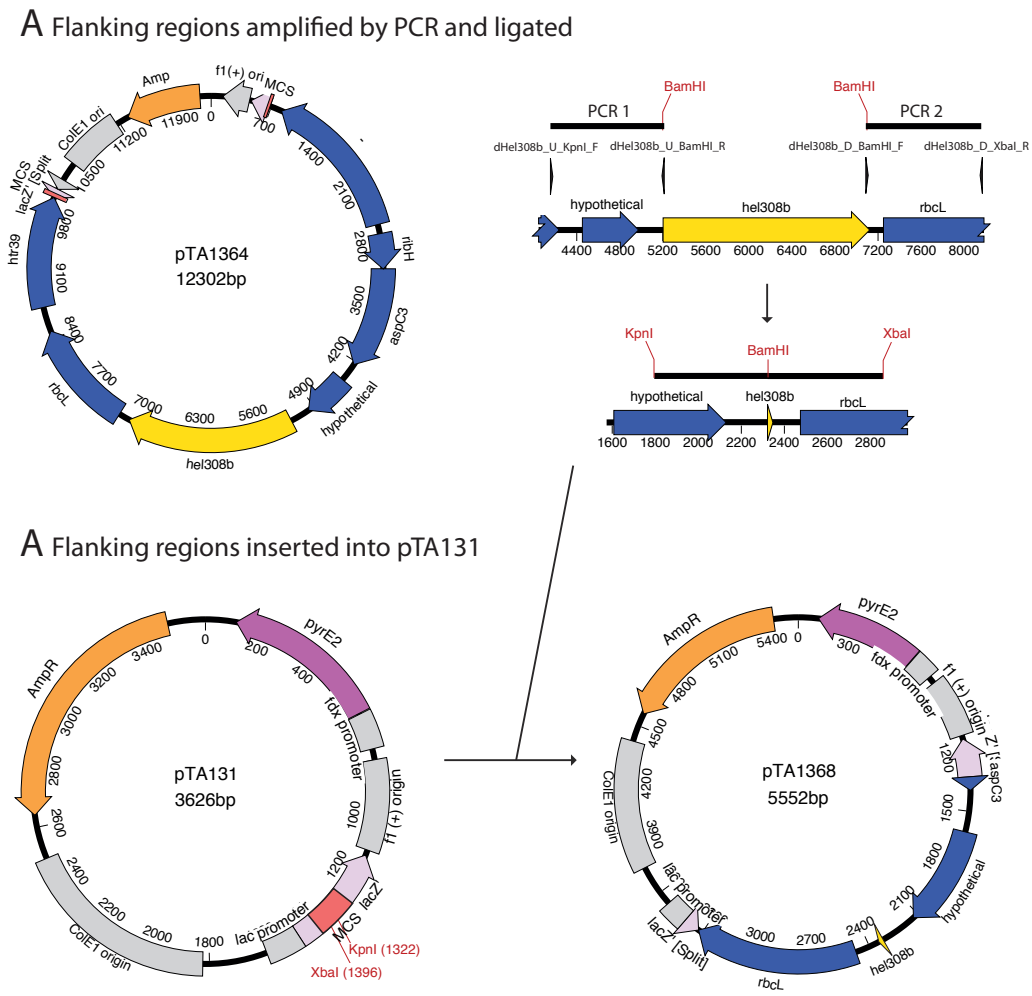


Figure 3.4: **Construction of *hel308b* deletion construct, pTA1368. (A)** Upstream and downstream flanking regions were amplified from pTA1364, digested with *Bam*HI and ligated to produce a construct with hel308b deleted **(B)** The resulting product was inserted into pTA131 at *Kpn*I and *Xba*I sites.

3.1.2.3 Construction of plasmids for Chapter 6: *in vitro* Analysis of Hel308

pTA1508, *tnaA* deletion construct

The *tnaA* genomic clone (pTA875) was digested with *Age*I, which cuts at positions 3267 and 4755 to remove 1332 bp of *tnaA*, leaving 15 bp of coding sequence remaining. The remaining plasmid was self ligated to remove the *tnaA* gene, producing the intermediate pTA1503. pTA1508 was digested with *Bam*HI and *Xba*I and the fragment containing *tnaA* upstream and downstream flanking regions (and *tnaA* deletion) was ligated into pTA131 at *Bam*HI and *Xba*I sites to create the *tnaA* deletion construct pTA1508 (Figure 3.5).

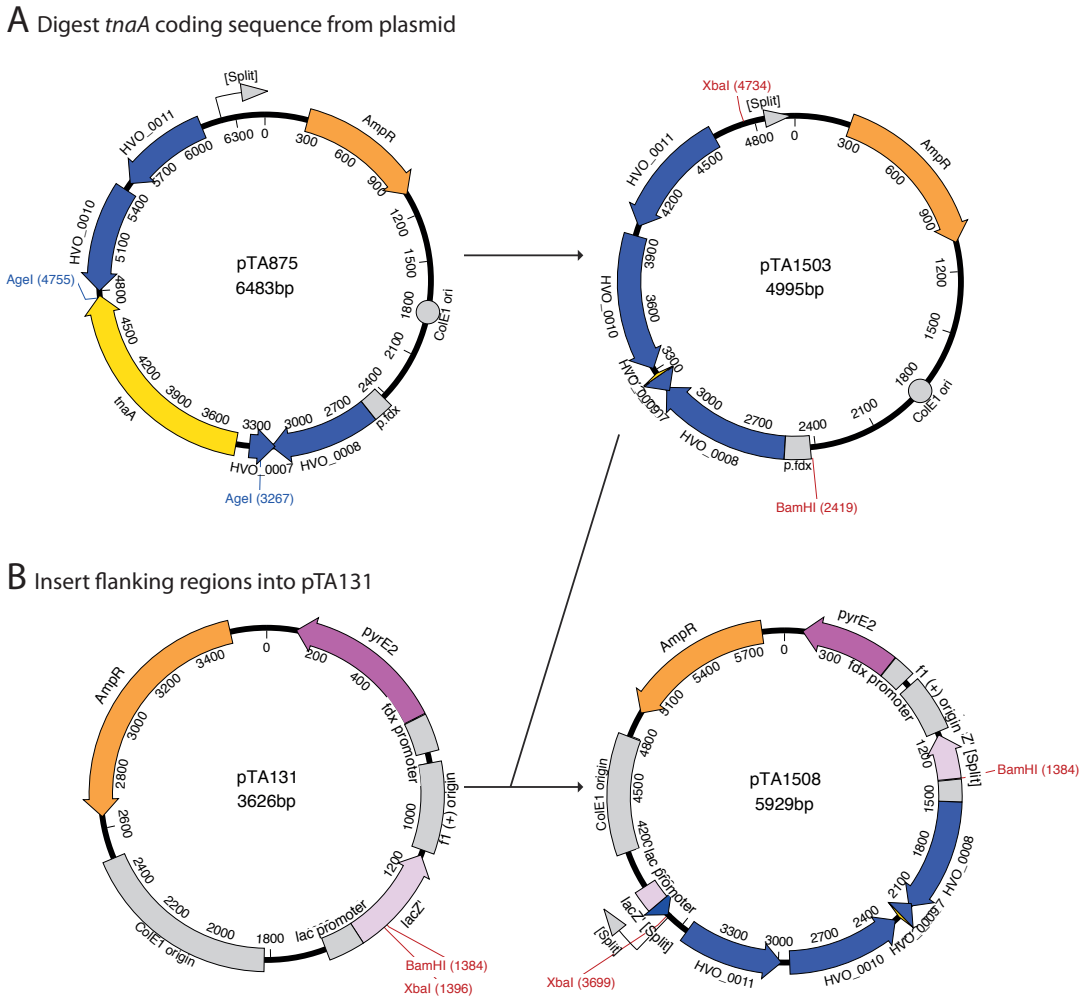


Figure 3.5: **Construction of *tnaA* deletion construct, pTA1508. (A)** A 1332 bp fragment of *tnaA* was removed from pTA875 by an *Age*I double digest to generate the intermediate plasmid pTA1503. A *Bam*HI and *Xba*I fragment containing *tnaA* flanking regions and therefore the *tnaA* deletion was ligated into pTA131.

pTA1615, *tnaA::hdrB+* deletion construct

pTA1185 was digested with *Bam*HI which cuts at positions 2419 and 3146 to liberate a fragment containing a *hdrB* marker, this fragment was blunt ended with Klenow. The *tnaA* deletion construct pTA1508 was digested with *Age*I, blunt ended with Klenow, dephosphorylated with Shrimp alkaline phosphatase and ligated to the *hdrB* containing fragment to generate a *hdrB+* marked *tnaA* deletion construct pTA1615 (Figure 3.6).

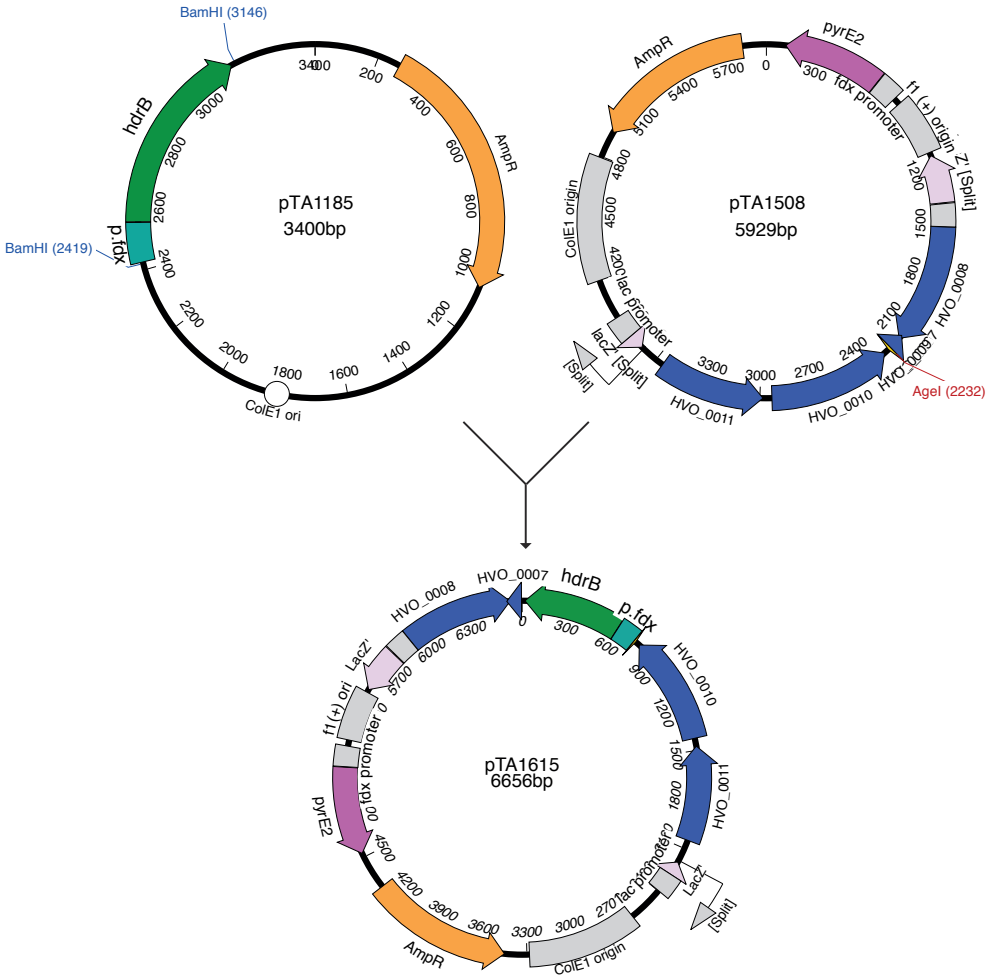


Figure 3.6: **Construction of *tnaA::hdrB+* deletion construct, pTA1615.** *hdrB* marker digested from pTA1185 at *Bam*HI sites and blunt end ligated into pTA1508 at an *Age*I site, to generate a *tnaA::hdrB+* deletion construct, pTA1615

3.1.2.4 Construction of plasmids for Chapter 5: Genetic Analysis of *hel038* Point Mutants

pTA1575, *hel308-H317G* gene replacement construct

The gene replacement plasmid pTA1575 was made to introduce an H317G mutation into *hel308* by PCR. *hel308* was amplified from the genomic clone pTA1316 by two PCRs (1 and 2 in Figure 3.7A). PCR 1 was carried out with the primers Hel308Fint and Hel308H317GR producing a 511 bp product, Hel308H317GR contains point mutations that will result in the generation of *hel308-H317G*. PCR 2 was carried out with the primers Hel308H317GF and hel308EcoR producing a 692 bp product; Hel308H317GF contains point mutations that will result in the generation of *hel308-H317G*. The binding sites in Hel308H317GF and Hel308H317GR overlap; therefore a third PCR was carried out mixing the products of PCR 1 and PCR 2 using the external primers Hel308Fint and hel308EcoR. This produced a 1183 bp product containing *hel308-H317G*. This PCR was digested with *Bsp*EI and *Eco*RI and resulting 977 bp fragment was used to replace the equivalent fragment in pTA1316.

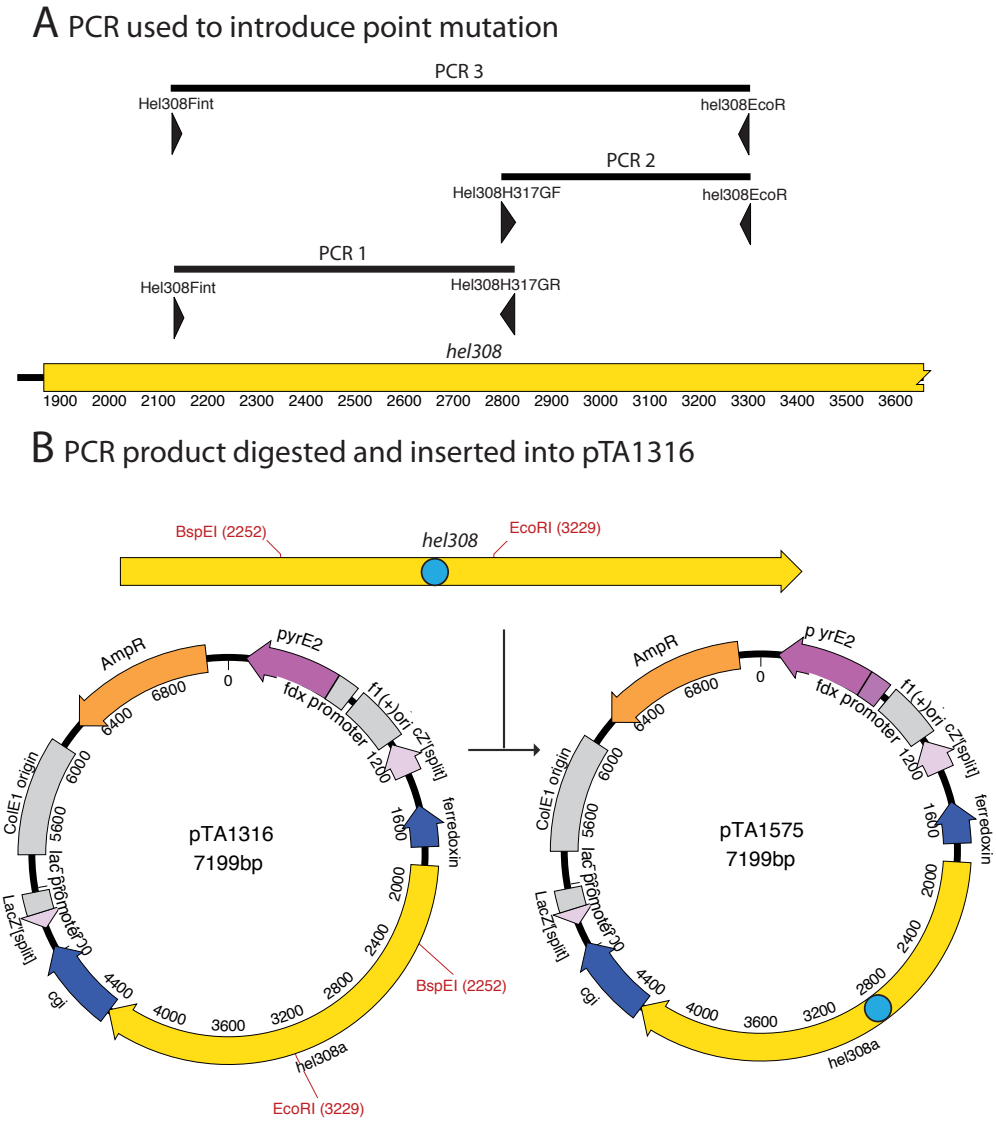


Figure 3.7: **Construction of *hel308-H317G* gene replacement construct, pTA1575. (A)** PCR was used to introduce a point mutation into the *hel308* gene, generating *hel308-H317G*. **(B)** The PCR product was digested with *Bsp*EI and *Eco*RI, the resulting fragment was used to replace the equivalent fragment in pTA1316.

Other *hel308* point mutant gene replacement constructs

All other hel308 point mutant gene replacement constructs were generated using the same principles as used to make pTA1575 (Figure 3.7). The binding site of the reverse primer in PCR 1 overlaps with that of the forward primer in PCR 2 to generate a point mutation. PCR 3 is carried out with external primers and uses the products from PCR 1 and 2 as a template. The product from PCR 3 is digested and used to replace the equivalent fragment in pTA1316, and so generating a *hel308* point mutant. Specific primers for PCRs 1, 2 and 3,

restriction sites used to digest PCR fragment and pTA1316 for each deletion construct are detailed in Table 3.1.

Table 3.1: **PCR and digests to generate *hel308* point mutant gene replacement plasmid constructs.**

Name	Point mutant	PCR1 Primers	PCR 2 Primers	PCR 3 Primers	PCR 3 and pTA1316 digests
pTA1545*	E422G	hel308NsiF, hel308-E422G-R	hel308-E422G-F, hel308EcoR	hel308NsiF, hel308EcoR	*Nsi*I, *Eco*RI
pTA1546*	D420A	hel308NsiF, hel308-D420A-R	hel308-D420A-F, hel308EcoR	hel308NsiF, hel308EcoR	*Nsi*I, *Eco*RI
pTA1576	E330G	Hel308FInt, Hel308E330GR	Hel308E330GF, Hel308EcoR	Hel308FInt, Hel308EcoR	*Bsp*EI, *Eco*RI
pTA1647	F316A	Hel308FInt, Hel308F316AR	Hel308F316AF, Hel308EcoR	Hel308FInt, Hel308EcoR	*Bsp*EI, *Eco*RI
pTA1648	R743A	Hel308-D420A-F, Hel308R743AR	Hel308R743AF, CgiBglR	Hel308-D420A-F, CgiBglR	*Eco*RI, *Blp*I

**dam+* plasmid originally constructed by Thorsten Allers, *dam-* version created during this study.

3.1.3 Episomal Plasmids for the Overexpression of Tagged Proteins

3.1.3.1 Construction of plasmids for Chapter 6: *in vitro* Analysis of Hel308

Plasmid constructs pTA1392 and pTA1403 were developed to allow for conditional overexpression of *H. volcanii* proteins with the option of a *his6*-tag and/or a *strepII* tag.

pTA1392, overexpression construct for N-terminally *his6-tagged* and/or a C-terminally *strepII-tagged* proteins

The 54 bp oligonucleotides StrepIIF and StrepIIR were annealed to create a double stranded DNA fragment containing a *strepII-tag* and the unique restriction sites *Pci*I, NspI, *Bmt*I, *Nhe*I, *Bst*XI, *Eco*RV and *Eco*RI. pTA1228 (overexpression vector containing an in frame *his6-tag* and a tryptophan inducible promoter, *p.tnaA*) (Brendel et al 2014) was digested with *EcoRI* and *Pci*I and into which the strepII oligonucleotide was ligated to create the *his6-tag* and *strepII-tag* containing overexpression vector pTA1392 (Figure 3.8).

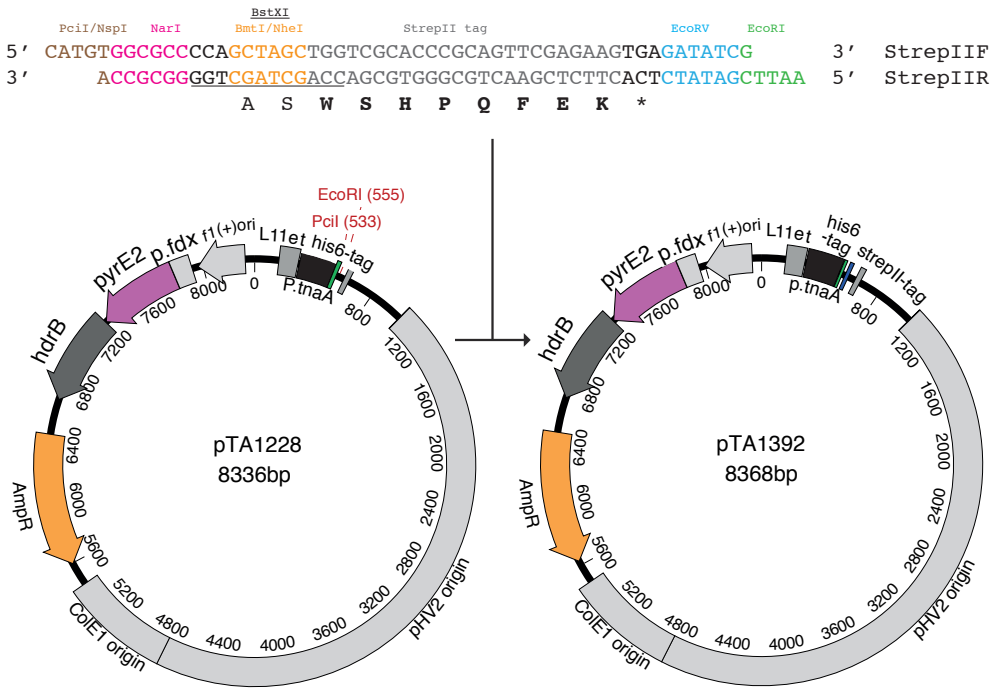


Figure 3.8: **Construction of the overexpression construct pTA1392.** Oligonucleotides StrepIIF and StrepIIR were annealed and ligated into pTA1228 at *Eco*RI and *Pci*I sites. pTA1392 is the overexpression construct that allows for N-terminally *his6-tagged* and/or a C-terminally *strepII-tagged* proteins under the control of a tryptophan inducible promoter *(p.tnaA)*

pTA1392 allows for in-frame insertion of a gene of interest with the option to have an N-terminal *his6-tag* and/or a C-terminal *strepII-tag* (Figure 3.9). To generate a construct without a N-terminal *his6-tag*, the start codon was replaced with an *Nde*I site and ligated with the *Nde*I end. For an N-terminal *his6-tag* the start codon was replaced with *Nco*I, *Bsp*HI, *Pci*I or an *Sph*I site and ligated with the *Pci*I or *Nsp*I end. To generate a construct without a C-terminal *strepII-tag*, an *Eco*RI *or Bam*HI site was incorporated after the stop codon and ligated, and for a C-terminal *strepII-tag* the stop codon of the gene was replaced with *Nhe*I or a compatible site *(Avr*II, *Spe*I *or Xba*I*)* and ligated with the *Nhe*I end.

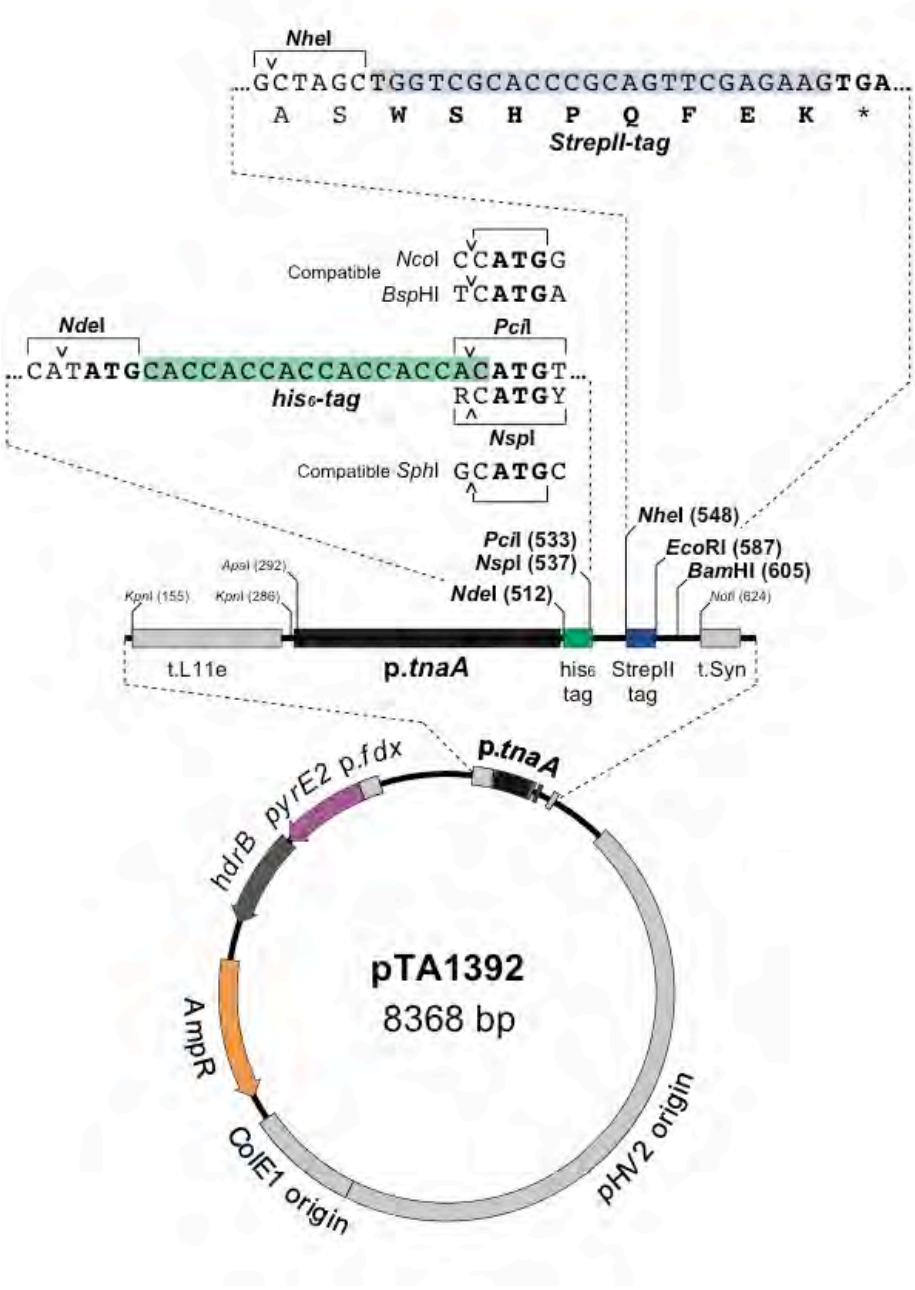


Figure 3.9: **pTA1392.** A plasmid for the conditional overexpression of N-terminally *his6* and/or C-terminally *strepII* tagged proteins. The plasmid contains a tryptophan inducible promoter (*p.tnaA*), *pyrE2* and *hdrB* markers for selection in *H. volcanii*. Restriction sites for inserting gene of interest are shown.

pTA1403, overexpression construct for N-terminally *strepII-tagged* and/or a C-terminally *his6-tagged* proteins

The 85 bp oligonucleotides StrepII_N_F and StrepII_N_R were annealed to create a double stranded DNA fragment containing a *strepII-tag* upstream of a *his6-tag* along with the unique restriction sites *Nde*I, *Pci*I, *Nsp*I, *Nar*I, *Nhe*I,

*Eco*RV and *Eco*RI. pTA1228 (overexpression vector containing an in frame *his6-tag* and a tryptophan inducible promoter, *p.tnaA*) was digested with *Nde*I and *Eco*RI , removing the existing *his6-tag* and into this plasmid the strepII oligo was ligated to create the *strepII-tag* and *his6-tag* containing overexpression vector pTA1403 (Figure 3.10).

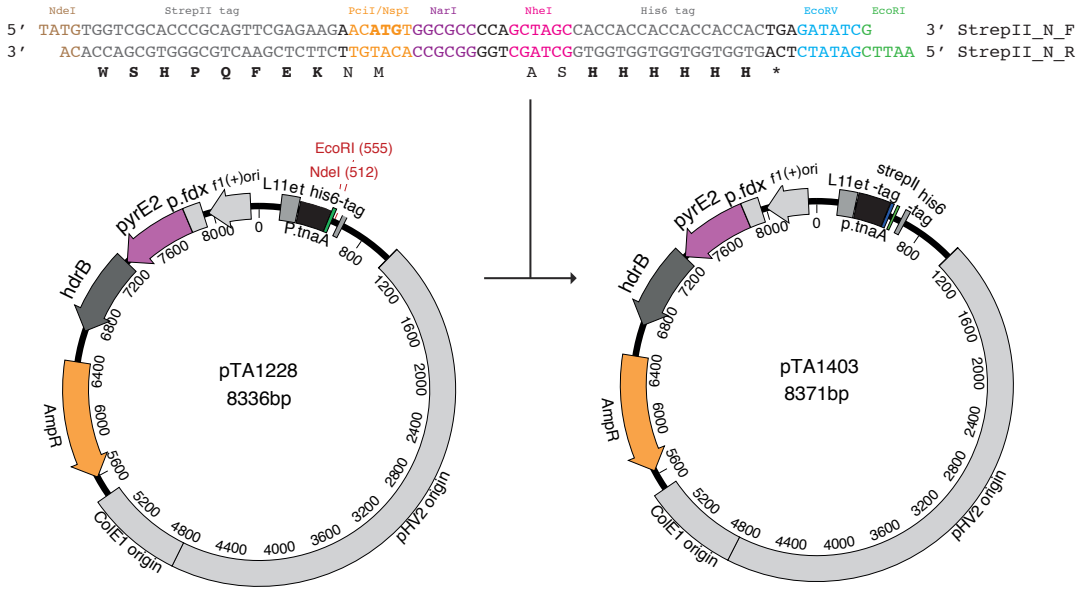


Figure 3.10: **Construction of the overexpression construct pTA1403.** Oligonucleotides StrepII_N_F and StrepII_N_R were annealed and ligated into pTA1228 at *Nde*I and *Eco*RI sites. pTA1103 is the overexpression construct that allows for N-terminally *strepII-tagged* and/or a C-terminally *his6-tagged* proteins under the control of a tryptophan inducible promoter *(p.tnaA)*

pTA1403 allows for in-frame insertion of a gene of interest with the option to have an N-terminal *strepII-tag* and/or a C-terminal *his6-tag* (Figure 3.11). To generate a construct without a N-terminal *strepII-tag* the start codon was replaced with an *Nde*I site and ligated with the *Nde*I end. For an N-terminal *strepII-tag* the start codon was replaced with an *Nco*I, *Bsp*HI, *Pci*I or an *Sph*I site and ligated with the *Pci*I or *Nsp*I end. To generate a construct without a C-terminal *his6-tag*, an *Eco*RI *or Bam*HI site was incorporated after stop codon and ligated, and for a C-terminal *his6-tag* the stop codon of the gene was replaced with an *Nhe*I or compatible site *(Avr*II, *Spe*I *or Xba*I*)* and ligated with the *Nhe*I end. In both pTA1392 and pTA1403 constructs the gene of interest is under the control of a tryptophan-inducible promoter, *p.tnaA*. Due to the

importance to this study, these plasmids will be discussed in more detail in Chapter 6: *in vitro Analysis of Hel308*.

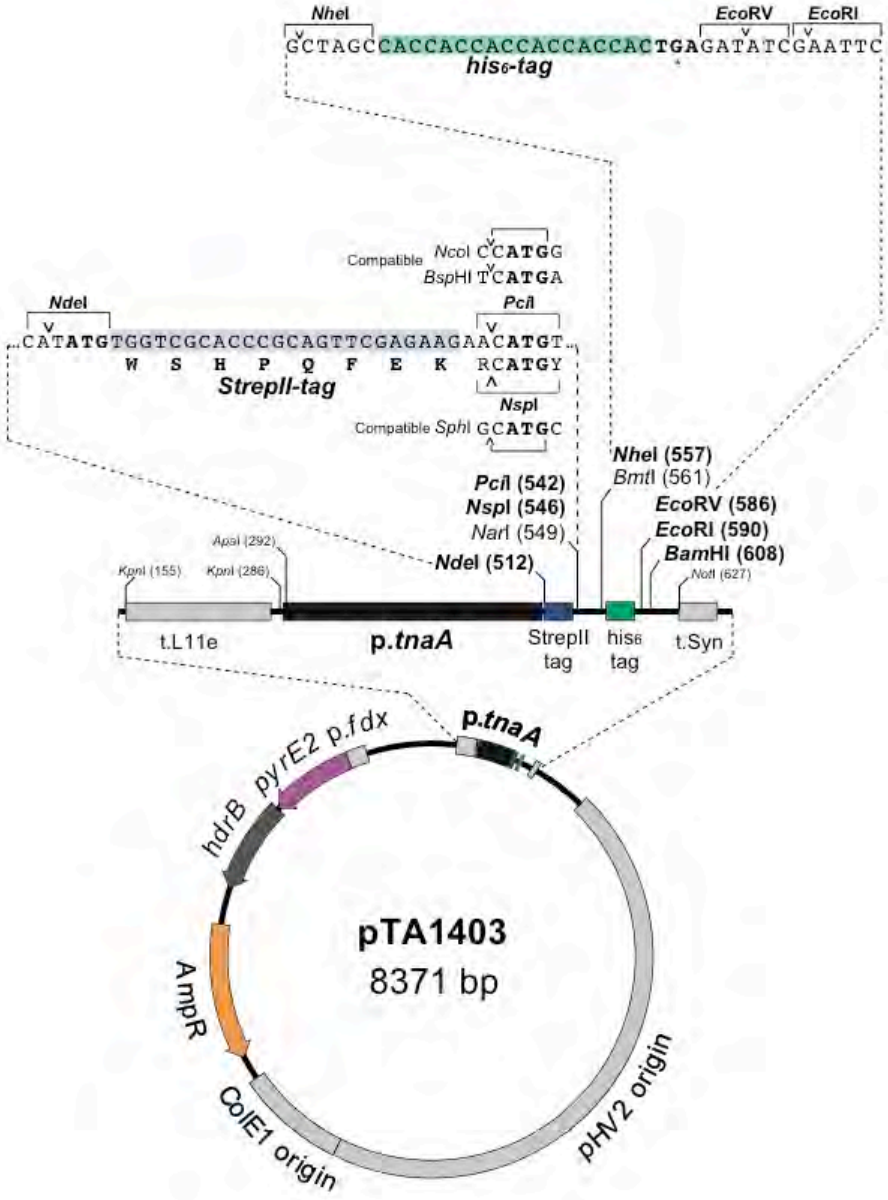


Figure 3.11: **pTA1403.** A plasmid for the conditional overexpression of N-terminally *strepII* and/or C-terminally *his6* tagged proteins. The plasmid contains a tryptophan inducible promoter (*p.tnaA*), *pyrE2* and *hdrB* markers for selection in *H. volcanii*. Restriction sites for inserting gene of interest are shown.

pTA1419, N-terminally *his6*-tagged and a C-terminally *strepII*-tagged *hel308* overexpression construct derived from pTA1392

hel308 was amplified from the genomic clone pTA415 using the primers Hel308SphF and Hel08R_NheI. The PCR product was digested with *Sph*I and

*Nhe*I and inserted into pTA1392 at *Nsp*I and *Nhe*I sites downstream of a *his6-tag* and upstream of a *strepII-tag* (Figure 3.12).

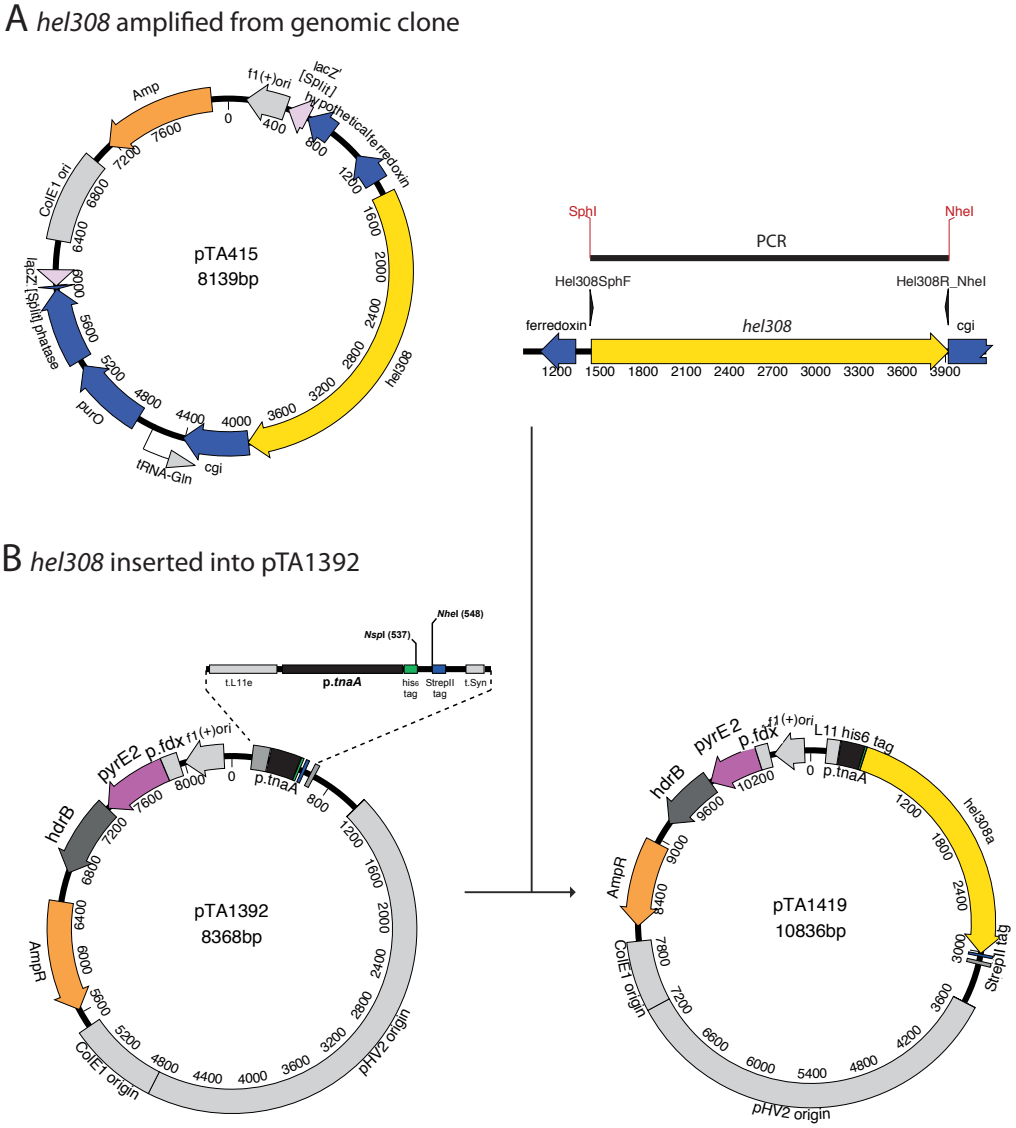


Figure 3.12: **Construction of N-terminally *his6*-tagged and a C-terminally *strepII*-tagged Hel308 overexpression construct, pTA1419. (A)** Hel308 amplified by PCR using Hel308SphF and Hel08R_NheI and digested with *SphI* and *NheI* **(B)** Ligation of PCR fragment into pTA1392 to generate pTA1419

Other N-terminally *his6-tagged* and/or C-terminally *strepII-tagged* protein overexpression constructs derived from pTA1392

All other overexpression constructs derived from pTA1392 were generated using the same principles as the ones used to make pTA1419. The gene of interest was amplified from a genomic clone with primers containing suitable

restriction sites to insert the gene into pTA1392 so that a N-terminal *his6-tag* and/or a C-terminal *strepII-tag* was in frame with the gene. Specific genomic clone plasmids, PCR primers and restriction sites used to digest the PCR fragment and pTA1392 for each over expression construct are detailed in Table 3.2.

Table 3.2: **PCR and digests to generate pTA1392 derived overexpression constructs.**

Name	Gene	Genomic clone used	PCR Primers	PCR digest	pTA1392 digest	Tags incorporated
pTA1422	hel308	pTA415	Hel308F_NdeI Hel308R_NheI	*Nde*I, *Nhe*I	*Nde*I, *Nhe*I	C-terminal *strepII tag*

pTA1425, N-terminally s*trepII*-tagged and a C-terminally *his6*-tagged *hel308* overexpression construct derived from pTA1403

hel308 was amplified from the genomic clone pTA415 using the primers Hel308SphF and Hel08R_NheI. The PCR product was digested with *Sph*I and *Nhe*I and inserted into pTA1403 at *Nsp*I and *Nhe*I sites downstream of a s*trepII-tag* and upstream of a *his6-tag* (Figure 3.13).

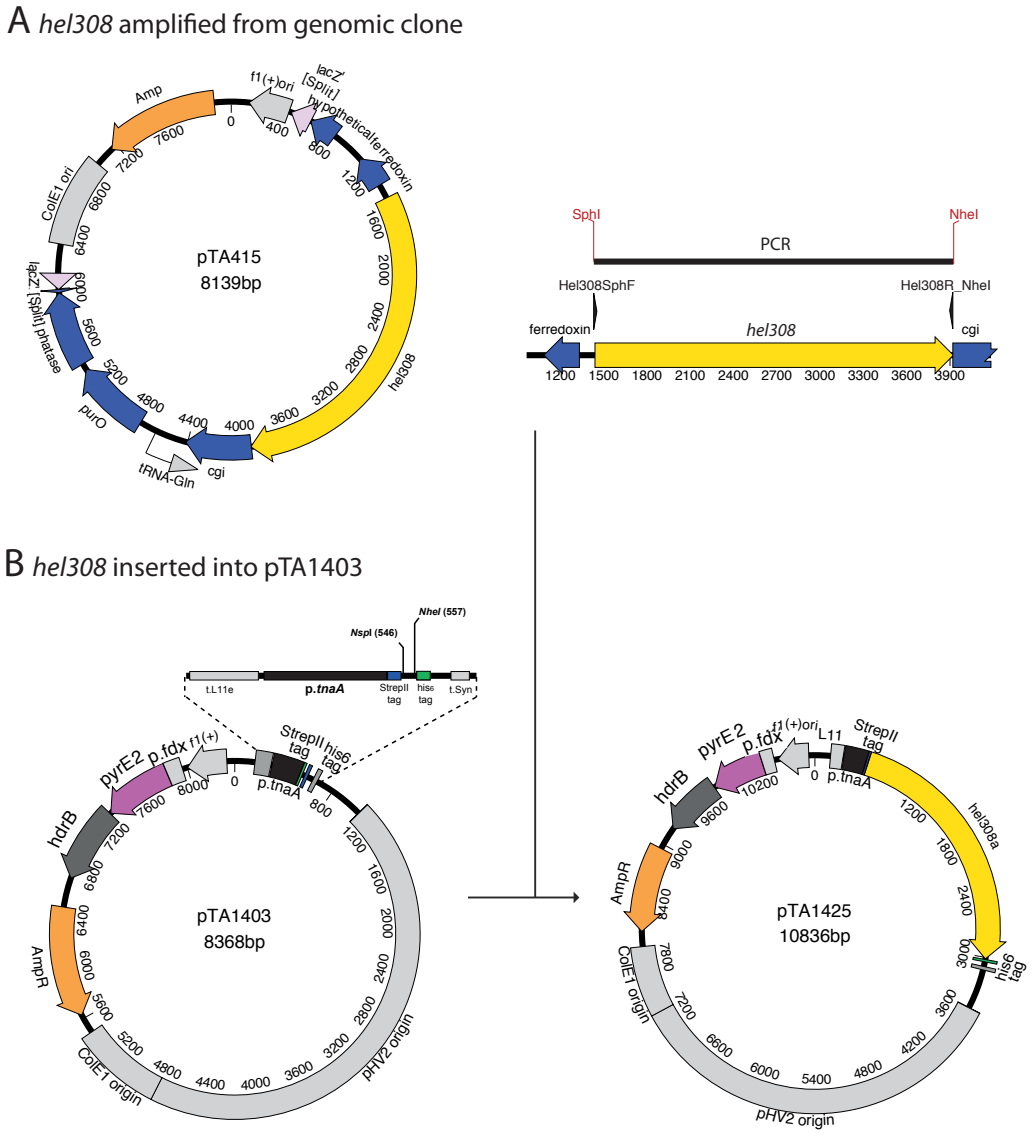


Figure 3.13: **Construction of N-terminally *strepII*-tagged and a C-terminally *his6*-tagged Hel308 overexpression construct, pTA1425. (A)** Hel308 amplified by PCR using Hel308SphF and Hel08R_NheI and digested with *Sph*I and *Nhe*I **(B)** Ligation of PCR fragment into pTA1403 to generate pTA1425.

Other N-terminally *strepII-tagged* and/or C-terminally *his6-tagged* protein overexpression constructs derived from pTA1403

All other overexpression constructs derived from pTA1403 were generated using the same principles as the ones used to make pTA1425. The gene of interest was amplified from a genomic clone with primers containing suitable restriction sites to insert the gene into pTA1403 so that a N-terminal *strepII-tag* and/or a C-terminal *his6-tag* was in frame with the gene. Specific genomic clone plasmids, PCR primers and restriction sites used to digest the PCR

fragment and pTA1403 for each over expression construct are detailed in Table 3.3.

Table 3.3: **PCR and digests to generate pTA1403 derived overexpression constructs.**

Name	Gene	Genomic clone used	PCR Primers	PCR digest	pTA1403 digest	Tags incorporated
pTA1428	hel308	pTA415	Hel308SphF helQ(R)BamHI02	*Sph*I, *Bam*HI	*Nsp*I, *Bam*HI	N-terminal s*trepII tag*

3.1.3.2 Construction of plasmids for Chapter 4: Genetic Analysis of *hel308*

pTA1669, *hel308* expression plasmid

A plasmid was constructed to allow for *in trans* complementation of *hel308* by expression from its native promoter. A 2442 bp *Sap*I to *Sca*I fragment of pTA354 containing *Hv oripHV1/4* was used to replace the equivalent fragment in pTA1316. This replication origin is low copy number and ensures that Hel308 will be expressed at native levels via the *Hv oripHV1/4* origin (Norais et al 2007). The resulting plasmid is pTA1669 (Figure 3.14).

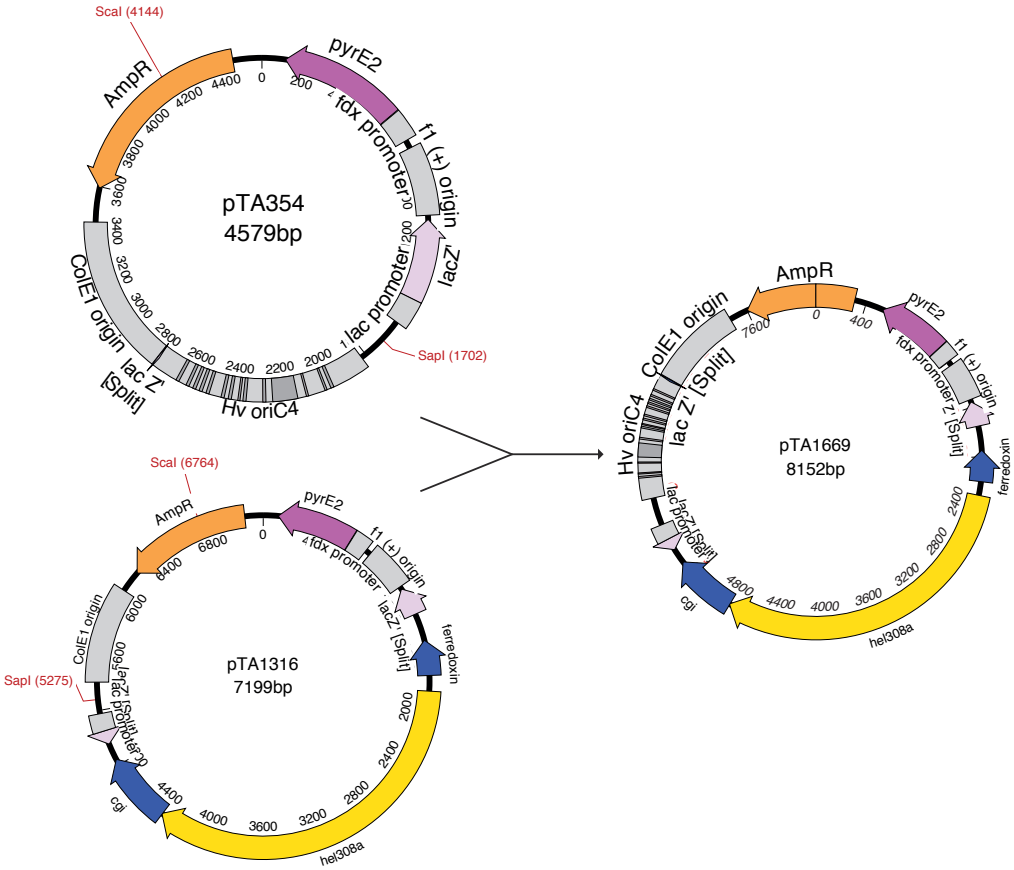


Figure 3.14: **Construction of *hel308* expression plasmid, pTA1669.** A 2442 bp *Sap*I to *Sca*I fragment of pTA354 was used to replace the equivalent fragment in pTA1316.

pTA1662, N-terminally *his6-tagged* and C-terminally *strepII-tagged hel308* gene replacement construct

This plasmid was constructed to allow for gene replacement of the native *hel308* with a N-terminally *his6-tagged* and C-terminally *strepII-tagged* version of *hel308* on the chromosome. This was so Hel308 could be protein purified following native levels of expression.

The backbone of pTA1316 was amplified using the primers Hel308_BamHI_F and hel308Nde5R, the PCR product was blunt ended and self ligated to result in a plasmid construct containing *hel308* upstream and downstream regions, pTA1661 (Figure 3.15A). An N-terminally *his6-tagged* and C-terminally *strepII-tagged hel308* 2561 bp

fragment was digested from pTA1419 using *Nde*I and *Bam*HI and ligated into pTA1661 at *Nde*I and *Bam*HI sites. This generated a construct containing an N-terminally *his6-tagged* and C-terminally *strepII-tagged hel308* with upstream and downstream flanking regions to be used for gene replacement, pTA1662 (Figure 3.15B).

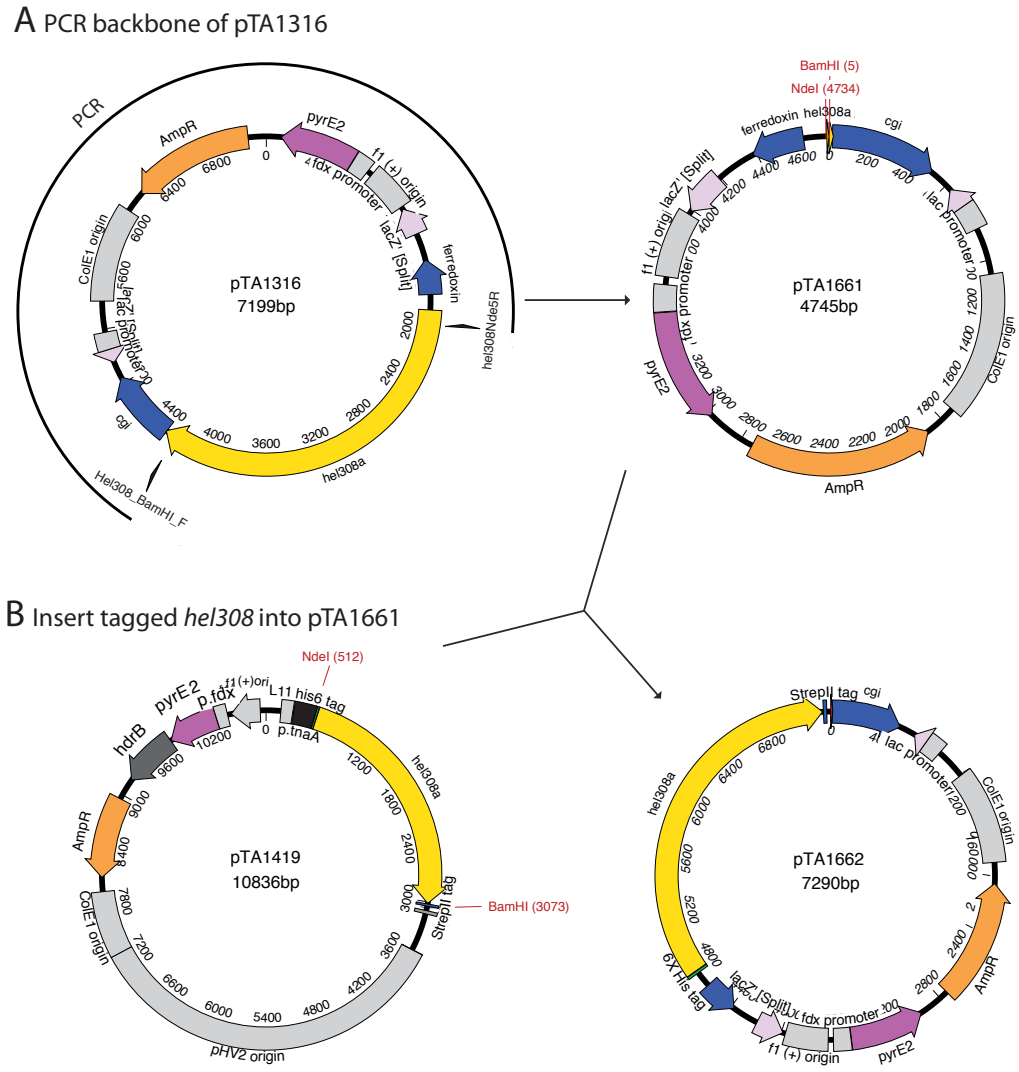


Figure 3.15: **N-terminally *his6-tagged* and C-terminally *strepII-tagged hel308* gene replacement construct, pTA1662. (A)** The backbone of pTA1316 was amplified the primers Hel308_BamHI_F and hel308Nde5R and product was blunt end self ligated to generate pTA1661. **(B)** *Nde*I to *Bam*HI fragment if pT1419 was ligated into pTA1661 to generate pTA1662

3.2 Strain Construction

Several *H. volcanii* parental strains were used in this study, each strain is used for different purposes, details of each strain is described below:

H26 (*ΔpyrE2*)

H26 is the standard wild-type laboratory strain; this strain was derived from the wild-type isolate DS2 by curing of the plasmid pHV2 and deletion of *pyrE2* (uracil biosynthesis) (Allers et al 2004, Wendoloski et al 2001). The pHV2 plasmid was lost to allow for use of its origin in episomal vectors. H26 derived strains are also used to study DNA damage sensitivity.

H53 (*ΔpyrE2, ΔtrpA*)

H53 contains the auxotrophic selection markers Δ*pyrE2* (uracil) and Δ*trpA* (tryptophan) the latter is utilised in the selection of integrated *trpA* marked gene replacement or deletion constructs for genes that are difficult to delete. This strain is derived from H26 (Allers et al 2004).

H164 (*ΔpyrE2, bgaHa-Bb, leuB-Ag1, ΔtrpA*)

H164 contains the auxotrophic selection markers Δ*pyrE2* (uracil) and Δ*trpA* (tryptophan). H164 also contains a mutant *leuB-Ag1* (leucine) allele and a mutant beta-galactosidase gene (*bgaHa-Bb*) from *Haloferax alicantei* that can also be used in recombination assays (Delmas et al 2009, Holmes & Dyall-Smith 2000, Lestini et al 2010).

H195 (*ΔpyrE2, bgaHa-Bb, leuB-Ag1, ΔtrpA, ΔhdrB*)

H195 is derived from H164 and contains the following auxotrophic selection markers: Δ*pyrE2* (uracil), Δ*hrdB* (thymidine), Δ*trpA* (tryptophan) *bgaHa-Bb* (beta-galactosidase) and *leuB-Ag1* (leucine) (Guy et al 2006). The Δ*hdrB* is utilised in the selection of integrated *hdrB* marked gene replacement or deletion constructs for genes that are difficult to delete, as an alternative to a *trpA* marker. H195 derived strains are also used to study DNA damage sensitivity.

H1424 (*ΔpyrE2, Nph-pitA, Δmrr, cdc48d-ct, ΔhdrB)*

H1424 was derived from H1209 (*ΔpyrE2, Nph-pitA, Δmrr, ΔhdrB)* for improved protein overexpression and purification (Allers et al 2010, Stroud et al 2012). *mrr* encodes a restriction enzyme that cleaves foreign DNA methylated at GATC residues, was deleted to generate H1209 to allow direct transformation of *H. volcanii* without the need to passage plasmid DNA through an *E. coli* dam mutant. This strain was developed in order to reduce the co-purification of naturally histidine rich proteins with *his6-tagged* recombinant proteins. The *pitA* gene of *H. volcanii*, which contains a histidine rich linker region, was previously replaced with a non-histidine rich orthologue from *Natronomonas* pharaonis (Allers et al 2010). This strain also contains a truncated version of the *cdc48d* gene which lacks its histidine rich C-terminus. The *mrr* gene was deleted for efficient transformation of plasmids into *H. volcanii*. Mrr is a restriction enzyme that cuts methylated 5′-GATC-3′ sequences. *E. coli* contains a Dam methylase that methylates 5′-GATC-3′ sequences and most plasmids are passaged through *E. coli* before being transformed into *H. volcanii mrr+* strains.

H1895 (*ΔpyrE2, Nph-pitA, Δmrr, cdc48d-ct, ΔhdrB, ΔpilB3C3)*

H1895 was derived from H1424 for improved protein overexpression and purification (Strillinger et al 2016). In this strain the genes Hvo_1033 (*pilC3)* and Hvo_1034 (*pilB3*) were deleted, these genes encode an ATPase and an integral membrane protein of the pilus assembly system respectively. These proteins are necessary for biofilm formation and deletion of these genes eliminates the attachment of *H. volcanii* to surfaces of any kind while leaving cells motile

H2047 (*ΔpyrE2, Nph-pitA, Δmrr, cdc48d-ct, ΔtrpA)*

H2047 is derived from H1424 via H1611 for improved protein expression and purification. To generate H1611, *trpA* was deleted from H1424 to allow for the selection of integrated trpA marked gene replacement or deletion constructs.

H1611 was then transformed with a linear fragment containing *hdrB* to make the resulting strain H2407 *hdrB*+.

3.2.1 Strains Containing Episomal Plasmids

Strains containing episomal plasmids were confirmed by selection on appropriate media, which depended on the genotype of the transformed strain and the markers present on the plasmid.

3.2.2 Gene Deletions and Replacements

Gene deletions in *H. volcanii* were achieved using the pop-in/pop-out gene knock-out system (Bitan- Banin *et al*. 2003) seen in Figure 3.16. A thorough description of this method can be found in Chapter 2: *Materials and methods*, Section 2.2.5: *Genetic Manipulation of Haloferax volcanii*.

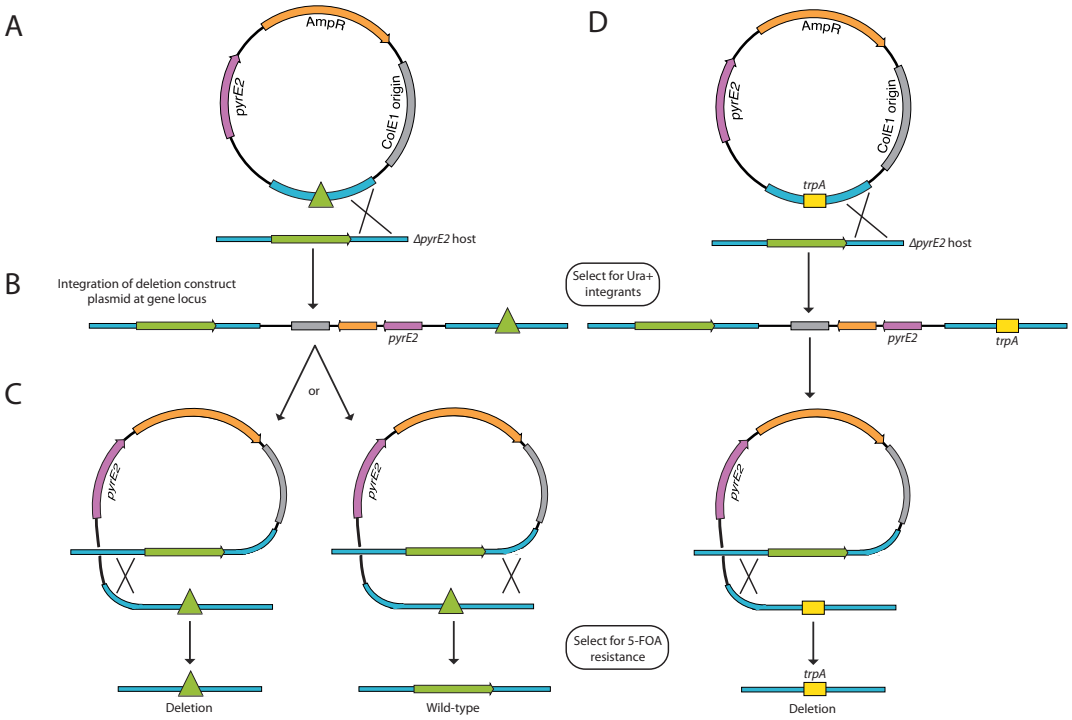


Figure 3.16: **Gene deletion by pop-in/pop-out. (A)** *ΔpyrE2* strains are transformed with a *pyrE2*+ deletion construct. **(B)** Pop-ins are selected for by their ability to grow on media lacking uracil **(C)** Relieving uracil selection allows for pop-out to occur, recombination between the regions of homology can be either upstream (left) or downstream (right). Pop-out is selected for by plating on 5-FOA The resulting gene locus will either be a deletion of wild-type **(D)** Replacing the gene with *trpA*+ marker enables direct selection for deletion mutants.

3.2.2.1 Generation of deletion strains for Chapter 4: Genetic Analysis of *hel308*

Deletion of Hvo_0014 (*hel308)*

Hvo_0014 (*hel308*) was deleted from H1804 (*ΔoriC1, ΔoriC2, ΔoriC3, Δori-pHV4-2*) and H164 using pTA1277 (*hel308::trpA+* deletion construct) to generate the strains H1953 and H2117 respectively. Deletion of *hel308* was confirmed by colony hybridisation using a 380 bp radiolabelled probe of *hel308* amplified from pTA415 using the primers Ski2F and Ski2R (Figure 3.17) Deletion of *hel308* was also confirmed by Southern blot (data not shown).

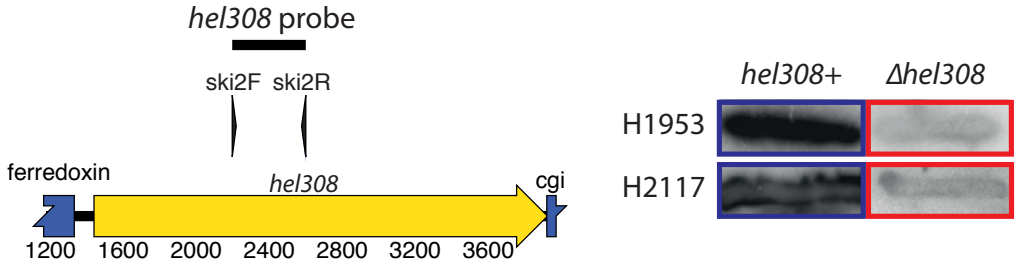


Figure 3.17: **Deletion of Hvo_0014 (*hel08*)** Candidate colonies were screened with a radiolabelled Ski2F to Ski2R PCR product of *hel308* from pTA415. Colonies that did not hybridise were *Δhel308*.

Deletion of Hvo_2383 (*radB)*

Hvo_2383 *(radB)* was deleted from H164 and H2117 and H1391 using pTA312 (*radB* deletion construct), to generate the strains H2641, H2417 and H1844 respectively. Deletion of *radB* was confirmed by colony hybridisation using a 286 bp radiolabelled probe of *radB* amplified from pTA50 using the primers EXTF and HEXTR, (Figure 3.18). Deletion of *radB* was also confirmed by Southern blot, Deletion was also confirmed by Southern blot using a 337 bp radiolabelled probe digested from pTA312 using *Kpn*I and *Not*I, Figure 3.18D shows H1844 as an example.

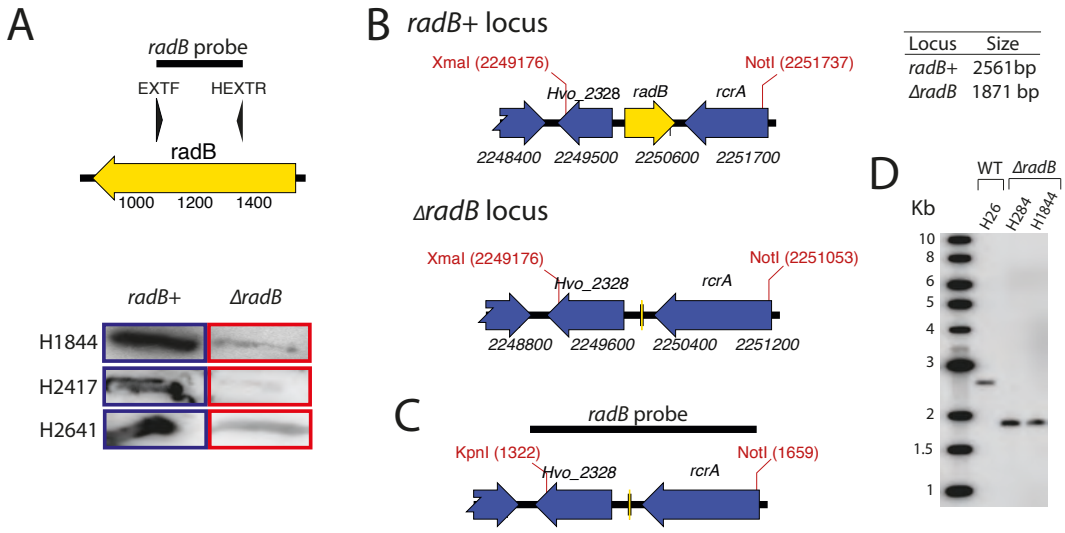


Figure 3.18: **Deletion of *radB* from H1391 (*Δhel308).* (A)** Candidate colonies were screened with a radiolabelled EXTF to HEXTR PCR product of *radB* from pTA50. Colonies that did not hybridise were *ΔradB*. **(B)** Gene locus for *radB*+ and *ΔradB***.** The *Xma*I and *Not*I sites used to digest genomic DNA are shown along with predicted locus size **(C)** A 337bp *ΔradB* probe for southern hybridization was made by digesting pTA312 with *Kpn*I and *Not*I **(D)** Southern blot, showing H1844 is deleted for *radB*. H284 is also deleted for *radB*.

3.2.2.2 Generation of deletion of strains for Chapter 7: Genetic and Phylogenetic Analysis of *hel308b*

Deletion of Hvo_0971 (*hel308b)*

Hvo_0971 (*hel308b*) was deleted from a number of backgrounds using the deletion construct pTA1368 to generate the strains H1843, H2007 and H2488. *hel308b* deletions were confirmed by colony hybridisation using a 993 bp probe amplified from genomic DNA using the primers Hel308bF and Hel308bR (Figure 3.19) Deletion of *hel308b* was also confirmed by Southern blot (data not shown).

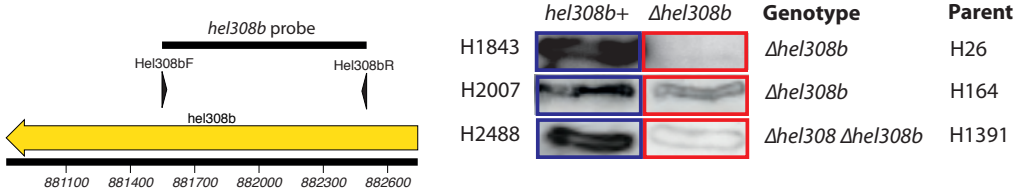


Figure 3.19: **Deletion of Hvo_0971 (*hel08b*)** Candidate colonies were screened with a radiolabelled Hel308bF to Hel308bR PCR product of *hel308b* from genomic DNA. Colonies that did not hybridise were *Δhel308b*.

H2643, *Δhel308b, Δhel308::trpA+*

hel308::trpA+ was deleted from H2007 using pTA1277 (*hel308::trpA+* deletion construct), deletion of *hel308* was confirmed by colony hybridisation using a 380 bp radiolabelled probe of *hel308* amplified from pTA415 using the primers Ski2F and Ski2R (Figure 3.20)

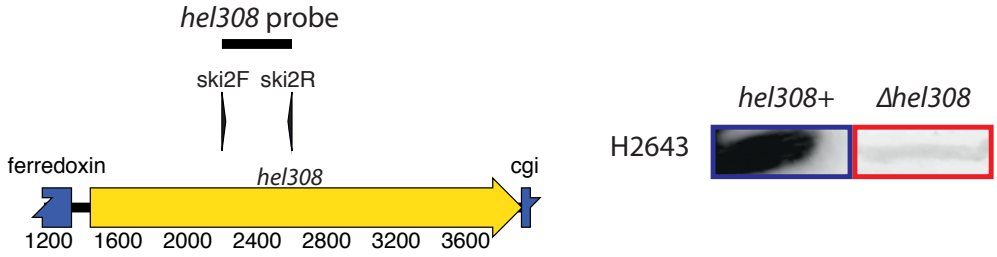


Figure 3.20: **Deletion of *hel308::trpA+* from H2007.** Candidate colonies were screened with a radiolabelled Ski2F to Ski2R PCR product of *hel308* from pTA415. Colonies that did not hybridise were *Δhel308*.

3.2.3 Gene Replacements

Gene replacements are carried out in the same manner as gene deletions, however the replacement plasmid features a gene of interest between the flanking regions of homology. Gene replacements were carried out to introduce point mutations in *hel308*.

3.2.3.1 Generation of gene replacement strains for Chapter 5: Genetic Analysis of *hel038* Point Mutants

hel308 point mutant gene replacement constructs (Table 3.4) were introduced in to H1392 and H2117 (*Δhel308* in a H26 background and *Δhel308* in a H164 background respectively). Deletion of *hel308* was confirmed by colony hybridisation using a 380 bp radiolabelled probe of *hel308* amplified from pTA415 using the primers Ski2F and Ski2R (Figure 3.21). The presence of *hel308* point mutants was also confirmed by Southern blot (data not shown).

Table 3.2: ***hel308* point mutant gene replacement constructs.**

Name	Details
pTA1335	*hel308-D154N* gene replacement construct
pTA1545	*hel308-E422G* gene replacement construct
pTA1546	*hel308-D420A* gene replacement construct
pTA1575	*hel308-H317G* gene replacement construct
pTA1576	*hel308-E330G* gene replacement construct
pTA1647	*hel308-F316A* gene replacement construct
pTA1648	*hel308-R743A* gene replacement construct

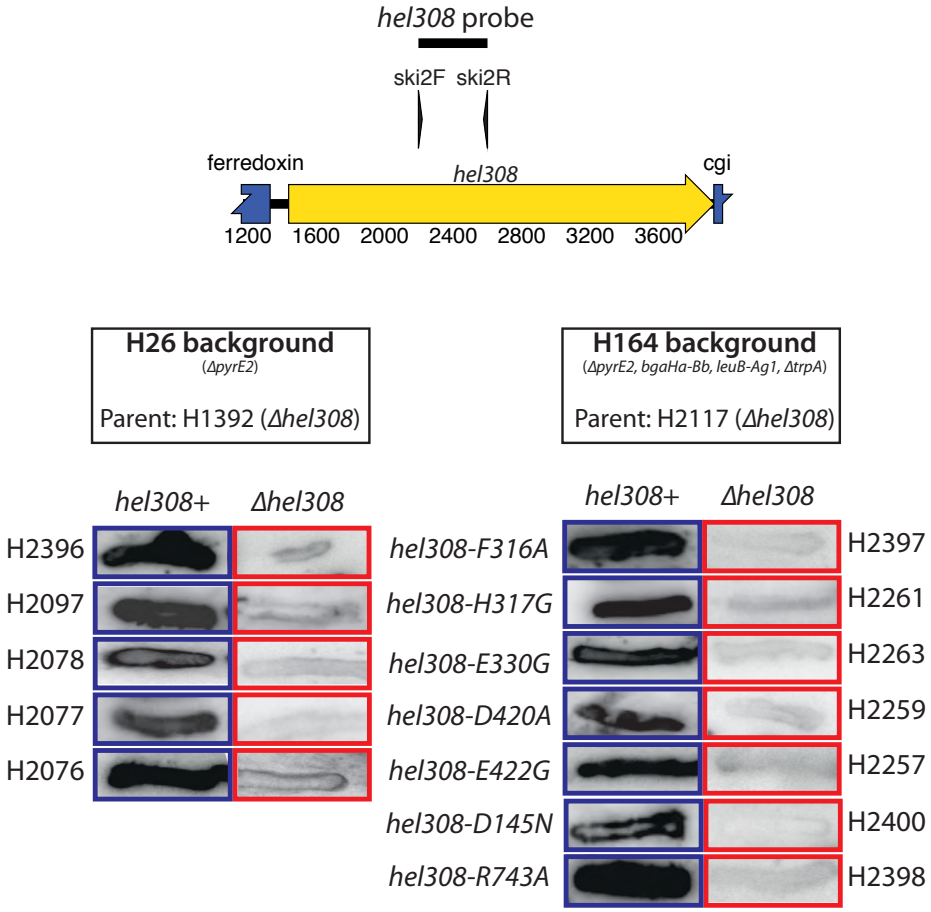


Figure 3.21: **Introduction of *hel308* point mutants into H1392 and H2117 (*Δhel308* in H26 and H164 backgrounds respectively).** Candidate colonies were screened with a radiolabelled Ski2F to Ski2R PCR product of *hel308* from pTA415. Colonies that hybridised to probe contained a *hel308* point mutant.

3.2.3.2 Generation of gene deletion and gene replacement strains for Chapter 6: *in vitro* Analysis of Hel308

To study Hel308 protein:protein interactions under conditions of native expression, a strain was generated where the chromosomal *hel308* was N-

terminally *his6-tagged* and C-terminally *strepII-tagged*. First, *hel308* was deleted from the improved protein-expression strain H2047 and then the N-terminally *his6-tagged* and C-terminally *strepII-tagged hel308* was introduced.

H2131, *Δhel308::trpA+*

hel308::trpA+ was deleted from H2047 using pTA1277 (*hel308::trpA+* deletion construct), deletion of *hel308* was confirmed by colony hybridisation using a 380 bp radiolabelled probe of *hel308* amplified from pTA415 using the primers Ski2F and Ski2R (Figure 3.22A). Deletion of *hel308* was also confirmed by Southern blot (data not shown).

H2418, *his6 tag-hel308-strepII tag*

An N-terminally *his6-tagged* and C-terminally *strepII-tagged hel308* was introduced into H2131 using the gene replacement construct pTA1662. Introduction of *his6 tag-hel308-strepII tag* was confirmed by colony hybridisation using a 380 bp radiolabelled probe of *hel308* amplified from pTA415 using the primers Ski2F and Ski2R (Figure 3.22B). Introduction of *hel308* was also confirmed by Southern blot (data not shown).

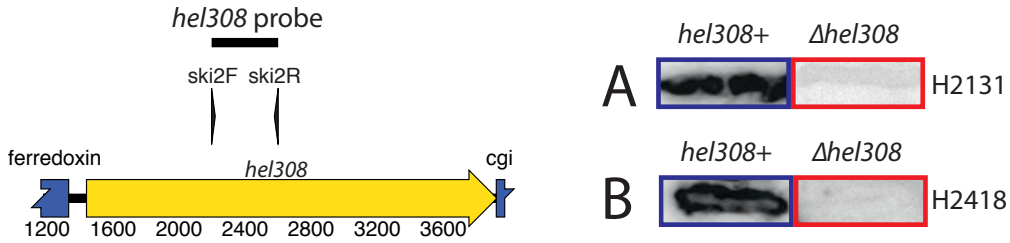


Figure 3.22: **Deletion of *hel308* from H2047 and replacement with N-terminally *his6-tagged* and C-terminally *strepII-tagged hel308*. (A)** *hel308* was deleted from the improved protein expression and purification strain H2047 to generate H2131. **(B)** An N-terminally *his6-tagged* and C-terminally *strepII-tagged hel308* was introduced to generate the strain H2418**.** Candidate colonies were screened with a radiolabelled Ski2F to Ski2R PCR product of *hel308* from pTA415. Colonies that did not hybridise were *Δhel308*.

Deletion of Hvo_0009 (*tnaA*)

Hvo_0009 (*tnaA*) was deleted from H1424 (improved protein overexpression and purification strain) and H1895 (H1424 *ΔpilB3C3*) using the *ΔtnaA::hdrB+* deletion construct pTA1615 to generate the pop-outs H2167 and H2169 respectively. *tnaA* deletions were confirmed by colony hybridisation using a 1488 bp radiolabelled probe of *tnaA* digested from pTA875 with *Age*I (Figure 3.23A). *tnaA* deletions were also confirmed by the ability of the strains to grow on media lacking Thy, PCR and by Southern blot (data not shown).

In order that *hdrB* selection can be utilised in episomal expression plasmids, H2167 and H2169 containing *ΔtnaA::hdrB+* were transformed with the *ΔtnaA* partial deletion construct, pTA1730. The *tnaAΔEcoNI* partial deletion was confirmed by colony hybridisation using a 519 bp radiolabelled probe digested from pTA875 with *Eco*NI, Figure 3.23B, Southern blot and by the inability ability of colonies to grow on media lacking Thy (data not shown).

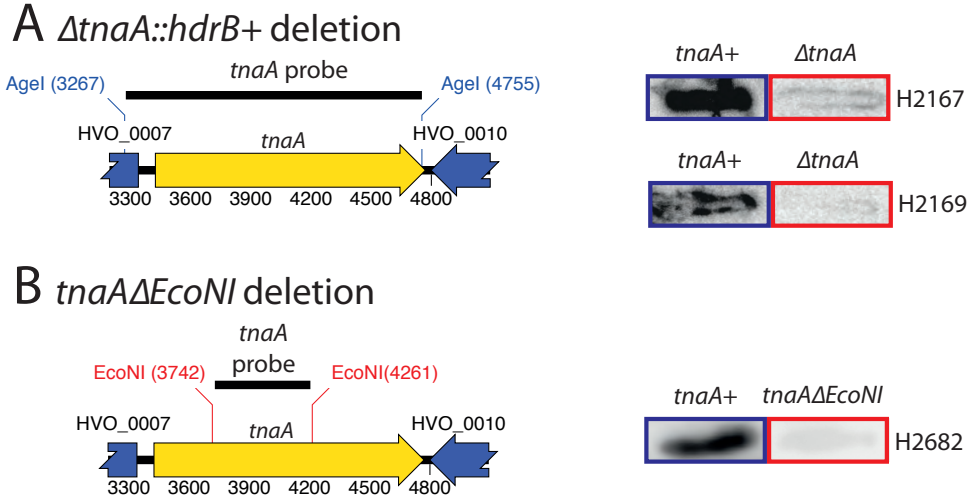


Figure 3.23: **Deletion of Hvo_0009 (*tnaA*). (A)** Candidate colonies were screened with a radiolabelled *Age*I digested fragment from pTA875. Colonies that did not hybridise were *ΔtnaA*. **(B)** Candidate colonies were screened with a radiolabelled *Age*I digested fragment from pTA875. Colonies that did not hybridise were *ΔtnaA*.

Chapter 4: Genetic Analysis of *hel308*

4.1 Background

Helicases are ubiquitous enzymes that play a fundamental role in nearly all DNA and RNA metabolic processes, including replication, recombination, DNA repair, transcription, translation and RNA splicing. Hel308 is a monomeric helicase that is conserved across metazoans and archaea but is absent in bacteria and fungi. Hel308 is implicated in replication fork restart and homologous recombination, however the exact cellular function and mechanism by which Hel308 acts is unknown.

4.1.1 Hel308 and Replication Forks

Replication forks can stall after collision with a lesion on the DNA and can be repaired by homologous recombination-dependent or independent pathways. Several studies have shown Hel308 to unwind replication fork structures with the ability to displace both leading and lagging strand structures with 5′ and 3′ overhangs, respectively (Fujii et al 2002, Guy & Bolt 2005). Hel308 has also been seen to localise to replication forks and interact with replication fork associated proteins such as RPA and PCNA (Fujikane et al 2006, Tafel et al 2011, Woodman et al 2011). It has been proposed that Hel308 may act to restart stalled replication forks but the exact role that Hel308 plays at replication forks is unknown.

4.1.2 Hel308 and Homologous Recombination

Hel308 was shown using genetic and biochemical techniques to interact with proteins that act in both early and late stages of homologous recombination such as the Rad51 recombinase and the Holliday junction resolvase Hjc respectively.

Early stages of homologous recombination

Recombinases are essential for homologous recombination and are found in all domains of life. Recombinases are named RecA in bacteria, Rad51 in eukaryotes and RadA in archaea. Recombinases form nucleoprotein filaments on ssDNA, which search for homologous dsDNA molecules. Once found, recombinases catalyse strand invasion of the dsDNA by the ssDNA, and D-loop formation (McEntee et al 1979).

The *C. elegans* Rad51 paralogue shows synthetic lethality when deleted in combination with the Hel308 homologue, HelQ (Taylor et al 2015). HelQ was also seen to disassemble Rad51 filaments from DNA *in vitro* (McClendon et al 2016, Ward et al 2010). Additionally, human Hel308 foci were seen to co-localise with Rad51 foci during immunoprecipitation studies, suggesting an interaction between the two proteins (Tafel et al 2011). Furthermore, the human Hel308 homologue HELQ was seen to interact with paralogues of the RAD51 recombinase, RAD51B/C/D and XRCC2, which are collectively known as the BCDX2 complex. The BCDX2 complex is required for homologous recombination (Adelman et al 2013, Takata et al 2013). Hel308 is clearly involved in the early stages of homologous recombination via interactions with recombinases, however the exact role that Hel308 plays is unknown.

Late stages of homologous recombination

Recombination intermediates such as Holliday junctions are formed during homologous recombination and require resolving, this can occur through several pathways. For a detailed description see Chapter 1: *Introduction*, Section 1.4.3.5. Resolution of Holliday junctions usually involves the action of endonucleases known as resolvases that are able to cleave branched DNA structures. Archaea contain a Holliday junction resolvase named Hjc, protein pull-down and yeast two-hybrid assays have demonstrated that the Hel308 homologue from *S. tokodaii* interacts with Hjc *in vitro* (Li et al 2008). The Hel308 homologue Hjm was also seen to prevent the formation of Hjc-

Holliday junction complexes (Hong et al 2012). This suggests that Hjm could regulate Hjc, however neither the mechanism of this interaction nor its function is known.

4.1.3 Hel308 in *Haloferax volcanii*

In *Haloferax volcanii*, deletion mutants of *hel308* are viable but slow growing, and like *hel308* mutants of other organisms are sensitive to DNA interstrand crosslinking agents such as MMC. *H. volcanii* also contains a second Hel308 helicase named Hel308b, which will be analysed in *Chapter 7: Genetic and Phylogenetic Characterisation of Hel308b*. *H. volcanii* is an ideal organism for the genetic analysis of Hel308 and its role in homologous recombination for several reasons. Firstly, *H. volcanii* is an easily-culturable archaeon and contains many genetic tools such as routine transformation, selectable markers, a well-established gene deletion/replacement system and a defined recombination assay. Secondly, several aspects of homologous recombination have already been studied in *H. volcanii*. For example, the regulation of recombination by the recombinase and mediator RadA and RadB, respectively, as well as the role of Hjc and Hef in resolution of homologous recombination, have been described (Haldenby 2007, Lestini et al 2010, Wardell 2013).

4.2 Aims

Genetic analysis of *hel308*, in combination with other recombination factors such as *radA, radB, hef* and *hjc*, will give insight into the role that Hel308 plays in homologous recombination and DNA repair. The aims of this chapter are to:

- Analyse the expression profile of *hel308* under native and DNA-damaging conditions.
- Generate strains deleted for *hel308* in combination with genes encoding homologous recombination and DNA repair proteins
- Analyse the generation time of strains deleted for *hel308*.

- Analyse the phenotypes of *hel308* deletion strains after treatment with DNA damaging agents such as UV and MMC, which cause double strand breaks and interstrand crosslinks, respectively.
- Analyse the recombination frequency and levels of crossover and non-crossover products formed by strains deleted for *hel308*.

4.3 Results

4.3.1 Analysis of *hel308* transcript levels

Expression of genes involved in DNA repair can be constitutive, or up-regulated in response to DNA-damaging agents. Expression of *hel308* in response to DNA damaging agents has not previously been studied. Therefore, expression of *hel308* was measured following treatment with UV and mitomycin C (MMC); these agents induce ss/dsDNA breaks and inter-strand crosslinking respectively. Homologous recombination and NER are involved in repairing DNA breaks and crosslinks. However, if the DNA damage is not repaired, replication forks can stall at these lesions. The stalled replication forks require restarting and this is achieved by homologous recombination.

Strains were grown to mid-exponential phase and either UV irradiated at 20 J/m^2 or incubated for 1 hour with 2 μg/ml MMC. Samples with no UV irradiation or MMC treatment were used as a control. This level of UV irradiation and MMC exposure results in a small amount of cell death. Following treatment, cells were resuspended in fresh growth media and allowed to recover. RNA was extracted and DNase treated to remove any contaminating DNA. End-point RT-PCR (reverse transcriptase PCR) was used to measure transcript levels of *hel308* using the primers Ski2F and Hel308RTR which gives a product size of 301 bp. As a control, expression of *rpoA* (RNA polymerase subunit A) was analysed, as transcript levels of this gene do not change depending on the growth phase of the cell (Tom Batstone, University of Nottingham, personal communication). *rpoA* expression levels were analysed

using the primers rpoARTR and rpoARTF (product size 328 bp). Results are shown in Figure 4.1.

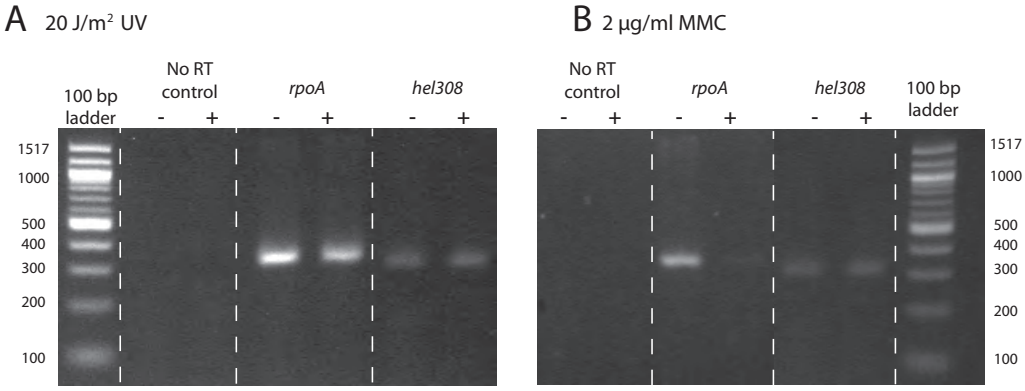


Figure 4.1: **RT-PCRs showing expression levels of *hel308* and *rpoA* (control). (A)** Expression following treatment with 20 J/m^2 UV-irradiation. *hel308* expression does not change. **(B)** Expression following treatment with 2 µg/ml MMC. *hel308* expression does not change. In all cases, '-' indicates no UV or MMC treatment, and '+' indicates treatment with UV or MMC.

This preliminary data shows that there is no difference in the expression of *hel308* following treatment with 20 J/m^2 UV-irradiation. This is not surprising, given that *hel308* does not have a growth phenotype when treated with UV-irradiation (Guy & Bolt 2005).

This data also shows that hel308 is also constitutively expressed after treatment of MMC (2 µg/ml). Further investigation may be required since these high doses may have a detrimental effect on the cells.

rpoA expression was used as a control for RT-PCRs as transcript levels of this gene do not change depending on the growth phase of the cell (Tom Batstone, University of Nottingham, personal communication). Results show that this control is not suitable for treatment with MMC, as the expression level of *rpoA* is drastically reduced following treatment. This was not the case after UV irradiation.

4.3.2 Genetic Interactions

In order to investigate in which pathway of DNA repair Hel308 acts, *hel308* was deleted in combination with genes known to be involved in DNA repair.

For details of plasmid and strain construction, see Chapter 3: *Plasmid and Strain Construction*.

4.3.2.1 Deletion in Combination with *radA*

Homologous recombination involves the exchange of nucleotide sequence between identical or near-identical DNA sequences. RadA is the archaeal recombinase that catalyses strand exchange during homologous recombination. With the aid of RadB, RadA forms a nucleoprotein filament on ssDNA that can then bind to dsDNA molecules and search for a region of homology. Once homology is found, RadA catalyses strand invasion and D-loop formation (Kil et al 2000), Figure 4.2.

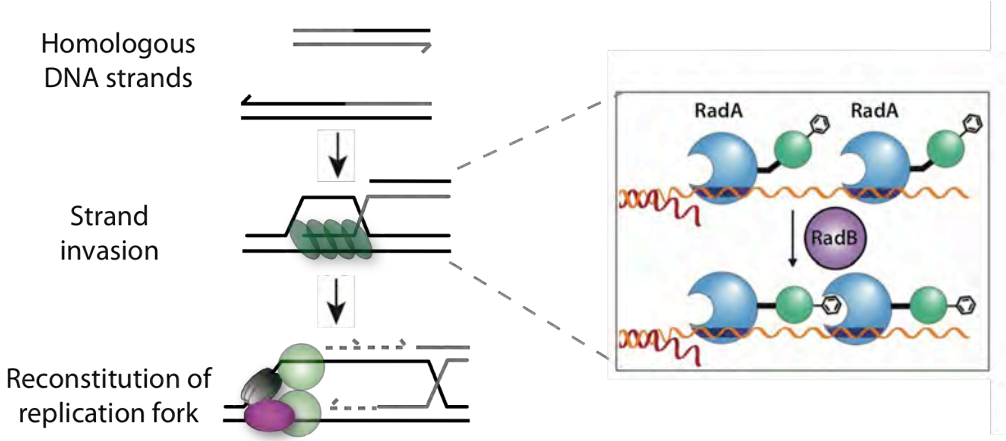


Figure 4.2: **Homologous recombination.** A brief overview of the strand invasion step of homologous recombination using a stalled replication fork as an example. With assistance from RadB, the recombinase RadA forms a nucleoprotein filament and catalyses strand invasion to generate a formation.

In two immunoprecipitation studies, the human Hel308 homologue HelQ was found to interact with RAD51C in human embryonic kidney 293T cells and with RAD51B, RAD51C, RAD51D and XRCC2 in HeLa S3 cells. (Adelman et al 2013, Takata et al 2013). Rad51 is the eukaryotic recombinase involved in strand exchange during homologous recombination. In vertebrates, a tetrameric complex involving the RAD51-paralogues RAD51B, RAD51C, RAD51D and XRCC2 (BCDX) functions as a recombination mediator. The BCDX2 complex has been implicated in loading Rad51 onto ssDNA (Sigurdsson et al 2001). A

dimer of RAD51B and RAD51C has also been shown to promote Rad51 strand exchange activity *in vitro* (Lio et al 2003, Sigurdsson et al 2001).Therefore it was of interest to analyse the interactions between Hel308 and the recombinase RadA to determine whether Hel308 is involved in homologous recombination in *H. volcanii*.

To study the relationship between *hel308* and homologous recombination, attempts were made to delete *hel308* in combination with *radA*. However, recombination is essential for the pop-in and pop-out steps of the gene deletion process, and RadA is the recombinase that catalyses the process of recombination (Woods & Dyall-Smith 1997). This poses a problem when attempting to delete *radA* from the genome. Δ*radA* strains of *H. volcanii* have a severe growth defect, and are unable to carry out recombination (Woods & Dyall-Smith 1997). Therefore, for a *radA* deletion to occur the chromosomal *radA* deletion must be complemented with an episomal copy of the *radA* gene (pTA637) during the pop-out step, this plasmid is later cured (Delmas et al 2009), Figure 4.3. Described further in Chapter 2: *Materials and Methods*, Section 2.2.5 *Genetic Manipulation of Haloferax volcanii, Deletion of RadA*.

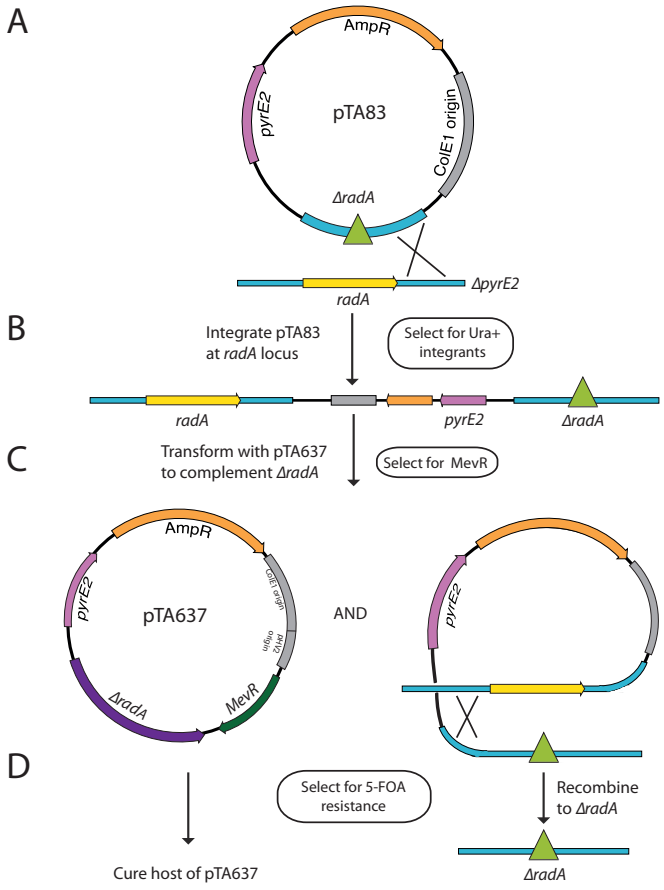


Figure 4.3: ***radA* gene deletion. (A)** A *ΔpyrE2* strain is transformed with pTA83, a deletion construct containing a *radA* deletion and *pyrE2* marker (pop-in). **(B)** The strain is then transformed with pTA637, an episomal plasmid containing *radA* and *pyrE2* and *MevR* markers. **(C)** Pop-out is able to occur due to the presence of *radA*+ on pTA637. **(D)** Strains are selected for both the pop-out of pTA83 and the loss of pTA637 by plating out on 5-FOA (Delmas et al 2009).

The *radA* deletion construct pTA83 was transformed into the H1391 (*Δhel308)* to generate the pop-in strain H1956, this strain was then transformed with the episomal *radA*+ plasmid pTA637 to generate the strain H2010. Pop-out was allowed to occur due to the presence of *radA*+ on pTA637 and then strains were selected for both the pop-out of pTA83 and the loss of pTA637 by plating out on 5-FOA. Colonies were patched and screened by colony hybridization using a *radA* radiolabelled probe to check for *radA* deletions.

Out of 320 colonies screened, no *radA* deletions in a H1391 (*Δhel308)* background were found.

4.3.2.2 Deletion in Combination with *radB*

RadB is a paralogue of the archaeal recombinase RadA, RadB is only found in euryarchaea and has been proposed to function as a recombination mediator (Haldenby 2007, Haldenby et al 2009, Wardell 2013). *ΔradB* strains display growth defects and sensitivity to DNA damaging agents, however these phenotypes are less severe than those of *ΔradA* strains (Guy et al 2006, Haldenby 2007). Furthermore, *ΔradB* strains recombine at approximately 5% of wild type unlike *ΔradA* strains, which are unable to carry out recombination.

Since *radA* could not be deleted in a *Δhel308* background, the strain H1844 (*Δhel308, ΔradB)* was generated to study the relationship between *hel308* and homologous recombination.

Growth Rate

In order to visually compare the growth rates, H1844 (*Δhel308 ΔradB*), H1845 (*hel308+ ΔradB*), and H1391 (*Δhel308 radB+*), were streaked onto complete media alongside wild-type H26 (*hel308+ radB+*), Figure 4.4.

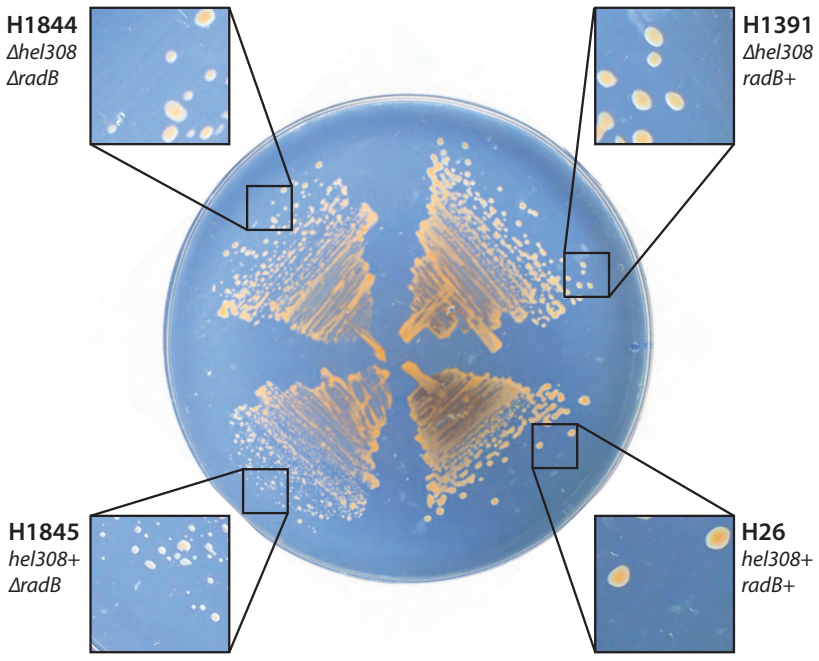


Figure 4.4: **Growth of strains deleted for *hel308*, *radB* or both.** H1391 (*Δhel308 radB+*) and H1845 (*hel308+ΔradB*) have a slight and severe growth defect respectively compared to wild type (*hel308+ radB+*). The double mutant H1844 (*Δhel308 ΔradB*) has a growth rate faster than that of H1845 (*hel308+ΔradB*).

H1845 (*hel308*+ Δ*radB*) has a severe growth defect compared to wild type (*hel308*+ *radB*+). The double mutant H1844 (Δ*hel308* Δ*radB*) has bigger colonies compared to H1845 (*hel308*+ Δ*radB*), these colonies are closer in size to H1391 (Δ*hel308 radB*+). The double deletion does not show synthetic lethality, in fact this assay shows the oposite, suggesting an antagonistic relationship between Hel308 and RadB.

A limitation of using growth on solid media is that only large variations can be readily observed (for example the difference between *hel308*+Δ*radB* and *hel308*+*radB*+ strains). Strains with similar growth rates cannot be distinguished by this qualitative method, and for this reason growth in liquid culture was measured. Strains were grown over two consecutive overnights to form a vigorously growing culture. Cell growth during exponential phase was measured every 15 minutes by the A_{600} of the culture over a time course of 48 hours using an Epoch2 Microplate Spectrophotometer (BioTek). A growth curve was plotted and generation time calculated during exponential phase for each strain, Figure 4.5.

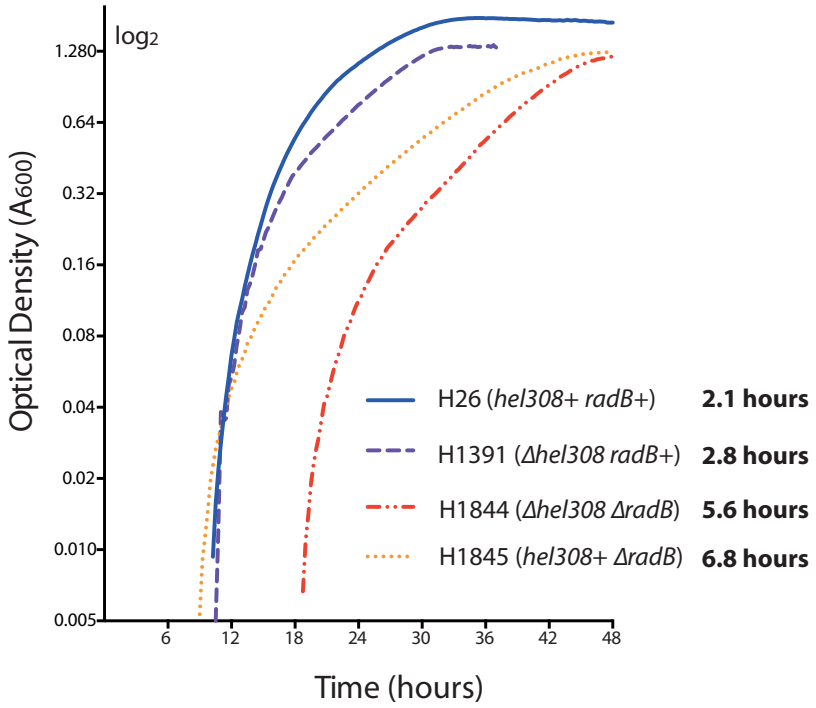


Figure 4.5: **Exponential growth rate of strains deleted for *hel308* and/or *radB*.** Growth was measured by A_{600}. Generation time is indicated at the side of the strain name. H1844 (Δ*hel308* Δ*radB*) has a faster generation time than H1845 (*hel308*+ Δ*radB*) but not as fast as H1391 (Δ*hel308 radB*+). Graph plotted on a $\log_2$ scale. 3 repeats were carried out for each strain. Generation time calculated from linear growth phase. All strains and repeats were incubated on the same 48 well plate and measured simultaneously using an Epoch 2 Microplate Spectrophotometer (BioTek).

H1844 (Δ*hel308* Δ*radB*) has a faster growth rate than H1845 (*hel308*+ Δ*radB*) with generation times of 5.6 hours and 6.8 hours respectively, however this is not as fast as H1391 (Δ*hel308 radB*+) which has a generation time of 2.8 hours. This shows that a double deletion of Δ*hel308* Δ*radB* improves the growth compared to a single Δ*radB* deletion. This suggests that Hel308 could be acting as an anti-recombinase and supressing homologous recombination. In H1844 (Δ*hel308* Δ*radB*), upon the deletion of *hel308* the inhibitory effect is relieved allowing for faster growth.

It was noticed that generation times of strains varied between experiments when measuring the growth using the Epoch 2 Microplate Spectrophotometer (BioTek). Therefore, in this study comparisons are ony made between sets of strains within the same experiment. i.e. strains that have been incubated on the same 48 well microtitre plate and with the A_{600} measured simultaneously during a single run on the the Epoch 2 Microplate Spectrophotometer (BioTek). Since generation times vary between experiments, the generation times stated are not absolute. However, the relationship between sets of strains has been observed to be consistent, meaning that comparisons of generation times between strains within a single experiment is reliable. Therefore, in this study growth curves generated by this method are used to illustrate the differences in generation times between a set of strains rather than an exact determination of each generation time.

Complementation of *hel308*

H26 (*hel308*+) and H1391 (Δ*hel308*) colonies are indistinguishable from each other on plates, Figure 4.4. Therefore, complementation of *hel308* was measured using liquid growth assays, the method for this assay has been described previously, Figure 4.6. The episomal plasmid for *in trans* expression of *hel308* pTA1669 (*hel308*+ *pyrE2*+) was reintroduced into Δ*hel308* strain generating H2572, H2573 containing the episomal empty vector plasmid pTA354 (*pyrE2*+) was used as a control.

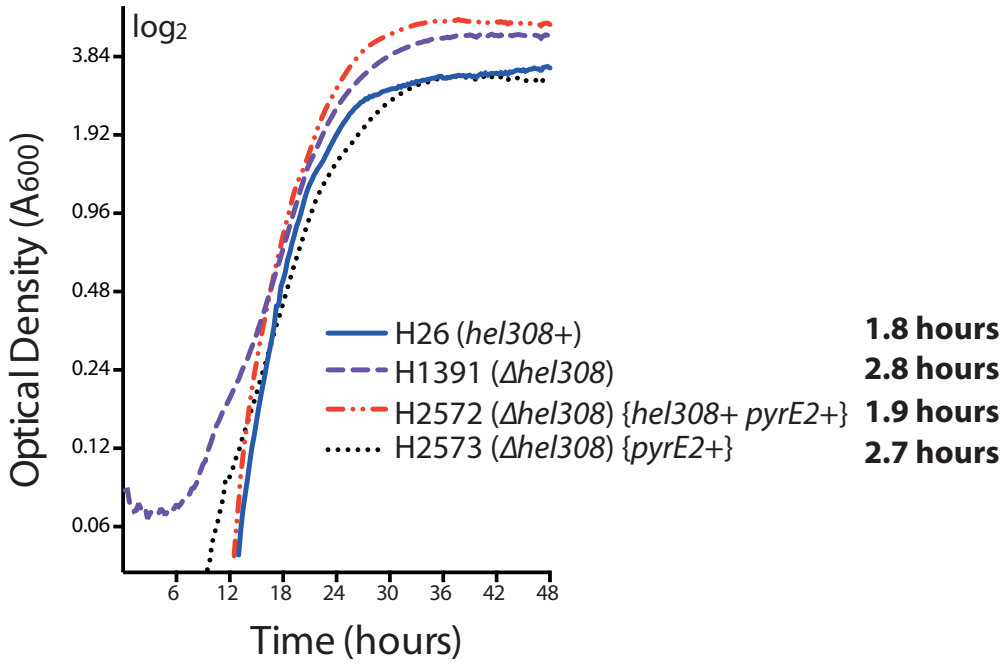


Figure 4.6: **Exponential growth rate of strains complemented with *hel308*.** Growth was measured by A_{600}. Generation time is indicated at the side of the strain name. H2572 (Δ*hel308)* with *hel308* episomally expressed (*hel308+ pyrE2+*) has a generation time comparable to that of wild type H26 (*hel308+*). H2573 (Δ*hel308)* with episomal empty vector (*pyrE2+*) has a generation time comparable to that of H1391 (Δ*hel308)*. Graph plotted on a $\log_2$ scale. 3 repeats were carried out for each strain. Generation time calculated from linear growth phase. All strains and repeats were incubated on the same 48 well plate and measured simultaneously using an Epoch 2 Microplate Spectrophotometer (BioTek).

H2572 (Δ*hel308)* with *hel308* episomally expressed (*hel308+ pyrE2+*) has a generation time comparable to that of wild type H26 (*hel308+*). H2573 (Δ*hel308)* with episomal empty vector (*pyrE2+*) has a generation time comparable to that of H1391 (Δ*hel308)*. This indicates that the deletion of *hel308* is resposible for the slow growing phenotype seen in H1391 (Δ*hel308)*.

To determine if *hel308* is truly responsible for the improved growth phenotype in the H1844 (Δ*hel308* Δ*radB*) strain compared with H1845 (*hel308+* Δ*radB)* (Figures 4.4 and 4.5), *hel308* was reintroduced on an episomal plasmid into H1844.

In order to visually compare the growth rates, H2426 (Δ*hel308* Δ*radB*) containing the episomal plasmid for *in trans* expression of *hel308* pTA1669 (*hel308+ pyrE2+*) was streaked onto complete media along side the control

strain H2427 (*Δhel308 ΔradB*) containing the episomal empty vector plasmid pTA354 (*pyrE2+*), Figure 4.7.

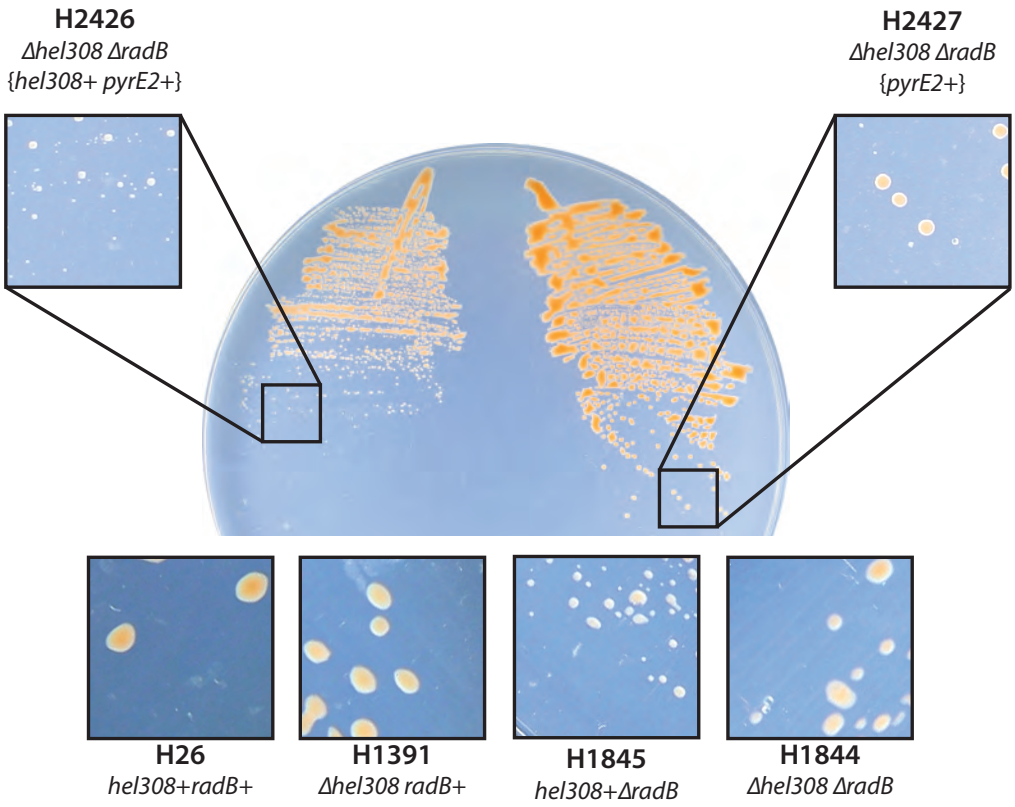


Figure 4.7: **Growth of *Δhel308 ΔradB* strains complemented with an episomally expressed *hel308* or an empty vector control.** H2426 (*Δhel308 ΔradB*) with episomally expressed *hel308* has smaller colonies than H2427 (*Δhel308 ΔradB*) with the empty vector control. H26, H1391, H1845, H1844 included for reference.

H2426 (*Δhel308 ΔradB* with episomally expressed *hel308*) has significantly smaller colonies than the control strain H2427 (*Δhel308 ΔradB* with empty vector). H2426 colonies resemble those of H1845 (*hel308+ ΔradB*), showing that the improvement in growth phenotype seen in H1844 (*Δhel308 ΔradB)* has been reversed upon the addition of *hel308*. This indicates that the deletion of *hel308* is resposible for the improved growth of *ΔradBΔhel308* strains. Additionally, as expected, H2427 colonies resemble H1844 (*Δhel308 ΔradB)*.

Survival Following Treatment with DNA-damaging Agents

UV light induces several types of DNA lesions which consequently result in single and double strand DNA breaks (Goosen & Moolenaar 2008). The most

common forms of DNA damage induced by UV irradiation are cyclobutane pyrimidine dimers (CPDs) and (6-4) photoproducts. These lesions distort the backbone of DNA which can lead to blockage of replication forks, resulting in single and double strand DNA breaks. Mechanisms that repair UV lesions include photo-reactivation (apart from in placental mammals), base excision repair (in some organisms), nucleotide excision repair (NER) and homologous recombination. Strains defective these repair pathways are sensitive to UV-irradiation.

Mitomycin C (MMC) is a chemical mutagen that causes inter and intra-strand DNA crosslinks (Tomasz et al 1987). MMC alkylates guanine bases which leads to monoadduct formation between the DNA molecule and the MMC molecule. An alkylation reaction with a second guanine leads to the formation of a bisadduct, resulting in crosslinking of DNA. Crosslinking across the two DNA strands (interstrand crosslinking) occurs more frequently than crosslinking between residues in the same DNA strand (intrastrand crosslink).

Strains deleted for *HelQ* have an increased sentisitvity to DNA crosslinking agents such as MMC (Adelman et al 2013, Takata et al 2013) and strains deleted for *radB* have an increased sensitivity to DNA damaging agents due to recombination defects (Haldenby 2007). In order to study the ability of a Δ*hel308*Δ*radB* strain to survive following treatment with DNA-damaging agents, cells were treated with UV-radiation or MMC.

To analyse survival of H1844 (Δ*hel308* Δ*radB*), H1845 (*hel308*+ Δ*radB*), H1391 (Δ*hel308 radB*+), and H26 (*hel308*+ *radB*+) following treatment with DNA damaging agents, cultures were grown to mid exponential phase, spotted onto complete media and treated with UV or on to complete media containing MMC. Plates were incubated for 4-7 days at 45℃ and colonies counted, Figure 4.8.

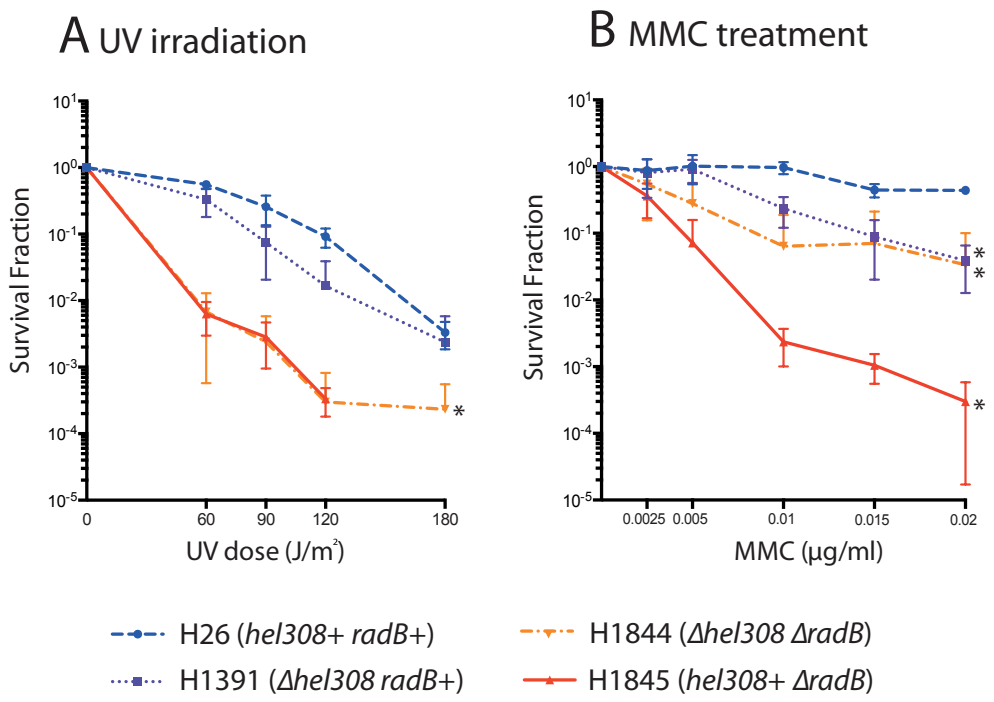


Figure 4.8: **Survival frequency of strains deleted for *hel308* and/or *radB* following treatment with DNA-damaging agents. (A)** Survival following treatment with UV irradiation. H1844 (Δ*hel308* Δ*radB*) and H1845 (*hel308*+ Δ*radB*) are highly sensitive to UV irradiation compared to wild type H26 (*hel308*+ *radB*+). H1391 (Δ*hel308 radB*+) is not sensitive to UV irradiation. **(B)** Survival following treatment with MMC. All strains are more sensitive to MMC than wild type H26 (*hel308*+ *radB*+). H1845 (*hel308*+ Δ*radB*) is highly sensitive to MMC, whereas H1844 (Δ*hel308* Δ*radB*) is only as sensitive as H1391 (Δ*hel308 radB*+). Survival fraction is calculated relative to untreated control. Each data point is generated as an average of at least 3 independent trials. Standard error is shown. Asterisk (*) indicates that the highest dose of UV (180 J/m^2) or MMC (0.02 μg/ml) is significantly different to H26 (wild type) with $P < 0.05$. P-value calculated from two-tailed *t*-test in mutated strains compared to H26 (wild-type).

Following UV irradiation both of the *radB* deleted strains H1844 (Δ*hel308* Δ*radB*) and H1845 (*hel308*+ Δ*radB*) show severe growth phenotypes; with P-values calculated from a two-tailed *t*-test being 0.0175 and 0.0125 respectively, when compared to the wild type strain H26. H1845 (*hel308*+ Δ*radB*) shows no growth after treatment with UV at 180 J/m^2. Whereas H1391 (Δ*hel308 radB*+) shows survival which is not significantly different to wild type (P-value of 0.6299). The lack of synthetic lethality demonstrates that *hel308* does not operate in the same pathway as *radB* when single and double strand DNA breaks are induced. Following treatment with MMC, H1845 (*hel308*+ Δ*radB*) shows a significant reduction in survival and H1391 (Δ*hel308 radB*+) shows a mild reduction in survival compared to wild type. P-values calculated from a two-tailed *t*-test for these strains are 0.0316 and 0.0338 respectively, when

compared to the wild type strain H26. Interestingly the double deletion H1844 (Δ*hel308* Δ*radB*) exhibits an improved survival fraction compared to the single *radB* deletion, and H1845 (*hel308*+ Δ*radB*) appears to show survival with no significant difference to H1391 (Δ*hel308 radB*+) (P-value of these two strains compared by a two-tailed *t*-test is 0.9089) suggesting a complex relationship between Hel308 and RadB. Hel308 could be acting as an antirecombinase and supressing homologous recombination. In H1844 (Δ*hel308* Δ*radB*), upon the deletion of *hel308* the inhibitory effect is relieved allowing for better survival. In this strain *radA* is still present and recombination is still occuring at low levels, allowing the cell to repair DNA crosslinks (at a reduced rate).

Recombination Frequency

Strains deleted for *radB* have a recombination defect of approximately 5% of the wild-type level (Haldenby 2007). In order to examine the effect of a *hel308* deletion in a Δ*radB* strain, homologous recombination between the chromosome and a closed circular plasmid was measured. A schematic for this assay is shown in Figure 4.9.

The recombination frequency between a plasmid and the chromosome can be measured by using a pair of mutant *leuB* alleles (Lestini et al 2010). *leuB* encodes for 3-isopropylmalate dehydrogenase, an enzyme required for leucine biosynthesis. Strains derived from a H164 background have a mutant *leuB* allele, *leuB-Ag1* that contains a point mutant at the 5' end of the gene, and these strains cannot grow on media lacking leucine. These strains are transformed with the non-replicative plasmid pTA163, which contains a *leuB-Aa2* allele that has a point mutant at the 3' of the gene. Recombination between these two *leuB* alleles results in a wild-type *leuB* allele. Since these strains are *leuB*+, they can be selected for by plating transformants on media lacking leucine. Transformants were also plated onto non-selective media to determine the viable cell count. Transformation frequency was calculated as the number of transformants on selective media relative to the viable cell count.

To determine the number of crossover and non-crossover events, colonies were patched in duplicate on Hv-Min+Trp and Hv-Min+Trp+Ura. Crossover

recombination events will generate a strain containing both wild-type *leuB* and *leuB-Aa2-Ag1*. The *pyrE2* marker also integrates onto the chromosome, resulting in a strain that is *leu+ ura+*. A non-crossover recombination event will generate a strain containing wild-type *leuB*, but the *pyrE2* marker is lost. The resulting strain would be *leu+ ura-*.

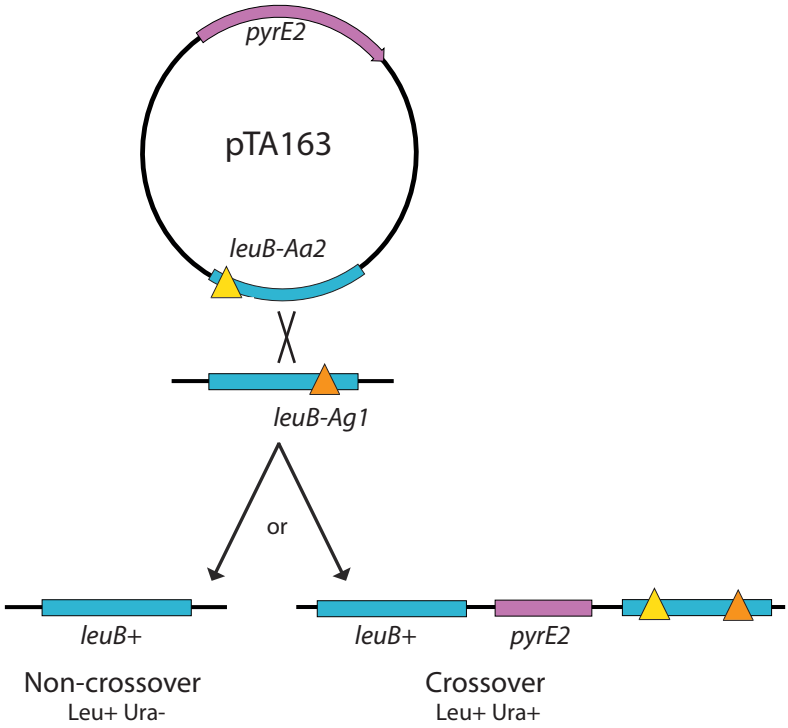


Figure 4.9: **Chromosome x plasmid recombination assays.** Δ*pyrE2* strains with a chromosomal *leuB-Ag1* allele (*leu-*) are transformed with pTA163, containing *pyrE2* and *leuB-Aa2*. A recombination event between the plasmid *leuB-Aa2* allele and chromosomal *leuB-Ag1* allele generates a wild-type *leu+* allele and strains can grow on media lacking leucine. Crossover and non-crossover events are measured by studying the proportion of transformants that have retained or lost the *pyrE2* marker found on pTA163: crossover recombinants are *pyrE2+* (*ura+*), and non-crossover recombinants Δ*pyrE2* (*ura-*).

The recombination frequency for H164 (*hel308+ radB+*), H2117 (Δ*hel308 radB+*), H2417 (Δ*hel308* Δ*radB*) and H2461 (*hel308+* Δ*radB*) was compared, transformants were screened for crossover or non-crossover recombination events, Table 4.1.

Table 4.1: **Recombination frequencies of strains deleted for *hel308* and/or *radB*.**

Strain	H164	H2117	H2461*	H2417
	hel308+ radB+	*Δhel308 radB+*	*hel308+ ΔradB*	*Δhel308 ΔradB*
Recombination Frequency (RF)	4.94×10^{-5} ($+/- 3.01\times10^{-5}$)	3.23×10^{-5} ($+/- 1.17\times10^{-5}$)	2.97×10^{-6} ($+/- 9.93\times10^{-7}$)	9.38×10^{-6} ($+/- 4.71\times10^{-6}$)
Transformation Efficiency (TE)	1.07×10^{-5} ($+/- 3.25\times10^{-6}$)	3.00×10^{-5} ($+/- 0.00$)	1.27×10^{-4} ($+/-2.30\times10^{-5}$)	2.44×10^{-4} ($+/- 1.57\times10^{-5}$)
Relative recombination frequency (normalised by TE)	4.62×10^{0}	1.08×10^{0}	2.23×10^{-2}	3.85×10^{-2}
	1×	**0.23×**	**0.048×**	**0.083×**
Crossover fraction	13.49% (126)	8.75% (120)	4.45 % (22)	4.35% (23)
Non-crossover fraction	86.51% (126)	91.25% (120)	95.45% (22)	95.65% (23)

*Recombination assay for this strain performed by Charlie Wickham-Smith (Doctoral Training Program rotation student) under my supervision.

Values in bold indicate the amount of recombination, crossover or non-crossover events compared to wild-type H164 (*hel308+ radB+*). Values are generated as an average of at least 3 independent trials, +/- standard error is shown in brackets. Cells are shaded blue to indicate recombination defect and red to indicate hyper-recombination. Fraction of crossover and non-crossover events represented as a percentage, cells are shaded where values differ significantly from the wild type (P =0.05), blue indicates a decrease, red indicates an increase. Number of colonies assayed for crossover and non-crossover is indicated in brackets underneeth the percentages.

The recombination frequency of H2117 (*Δhel308 radB+*) is lower than that of wild type H164 (*hel308+ radB+*) (0.23×). This suggests that Hel308 may play a role in homologous recombination. As expected, H2461 (*hel308+ ΔradB*) has an extremely low recombination frequency of 0.0048× that of wild type. However, upon the deletion of *hel308* in H2417 (*Δhel308 ΔradB*) the recombination frequency increases slightly to 0.083× that of wild type. The crossover and non-crossover fractions in H2461 (*hel308+ ΔradB*) and H2417 (*Δhel308 ΔradB*) are not significantly different to wild type (with two degrees of freedom with a chi-squared test). However, only a small number of colonies were available to calculate these fractions (22 and 23 respectively for H2461 and H2417), therefore a larger samping size of above 60 colonies is needed to confirm if these results are acurate.

DNA Content and Cell Size

In order to determine the effect of *hel308* and *radB* deletions on DNA content and cell size, flow cytometry was used, Figure 4.10.

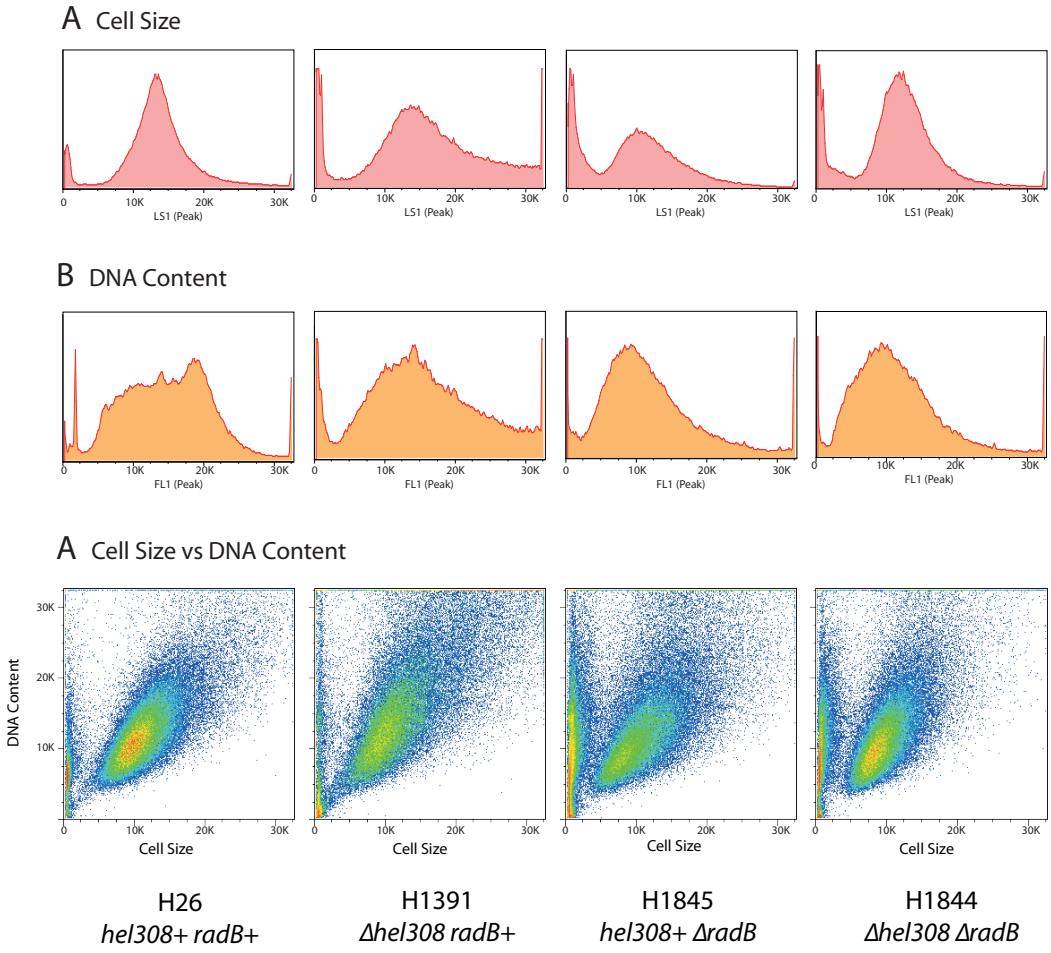


Figure 4.10: **Flow cytometry analysis of strains deleted for *hel308* and/or *radB*. (A)** Determination of cell size, H1844 (Δ*hel308* Δ*radB*) is more similar to wild type than either H1391 or H1845. **(B)** Determination of DNA content, H1844 (Δ*hel308* Δ*radB*) closely resembles H1845 (*hel308*+ Δ*radB*). **(C)** Cell size vs DNA content.

H1391 (Δ*hel308 radB*+) has a broader range of cell sizes and higher DNA content compared to wild type H26 (*hel308*+ *radB*+). H1845 (*hel308*+ Δ*radB*) has smaller sized cells and a smaller DNA content compared to wild type. Cell size of H1844 (Δ*hel308* Δ*radB*) is more similar to wild type than either H1391 or H1845, however DNA content closely resembles H1845 (*hel308*+ Δ*radB*).

4.3.2.3 Deletion in Combination with *hjc* and *hef*

Hjc and Hef are endonucleases that have been proposed to restart stalled replication forks in *H. volcanii* (Lestini et al 2010). Hjc is a Holliday junction resolvase that binds specifically to Holliday junctions and cleaves two opposing strands symmetrically to generate two recombinant duplexes. Hef comprises two distinct domains: an N-terminal domain of the DEAH helicase family and a C-terminal domain of the XPF endonuclease family. Hef acts on nicked, flapped or forked DNA and can convert a Holliday junction into a forked structure by introduction of an incision near the branch point (Komori et al 2002, Komori et al 2004). In *H. volcanii* deletion of *hef* results in only a moderate sensitivity to DNA crosslinking agents, whereas deletion of *hjc* has no effect. However, a Δ*hjc* Δ*hef* double mutant is synthetically lethal. Deletion of *hef* in a Δ*radA* background is highly deleterious but deletion of *hjc* has no effect. This suggests that Hjc acts exclusively in homologous recombination whereas Hef can act in a pathway that avoids the use of homologous recombination (Lestini et al 2010), Figure 4.11.

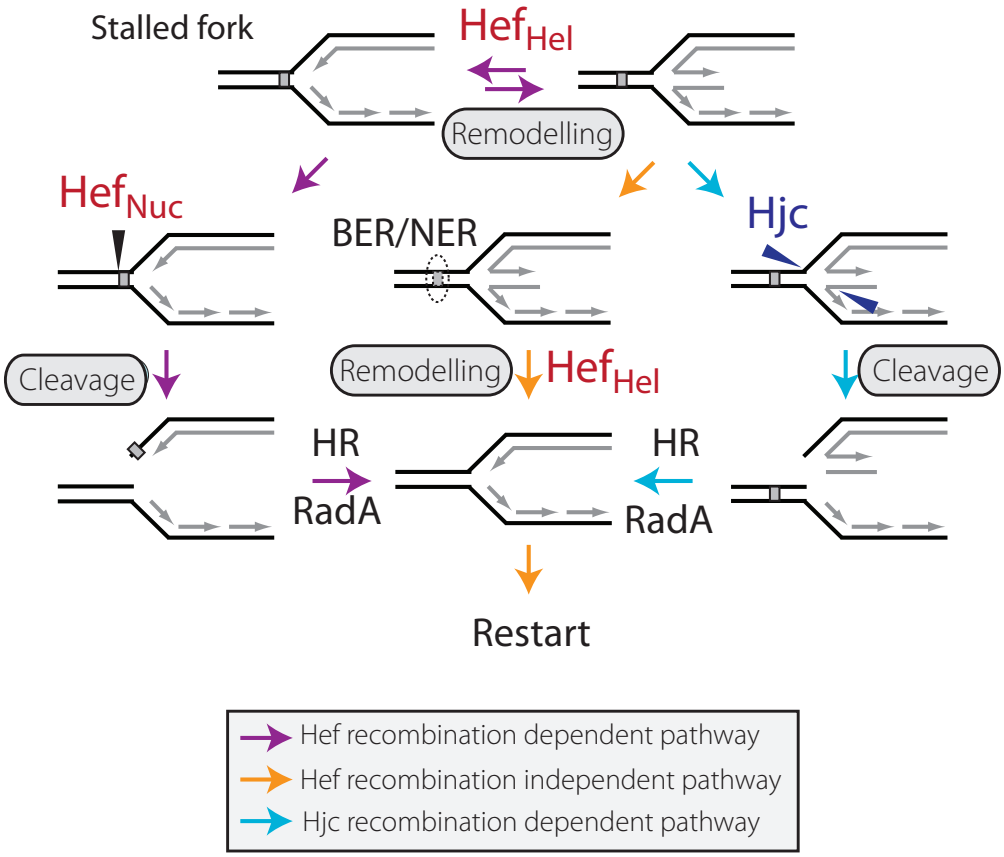


Figure 4.11: **Hef and Hjc are proposed to act in alternative pathways of replication fork restart.** Hef is proposed to act in a recombination dependent pathway involving cleavage of the replication fork (purple arrows) and a recombination independent pathway involving fork remodelling (orange arrows). Hjc is proposed to play a role exclusively in a recombination dependent pathway of replication fork restart (blue arrows). Hef_{Hel} indicates Hef helicase activity, Hef_{Nuc} indicates Hef nuclease activity. Figure adapted from (Lestini et al 2010).

The Hel308/Hjm homologue from the hyperthermophilic archaeon *Sulfolobus tokodaii* was found to physically interact with Hjc via gel filtration, affinity pull down, and yeast two-hybrid analyses (Li et al 2008). Furthermore, the unwinding activity of the Hel308/Hjm homologue was inhibited by Hjc *in vitro*. The polarity of Hel308 unwinding in *H. volcanii* is 3' to 5', however the Hel308/Hjm homologue in *S. tokodaii* can unwind DNA in both 3' to 5' and 5' to 3' directions. These results may suggest that the Hel308/Hjm family helicases, in association with Hjc endonucleases, are involved in processing of stalled replication forks. Therefore, it was of interest to analyse the genetic interactions between *hjc*, *hef* and *hel308* to determine whether Hel308 acts in the Hef or Hjc pathway in *H. volcanii*.

Growth Rate

In order to compare growth rate strains were streaked onto solid media alongside wild-type H26 (*hel308*+ *hjc*+ *hef*+). H1468 (Δ*hel308* Δ*hef*), was also streaked alongside H358 (*hel308*+ Δ*hef*), and H1391 (Δ*hel308 hef*+). H1467 (Δ*hel308* Δ*hjc*) was also streaked alongside H1049 (*hel308*+ Δ*hjc*) and H1391 (Δ*hel308 hjc*+), Figure 4.12.

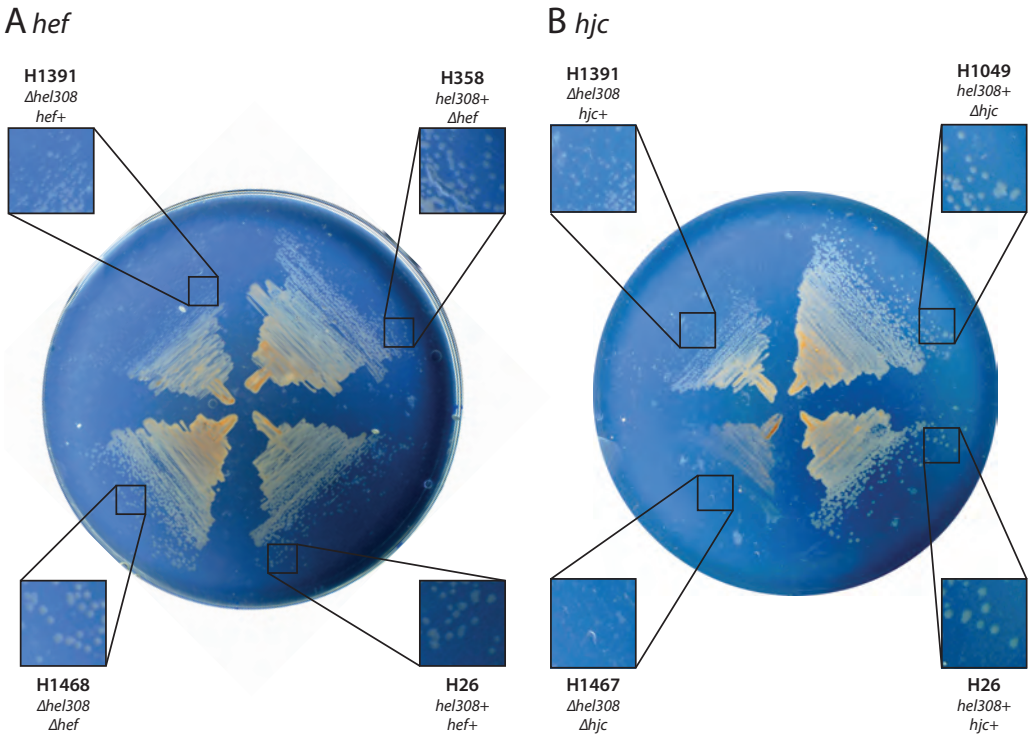


Figure 4.12: **Growth of strains deleted for *hel308* and *hef* or *hjc*. (A)** Growth rate for *hef* strains. No difference in single colony size can be seen between either strain deleted for *hef*, H358 (*hel308*+ Δ*hef*) and H1468 (Δ*hel308* Δ*hef*) and both strains have growth comparable to wild type. **(B)** H1467 (Δ*hel308* Δ*hjc*) colonies are significantly smaller than H1049 (*hel308*+ Δ*hjc*) or H1391 (Δ*hel308 hjc*+).

No difference in growth can be seen between H1468 (Δ*hel308* Δ*hef*) and H358 (*hel308*+ Δ*hef*), H1391 (Δ*hel308 hef*+) has slightly smaller colonies than either of these two strains. H1467 (Δ*hel308* Δ*hjc*) has a severe growth defect; colonies are significantly smaller than either H1049 (*hel308*+ Δ*hjc*) or H1391 (Δ*hel308 hjc*+) single colonies. This suggests a synthetic defect in *hel308 hjc* mutants.

The growth defect seen in H1467 (Δ*hel308* Δ*hjc*) was quantified; furthermore, strains with similar growth rates such as Δ*hef* cannot be distinguished by this qualitative method. For these reasons growth assays in liquid culture were performed for strains deleted for *hel308* and *hef* or *hjc*, Figure 4.13. The method for this assay has been described previously.

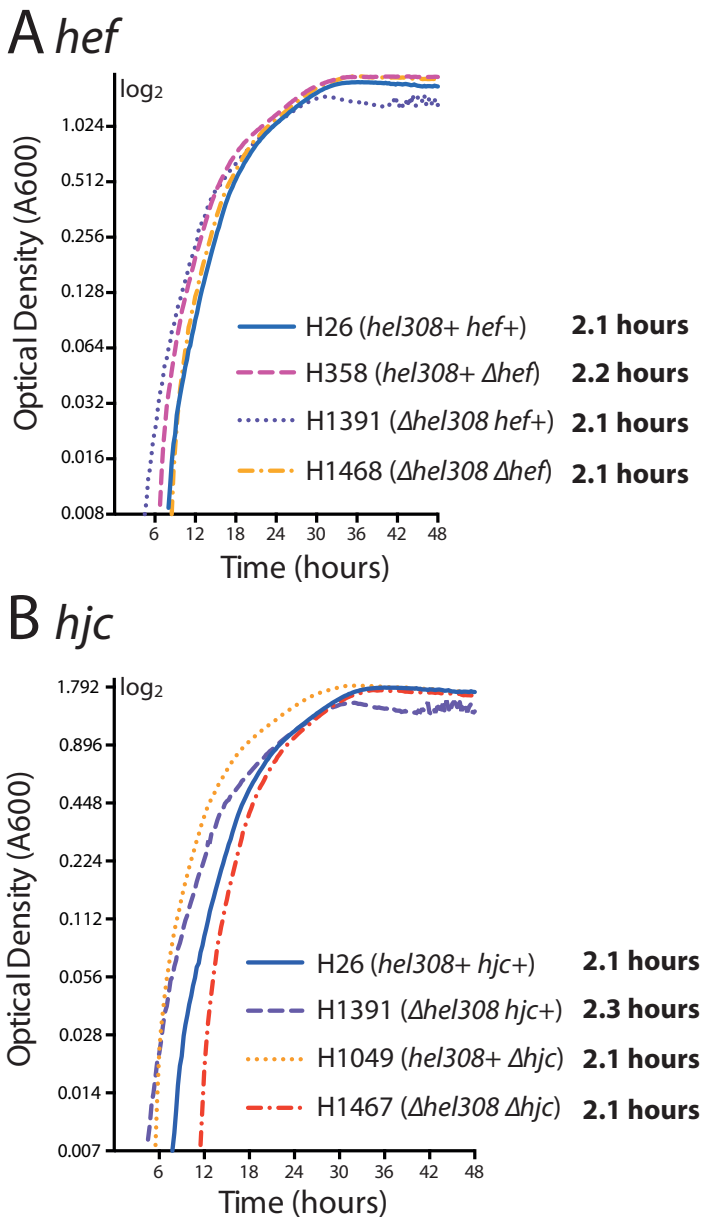


Figure 4.13: **Exponential growth rate of strains deleted for *hel308* and *hef* or *hjc*.** Growth was measured by A_{600}. Generation time is indicated at the side of the strain name. **(A)** H1391 (Δ*hel308 hef*+) shows a slight growth defect, however H358 (*hel308*+ Δ*hef*), H1468 (Δ*hel308* Δ*hef*) show no difference in growth compared to wild type H26 (*hel308*+ *hef*+). **(B)** H1391 (Δ*hel308 hjc*+) shows a slight growth defect, however H1049 (*hel308*+ Δ*hjc*) and H1467 (Δ*hel308* Δ*hjc*) show no difference in growth compared to wild type H26 (*hel308*+ *hjc*+). Graph plotted on a $\log_2$ scale. 3 repeats were carried out for each strain. Generation time calculated from linear growth phase. All strains and repeats were incubated on the same 48 well plate and measured simultaneously using an Epoch 2 Microplate Spectrophotometer (BioTek).

H1468 (Δ*hel308* Δ*hef*) has a generation time identical to H358 (*hel308*+ Δ*hef*) and H26 (*hel308*+ *hef*+) of 2.1 hours. However, an improved generation time is seen when H1391 (Δ*hel308 hef*+) is compared to H1468 (Δ*hel308* Δ *hef*) of 2.2 hours to 2.1 hours respectively. These observations indicate that there is no synthetic lethality between *hel308* and *hef*.

In contrast with growth on solid media, the growth assay in liquid culture shows that H1467 (Δ*hel308* Δ*hjc*) has a generation time identical to H1049 (*hel308*+ Δ*hjc*) and H26 (*hel308*+ *hjc*+) of 2.1 hours. However, an improved generation time is seen when H1391 (Δ*hel308 hjc*+) is compared to H1467 (Δ*hel308* Δ*hjc*) of 2.3 hours to 2.1 hours respectively. These observations indicate that there is no synthetic lethality between *hel308* and *hjc*.

Survival Following Treatment with DNA-damaging Agents

Strains deleted *hjc* or *hef* are not sensitive to UV-irradiation but are slightly sensitive to MMC (Lestini et al 2010). In order to test whether there was a synthetic defect between *hel308* and *hef* or *hjc*, survival frequency of H1468 (Δ*hel308* Δ*hef*) and H1467 (Δ*hel308* Δ*hjc*) was analysed following treatment with UV irradiation or MMC, Figure 4.14.

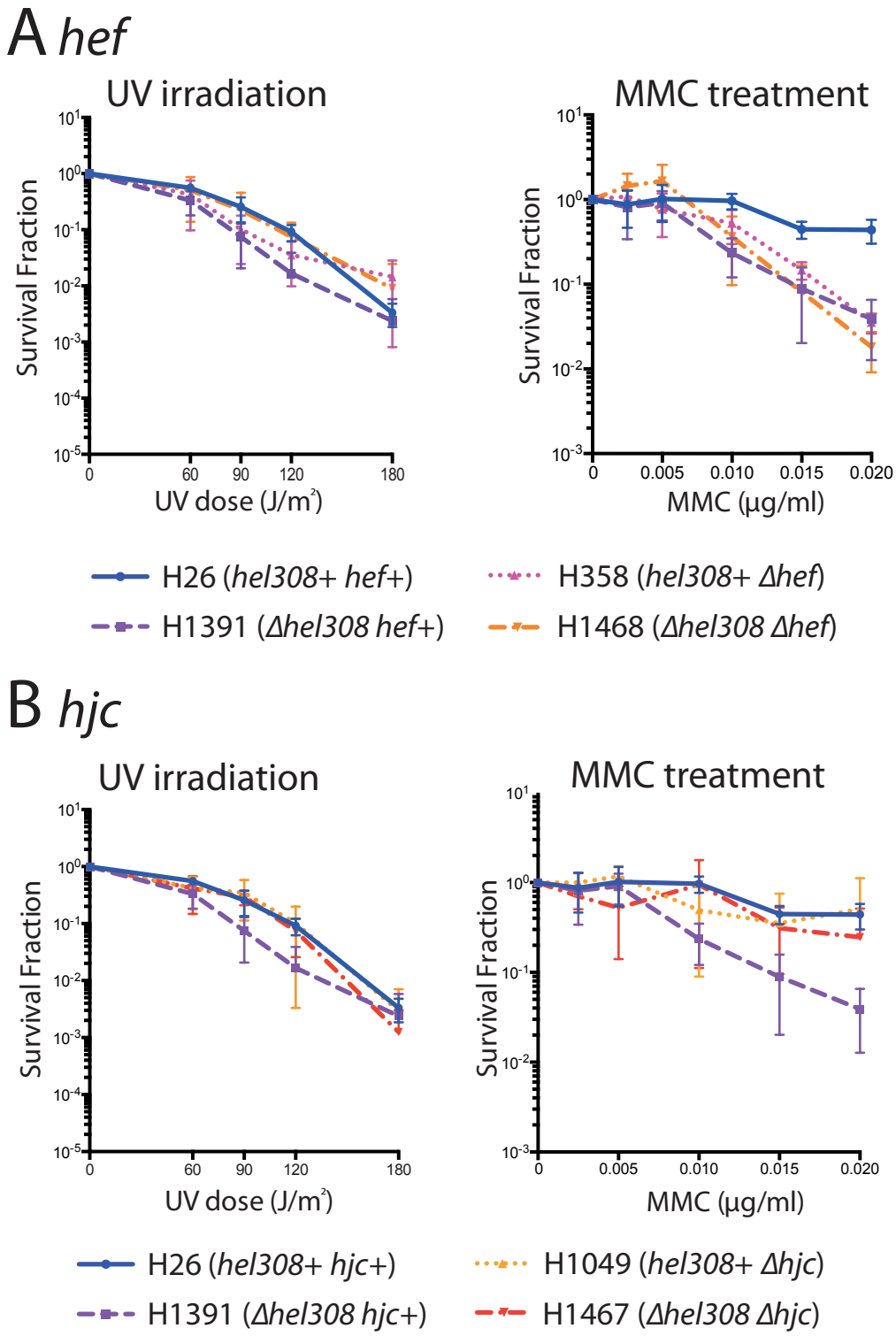


Figure 4.14: **Survival frequency of strains deleted *hel308* and *hef* or *hjc* following treatment with DNA-damaging agents. (A)** Survival for *hef* strains. No observed difference in survival fraction of any strain compared to wild type H26 (*hel308*+ *hef*+) after UV irradiation. H358 (*hel308*+ Δ*hef*) and H1468 (Δ*hel308* Δ*hef*) show survival fraction similar to H1391 (Δ*hel308* *hef*+) after treatment with MMC. **(B)** Survival for *hjc* strains. No observed differences in survival fraction of any strain compared to wild type H26 (*hel308*+ *hjc*+) after UV irradiation. H1391 (Δ*hel308* *hjc*+) shows a reduction in survival fraction whereas H1467 (Δ*hel308* Δ*hjc*) and H1049 (*hel308*+ Δ*hjc*) show survival similar to wild type H26 (*hel308*+ *hjc*+) after treatment with MMC. Survival fraction is calculated relative to un-treated control. Each data point is generated as an average of at least 3 independent trials. Standard error is shown.

Following treatment with UV irradiation H1391 (Δ*hel308 hef+ hjc+*) shows survival similar to wild type H26. Furthermore, both double mutants H1468 (Δ*hel308* Δ*hef*) and H1467 (Δ*hel308* Δ*hjc*) show similar survival compared to wild type H26. The P-values are > 0.05 for all strains compared to wild type (H26) (calculated by a two-tailed *t*-test) indicating no significant difference to wild type. This suggests that Hel308 does not play a role in the repair of UV-induced lesions in the presence or absence of Hjc or Hef

Following treatment with MMC both H1468 (Δ*hel308* Δ*hef*) and H1467 (Δ*hel308* Δ*hjc*) show survival similar to the single Δ*hef* or Δ*hjc* strains (H358 and H1049 respectively). The P-values are > 0.05 for all strains compared (calculated by a two-tailed *t*-test) indicating no significant difference between strains. In Figure 4.14B It appears that the rate of survival is improved upon the double deletion of *hel308* and *hjc* (H1467) compared to H1391 (Δ*hel308 hjc+*). However, upon comparison of these strains using a two-tailed *t*-test, the survival fractions show no significant difference (P-value > 0.05). This suggests that in the absence of *hef and hjc*, *hel308* deletion does not lead to enhanced survival following treatment with MMC.

4.3.2.4 Deletion in Combination with Origins of Replication

Origins of replication are defined sites on the chromosome where DNA replication initiates; origins are the sites where initiator proteins and replication machinery assemble (Bell & Dutta 2002). In *H. volcanii* a strain that lacks all origins of replication (*oriC1*Δ*oriC2* Δ*oriC3* Δ*ori-pHV4-2*) shows no apparent defects and has growth faster than wild type (+7.5%) (Hawkins et al 2013). This strain initiates replication at dispersed sites across the chromosome rather than at discrete origins. This strain also has an absolute dependence on RadA for replication of the entire genome via homologous recombination.

Since *hel308* might have anti-recombinase activity, it would be interesting to observe the effect of a *hel308* deletion in an *oriC1*Δ*oriC2* Δ*oriC3* Δ*ori-pHV4-2* background in *H. volcanii*.

Growth Rate

In order to visually compare the growth rates, H1953 (*Δhel308 ΔoriC1ΔoriC2 ΔoriC3 Δori-pHV4-2*), H1804 (*hel308+ ΔoriC1ΔoriC2 ΔoriC3 Δori-pHV4-2*), and H1391 (*Δhel308 oriC1+ oriC2+ oriC3+ ori-pHV4-2+*), were streaked onto complete media alongside wild-type H26 (*hel308+ oriC1+ oriC2+ oriC3+ ori-pHV4-2+*), Figure 4.15.

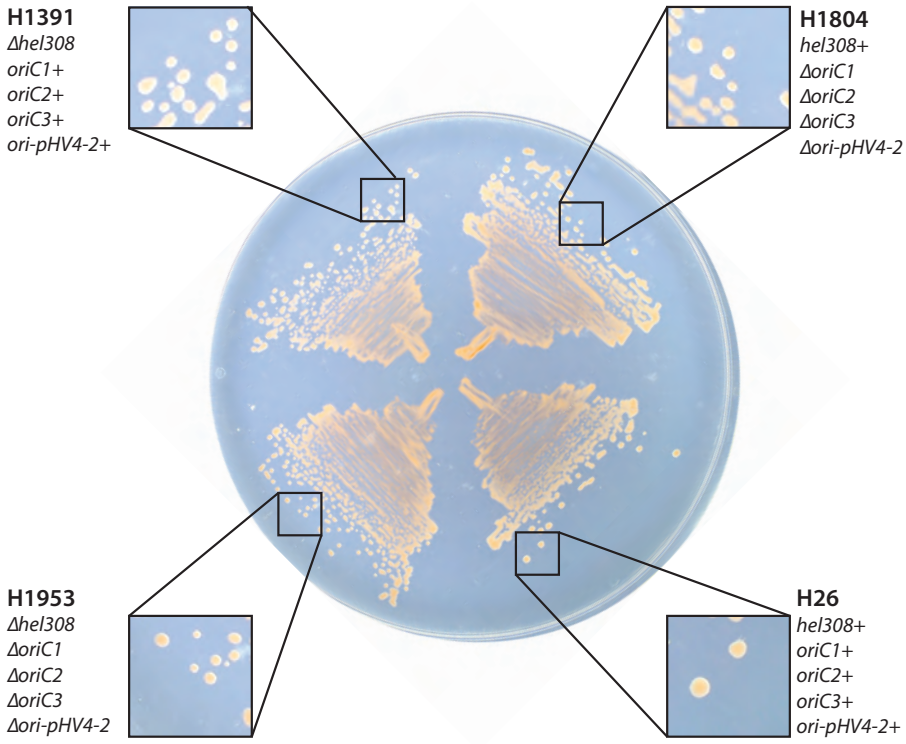


Figure 4.15: **Growth of strains deleted for *hel308* and/or *ΔoriC1 oriC2 oriC3 ori-pHV4-2*.** No difference in single colony size can be seen between any of these strains.

No difference in growth between H1953 (*Δhel308 ΔoriC1ΔoriC2 ΔoriC3 Δori-pHV4-2*), H1804 (*hel308+ ΔoriC1ΔoriC2 ΔoriC3 Δori-pHV4-2*), and H1391 (*Δhel308 oriC1+ oriC2+ oriC3+ ori-pHV4-2+*) can be observed.

As mentioned previously, strains with similar growth rates cannot be distinguished by this qualitative method, and for this reason growth assays in liquid culture were performed, Figure 4.16. The method for this assay has been described previously.

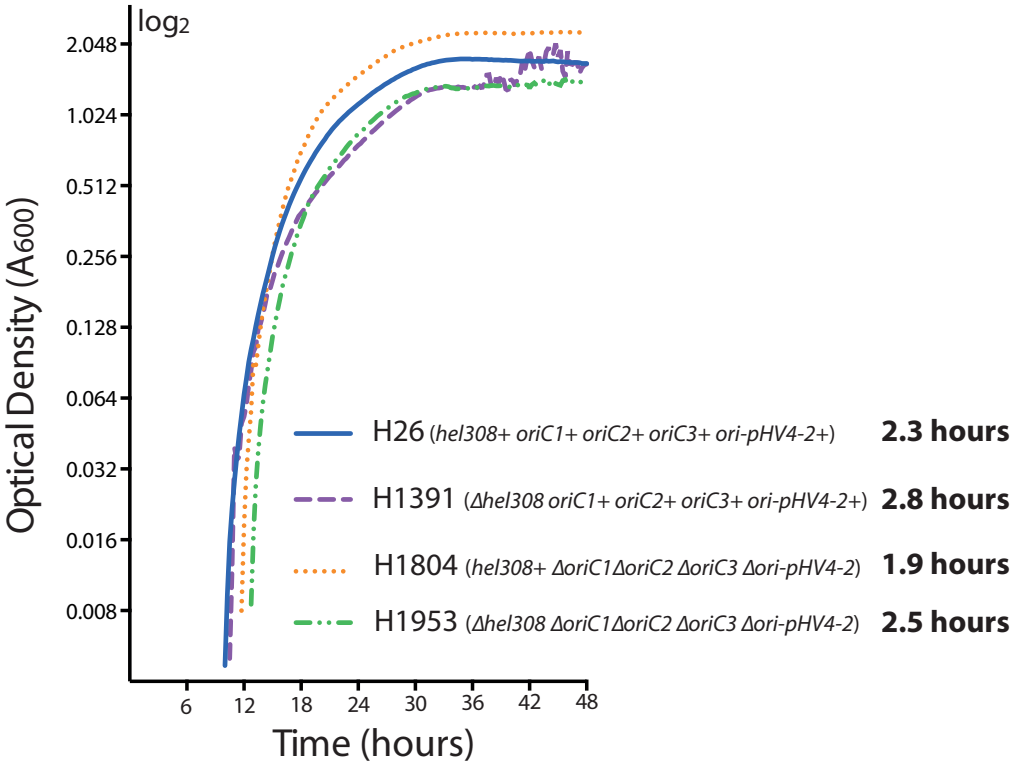


Figure 4.16: **Exponential growth rate of strains deleted for *hel308* and/or *ΔoriC1 oriC2 oriC3 ori-pHV4-2*.** Growth was measured by A_{600}. Generation time is indicated at the side of the strain name. H1804 (*hel308+ ΔoriC1ΔoriC2 ΔoriC3 Δori-pHV4-2*) grows faster than wild type H26 (*hel308+ oriC1+ oriC2+ oriC3+ ori-pHV4-2+*). H1953 (*Δhel308 ΔoriC1ΔoriC2 ΔoriC3 Δori-pHV4-2*) has improved growth compared to H1391 (*Δhel308 oriC1+ oriC2+ oriC3+ ori-pHV4-2+*). Graph plotted on a $\log_2$ scale. 3 repeats were carried out for each strain. Generation time calculated from linear growth phase. All strains and repeats were incubated on the same 48 well plate and measured simultaneously using an Epoch 2 Microplate Spectrophotometer (BioTek).

In agreement with published observations, H1804 (*hel308+ ΔoriC1ΔoriC2 ΔoriC3 Δori-pHV4-2*) shows to have a faster generation time than wild type H26 (*hel308+ oriC1+ oriC2+ oriC3+ ori-pHV4-2+*) (Hawkins et al 2013). H1953 (*Δhel308 ΔoriC1ΔoriC2 ΔoriC3 Δori-pHV4-2*) has a faster generation time compared to H1391 (*Δhel308 oriC1+ oriC2+ oriC3+ ori-pHV4-2+*). This observation is likely due to the improved growth due to the lack of origins.

Survival Following Treatment with DNA-damaging Agents

The survival of H1953 (*Δhel308 ΔoriC1ΔoriC2 ΔoriC3 Δori-pHV4-2*), H1804 (*hel308+ ΔoriC1ΔoriC2 ΔoriC3 Δori-pHV4-2*), H1391 (*Δhel308 oriC1+ oriC2+ oriC3+ ori-pHV4-2+*) and wild type H26 (*hel308+ oriC1+ oriC2+*

oriC3+ ori-pHV4-2+) following treatment with DNA damaging agents was analysed, methods described previously, Figure 4.17.

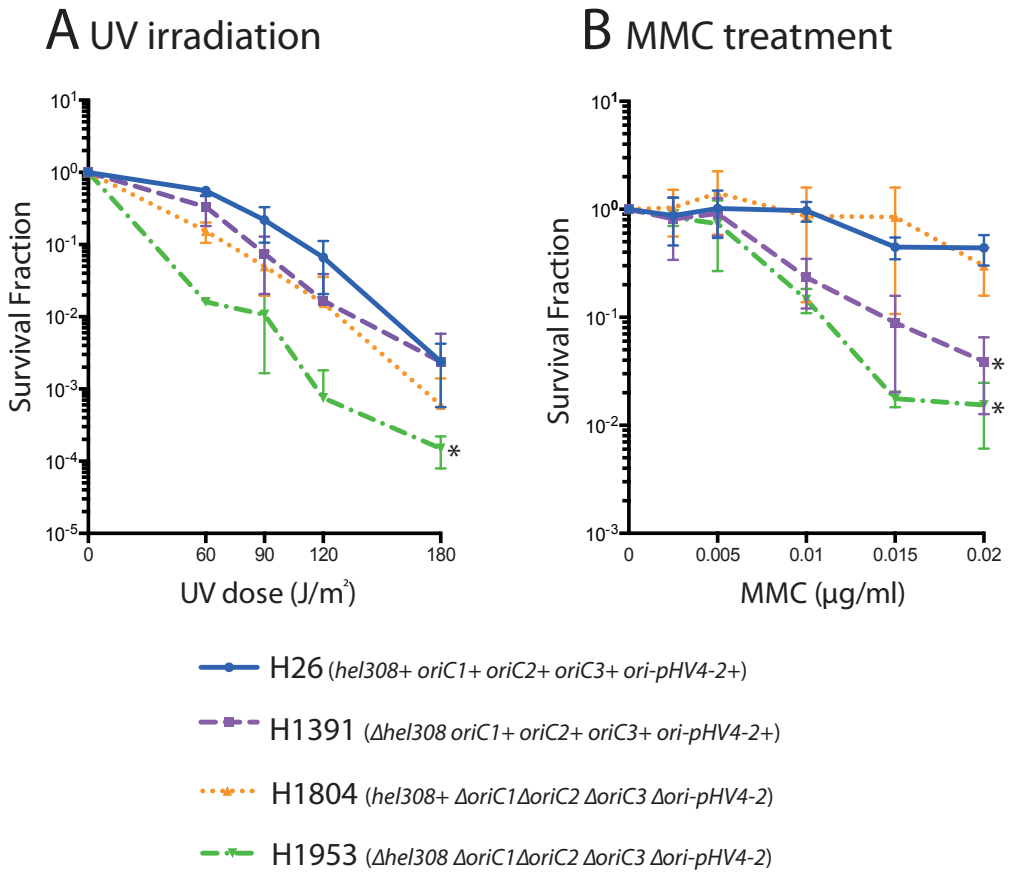


Figure 4.17: **Survival frequency of strains deleted for *hel308* and/or *ΔoriC1 oriC2 oriC3 ori-pHV4-2* following treatment with DNA-damaging agents. (A)** Survival following treatment with UV irradiation. H1953 (*Δhel308 ΔoriC1ΔoriC2 ΔoriC3 Δori-pHV4-2*) shows a slight decrease in survival whereas H1804 (*hel308+ ΔoriC1ΔoriC2 ΔoriC3 Δori-pHV4-2*) and H1391 (*Δhel308 oriC1+ oriC2+ oriC3+ ori-pHV4-2+*) show survival comparable to wild type H26 (*hel308+ oriC1+ oriC2+ oriC3+ ori-pHV4-2+*). **(B)** Survival following treatment with MMC. H1953 (*Δhel308 ΔoriC1ΔoriC2 ΔoriC3 Δori-pHV4-2*) shows a reduction in survival compared to wild-type but shows survival no worse than H1391 (*Δhel308 oriC1+ oriC2+ oriC3+ ori-pHV4-2+*). Survival fraction is calculated relative to un-treated control. Each data point is generated as an average of at least 3 independent trials. Standard error is shown. Asterisk (*) indicates that the highest dose of UV (180 J/m^2) or MMC (0.02 μg/ml) is significantly different to H26 (wild type) with $P < 0.05$. P-value calculated from two-tailed *t*-test in mutated strains compared to H26 (wild-type).

Following UV irradiation, H1953 (*Δhel308 ΔoriC1ΔoriC2 ΔoriC3 Δori-pHV4-2*) shows a significant decrease in survival compared wild type H26 (*hel308+ oriC1+ oriC2+ oriC3+ ori-pHV4-2+*), (P-value = 0.0301, calculated from a two-tailed *t*-test). Whereas H1804 (*hel308+ ΔoriC1ΔoriC2 ΔoriC3 Δori-pHV4-2*) and H1391 (*Δhel308 oriC1+ oriC2+ oriC3+ ori-pHV4-2+*) exhibit survival not significantly different to wild type (P-values > 0.05, calculated from a two-

tailed *t*-test). After treatment with MMC, H1953 (Δ*hel308* Δ*oriC*1Δ*oriC2* Δ*oriC3* Δ*ori-pHV4-2*) exhibits survival significantly lower to that of wild type (P-value = 0.0330, calculated from a two-tailed *t*-test) but shows a survival with no significant difference to H1391 (Δ*hel308 oriC1+ oriC2+ oriC3+ ori-pHV4-2+*) (P-value = 0.2607, calculated from a two-tailed *t*-test).

4.4 Discussion

Expression of Hel308

Hel308 was shown to be constitutively expressed under native and DNA-damaging conditions (treatment with 20 J/m^2 UV-irradiation or 2 μg/ml MMC), with no apparent differences in transcript levels under all conditions. In response to DNA damage, bacteria undergo an SOS response, which involves the up-regulation of DNA repair genes (Little & Mount 1982). Archaea do not appear to have a cellular SOS response, which could account for the lack of up-regulation of Hel308 following DNA damage (Frols et al 2009, McCready et al 2005). Hel308 from the archaeon *Sulfolobus acidocaldarius* shows a cyclic transcription pattern during cell cycle and the transcription of this gene is up-regulated in early S phase (Bernander et al 2010). However, *H. volcanii* does not appear to have a defined cell cycle (Iain Duggin, personal communication).

Hel308 as an Antirecombinase?

RadB is a recombination mediator that asssists the filament formation of RadA on ssDNA during homologous recombination (HR), and deletion of RadB results in slow growth and a marked reduction in HR (~ 5% of a wild type strain) (Haldenby 2007). In all phenotypic assays performed in this study, deletion of *hel308* in combination with *radB* resulted in a reduction in the severity of the defects, compared to a *radB* strain. For example, the generation time of a Δ*hel308* Δ*radB* strain was around 0.8 hours faster than a Δ*radB* strain. Complementation of *hel308* by an episomal plasmid in a Δ*hel308* Δ*radB* background resulted in slow-growing colonies (comparable to a Δ*radB* strain), confirming that the absence of *hel308* is the cause of the defect suppression observed in a Δ*hel308* Δ*radB* strain. Following treatment with MMC, the survival fraction of the Δ*hel308* Δ*radB* strain was markedly increased, compared to that of a Δ*radB* strain. This trend was also observed in flow cytometry profiles, where the Δ*hel308* Δ*radB* strain more closely resembled a wild type strain than a Δ*radB* strain. The suppression effect of deleting *hel308* in a Δ*radB* background can be explained if Hel308 acts as an anti-recombinase

and RadB has more than one function during homologous recombination, Figure 4.18.

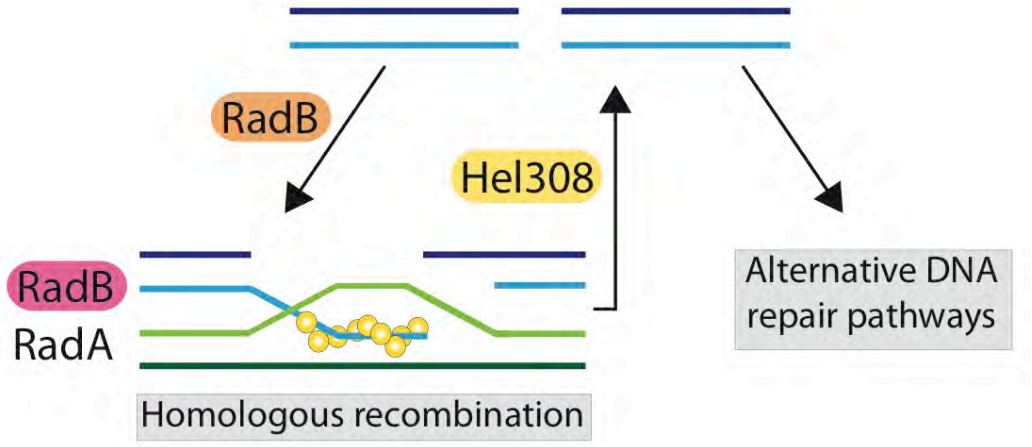


Figure 4.18: **Possible function of Hel08 as an anti-recombinase.** DNA damage (this example depicts a double strand DNA break) can be repaired by homologous recombination or by alternative DNA repair pathways. RadB could act to channel DNA damage into the homologous recombination pathway as well as acting as a recombination mediator to help polymerise RadA onto ssDNA. Hel308 could act as an anti-recombinase to disassemble D-loops.

The first role of RadB involves the catalysis of RadA polymerisation on ssDNA during homologous recombination (pink RadB, Figure 4.18). The second proposed role of RadB is as a regulator of homologous recombination (orange RadB, Figure 4.18). In this scenario, damaged DNA (depicted in Figure 4.18 as a double strand break) is channeled down the homolgous recombination pathway by RadB, RadA assisted by RadB then polymerises onto ssDNA and undergoes strand invasion, and the resuting D-loop may be dissasembled by Hel308. In a Δ*radB* deleted strain, a small amount of homologous recombination is carried out by RadA alone (~ 5% of a wild type strain) (Haldenby 2007). In the absence of RadB, these D-loops are easily dismantled by the anti-recombinase activity of Hel308, resulting in severe growth and DNA damage defects. Upon the deletion of *hel308* in a Δ*radB* background, the small amount of recombination that is carried out solely by RadA is unopposed (due to the absence of the Hel308 anti-recombinase) and so the growth and DNA damage defects are less severe. This hypothesis is supported by the slight increase in recombination frequency observed for a Δ*hel308* Δ*radB* strain compared to a Δ*radB* strain (0.083 and 0.048 times that of wildtype levels, respectively). Due to the poor growth of Δ*radB* deleted

strains, only 22 colonies of the Δ*radB* strain (compared to 64 colonies of the Δ*hel308* Δ*radB* strain) were available to calculate this frequency. Repetition of this assay with more data is necessary to confirm these results.

Hel308 is a 3' to 5' RecQ family helicase and other RecQ helicases with the same polarity also exhibits anti-recombinase activity, for example Srs2 from *S. cerevisiae,* which is homologous to the *E. coli* helicase UvrD. Srs2 is capable of dismantling Rad51 filaments from D-loops (Krejci et al 2012), and mammalian RecQ helicases such as RecQ5, BLM, WRN and FANCJ have also demonstrated the ability to disrupt Rad51-ssDNA filaments (Bugreev et al 2007, Hu et al 2007, Sommers et al 2009). The observation that other RecQ helicases are capable of anti-recombinase activity suggests that Hel308 in archaea could also act as an anti-recombinase. Since other anti-recombinases are present, Hel308 could have a redundant role in mammalian species.

Hel308 does not Appear to Regulate Hjc or Hef During ICL Repair.

Deletion of Hel308 in combination with Holliday junction resolvases Hef and Hjc has no effect on generation time. The lack of a synthetic defect suggests that Hel308 does not function in the later stages of homologous recombination alongside Hef and Hjc. Although the strain deleted for Δ*hel308* Δ*hjc* displays a severe growth defect on solid media but no growth defect in liquid culture. Since strains are growing vigorously in exponential phase in liquid culture, this growth assay is more reliable. Repetition of Δ*hel308* Δ*hjc* growth assay on solid media may be required.

Following treatment with MMC, both Δ*hel308*Δ*hef* and Δ*hel308*Δ*hjc* strains show survival similar to the single Δ*hef* and Δ*hjc* strains respectively suggesting that Hel308 does not act alongside Hjc or Hef during the repair of interstrand cross links in *H.volcanii*. However, yeast two-hybrid analyses revealed that the Hel308 homologue Hjm from *Sulfolobus tokodaii* physically interacts with Hjc *in vitro* (Li et al 2008), suggesting further investigation into this relationship in *H.volcanii* may be required.

Is Hel308 Involved in Recombination Mediated DNA Replication?

A *H. volcanii* strain deleted for the four chromosomal origins of replication (Δ*oriC*1Δ*oriC2* Δ*oriC3* Δ*ori-pHV4*) grows 7.5% faster than wild type and exhibits an absolute dependence on RadA for replication of the entire genome via homologous recombination (Hawkins et al 2013). In this study, a strain deleted for *hel308 oriC*1*oriC2 oriC3* and *ori-pHV4-2* was shown to grow faster than a Δ*hel308* strain, which is likely due to faster growth due to the lack of origins. However, the Δ*hel308* Δ*oriC*1Δ*oriC2* Δ*oriC3* Δ*ori-pHV4-2* strain had a longer generation time than a strain deleted for all origins but not deleted for *hel308*. If Hel308 solely acts as an anti-recombinase, then in the origin-less strain where DNA replication relies on recombination, the presence of Hel308 should be detrimental; upon deleting *hel308,* recombination will not be antagonised by Hel308 and the cell should grow faster. However the opposite is seen: strains grow slower after *hel308* is deleted in an origin-less background.

This suggests that Hel308 could have a second function in the cell. If Hel308 is also involved in promoting the (re)initiation of DNA replication, then deletion of *hel308* will be detrimental to a strain that only has one route of replicating its genome (via homologous recombination). A strain deleted for *hel308, oriC*1*oriC2 oriC3* and *ori-pHV4-2* displayed a decrease in survival after treatment with UV, relative to that seen in either a *hel308* deleted strain or an origin-less strain. Hel308 is not known to be involved in the repair of double strand DNA breaks or other forms of UV-induced damage, but if DNA replication is impaired upon deletion of *hel308* in the origin-less background, then this would account for the DNA damage phenotype.

4.5 Future perspectives

To investigate the role of Hel308 as a regulator of homologous recombination, it would be of interest to analyse a Δ*hel308*Δ*radA* strain. However, the generation of such strain was not possible during this study. This could mean that RadA is essential in a *hel308*-deleted background but the more likely reason is due to the techincal difficulties in generating *radA* deletions. The

pop-in pop-out gene deletion method used in *H. volcanii* utilises homologous recombination (and therefore RadA), which makes the deletion of *radA* challenging. If further efforts to delete *radA* should prove unsuccessful, then *radA* could be placed under the tryptophan-inducible p.*tnaA* promoter and assayed under conditions lacking tryptophan. This method was used by Hawkins and colleagues to demonstrate that *radA* deletion is impossible in an origin-less background (Hawkins et al 2013). If a Δ*hel308*Δ*radA* strain can be generated, then based on the scenario depicted in Figure 4.18 it is expected that the growth and DNA damage defects would be no worse than a Δ*radA* strain. This is because Δ*radA* strains are unable to carry out homologous recombination (Woods & Dyall-Smith 1997). Therefore, no D-loop intermediates would be formed for Hel308 to dismantle and so deleting *hel308* would have no effect on this strain.

Another method to investigate the role of Hel308 in the regulation of homologous recombination would be to combine *hel308* deletions with point mutations in *radA* that act as supressors of the Δ*radB* phenotype. RadB is a recombination mediator, which assists RadA polymerisation onto ssDNA during homologous recombination. Strains deleted for *radB* are slow-growing and carry out homologous recombination at ~5% of the level seen in a wild type strain. The point mutants *radA-S101P* and *radA-A196V* were identified as supressors of the ΔradB phenotype (Haldenby 2007, Wardell 2013), Figure 4.19 .

RadA monomers polymerise onto ssDNA through a 'ball and socket' mechanism, where an invariant phenylalanine residue inserts into a hydrophobic socket of an adjacent monomer (Shin et al 2003). It is thought that RadB may activate RadA into a polymerisation-competent conformation by decreasing the flexibility of the N-terminal domain, thereby 'locking' RadA into an active state (Wardell 2013). The RadA-S101P mutation is located in the linker region between the core domain and the N-terminal domain of RadA. It is thought that this mutation locks the RadA-S101P monomer in a competent conformation for polymerisation. By contrast, the RadA-A196V mutation is

located in the monomer:monomer binding pocket in the core domain of RadA. This mutation increases the hydrophobicity of the binding pocket, resulting in stronger interactions between RadA monomers; as a consequence, RadA-A196V does not require RadB for polymerisation (Haldenby 2007, Wardell 2013).

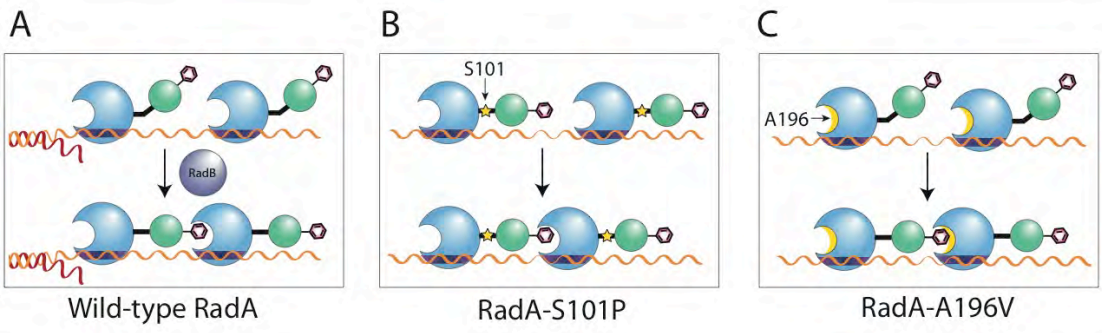


Figure 4.19: **Δ*radB* suppressors are *radA-S101P* and *radA-A196V*. (A)** Wild-type RadA may not be in the correct conformation to form a nucleoprotein filament on ssDNA. RadB may be required for efficient polymerisation. **(B)** RadA-S101P may already be in the correct conformation for polymerisation, and not require RadB. S101P is indicated by yellow star. **(C)** RadA-A196V increases the hydrophobicity of the RadA binding pocket, which could result in stronger hydrophobic interactions between RadA monomers, negating the need for RadB. A196V indicated by yellow arc. RadA core domain in blue and N-terminal domain in green. Adapted from (Li et al 2008, Wardell 2013)

Combining deletions of *hel308* and *radB* with these chromosomal alleles of *radA* that act as Δ*radB* suppressors may give insights into the mechanistic role of Hel308. In these strains, *radB* is deleted but homologous recombination occurs at near-wild type levels, since RadA is able to function 'normally' by itself. Utilising this genetic background may help determine how Hel308 interacts with the recombination apparatus and answer the question of whether Hel308 interacts with RadB, RadA or both. Furthermore, this genetic background may help to address the hypothesis that RadB could have two roles in homologous recombination as described in Figure 4.18.

4.6 Conclusion

Hel308 is a helicase that has previously been proposed to act in homologous recombination. Through genetic analysis of *hel308* deletions in combination with mutations in several different recombination proteins, an insight into the role of Hel308 in homologous recombination has been gained.

Chapter 5: Genetic Analysis of *hel308* Point Mutants

5.1 Background

The phenotypic analysis of amino acid point mutations is a powerful tool to dissect the function of a protein (Sinha & Nussinov 2001).

Biochemical analysis of point mutations within Hel308 has been carried out in a number of archaeal organisms. This is discussed in Chapter 1: *Introduction*, Section 1.7: *Hel308,* but mutations relevant to this study are discussed in further detail in the results sections of this chapter. However, relatively few point mutations in Hel308 have been studied using a genetic approach. *Haloferax volcanii* has available many genetic tools and is easily cultured, and therefore is an ideal organism to use for the genetic analysis of Hel308 point mutants. However, biochemical characterization of *Haloferax volcanii* Hel308 point mutations is experimentally demanding as all biochemical assays will require adapting to 2.5 M NaCl conditions.

The protein sequence of Hel308 from *H. volcanii* is well conserved with respect to other archaeal Hel308 helicases from *Archaeoglobus fulgidus*, *Sulfolobus solfataricus*, *Pyrococcus furiosus* and *Methanothermobacter thermautotrophicus,* Figure 5.1. Hel308 from *A. fulgidus*, *S. solfataricus* and *P. furiosus* have been included in this protein alignment as crystal structures of Hel308 are available from these species (Buttner et al 2007, Oyama et al 2009, Richards et al 2008). The crystal structures will give insights into possible effects that point mutations will have on Hel308 from *H. volcanii*. *M. thermautotrophicus* has been included in this alignment because several biochemical characterisations of Hel308 within this organism have already been reported (Woodman & Bolt 2011, Woodman et al 2007). Furthermore, any Hel308 point mutations giving interesting genetic phenotypes in *H. volcanii* could be biochemically characterised in this organism in the future.

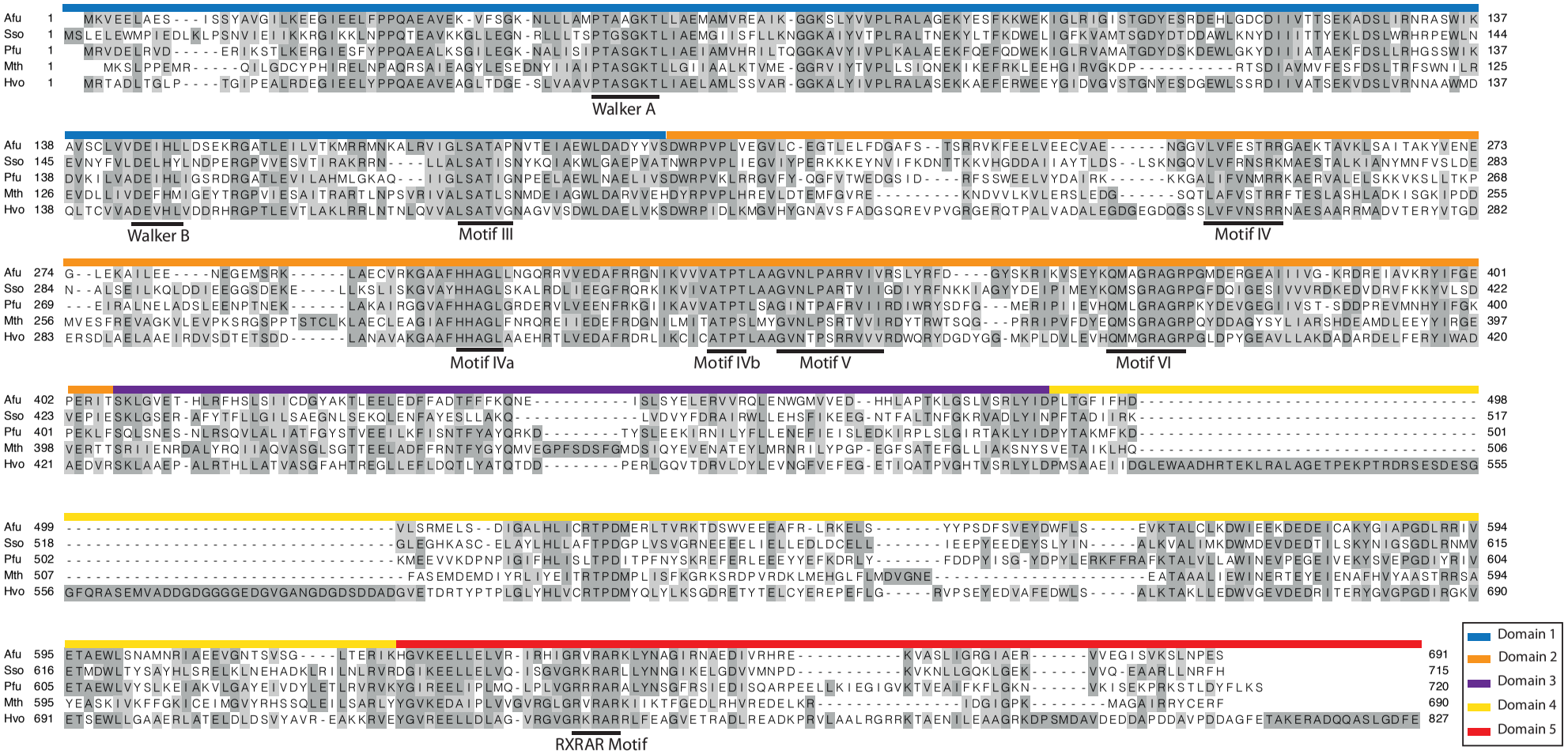


Figure 5.1: **Hel308 protein sequence alignment.** Multiple sequence alignment of Hel308 from *Archaeoglobus fulgidus* (Afu), *Sulfolobus solfataricus* (Sso), *Pyrococcus furiosus* (Pfu), *Methanothermobacter thermautotrophicus* (Mth) and *Haloferax volcanii* (Hvo). Hel308 domains and motifs are indicated. Alignment carried out in MacVector using ClustalW (Gonet; penalty for open gap = 110, extend gap = 0.2)

As a consequence of a conserved protein sequence, the predicted structure of Hel308 from *H. volcanii* is also conserved with other archaeal Hel308 proteins such as Hel308 from *A. fulgidus*, Figure 5.2.

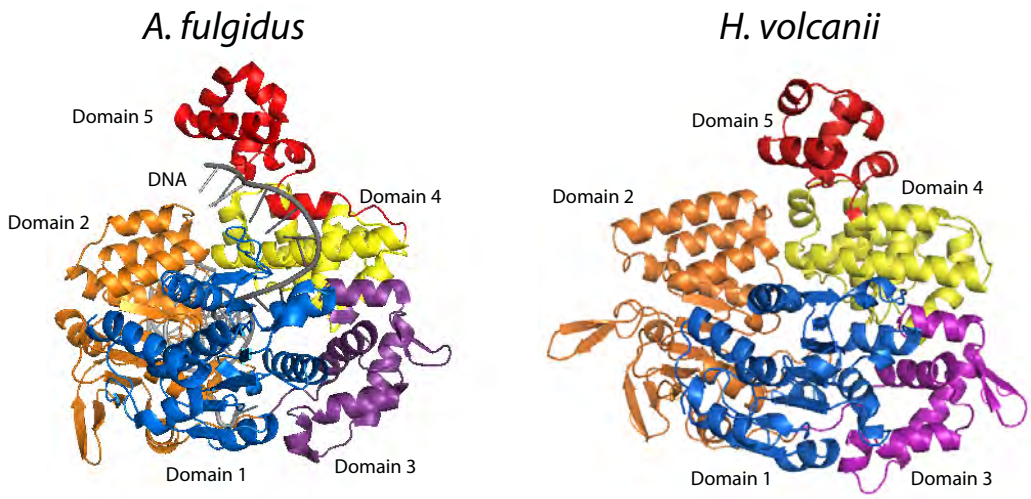


Figure 5.2: **Structure of Hel308 from *A. fulgidus* and *H. volcanii*.** The structure of Hel308 from *H. volcanii* was predicted from the amino acid sequence using Phyre2 (Kelley et al 2015) and is similar to crystal structure of Hel308 from *A. fulgidus* (in complex with DNA). Domains are indicated.

The mutations generated during this study are found in domains 1, 2, the linker between domains 2-3, and domain 5. The details and rationale behind each mutation are described in the relevant results sections.

5.2 Aims

Phenotypic analysis of Hel308 point mutants will give insights into the role that Hel308 plays within the cell. The aims of this chapter are to:

- Generate point mutants in the key motifs in domains 1, 2 and 5, as well as within the previously unstudied domain 2-3 linker region of Hel308.
- Analyse the phenotypes of strains containing the Hel308 point mutants after treatment with DNA damaging agents such as UV and MMC, which lead to double strand breaks and interstrand crosslinks, respectively.
- Analyse the growth rate of strains containing the Hel308 point mutants.

- Analyse the recombination frequency and ratio of crossover vs. non-crossover products formed by strains containing the Hel308 point mutants.

5.3 Results

5.3.1 K53R Walker A and D145N Walker B mutations

The Walker A motif with the consensus sequence of GXXXXGKT/S (where X can be any residue) is the site of ATP binding in many proteins. Crystal structures reveal that Walker A motifs have a looped structure which wraps around nucleotides. Within this motif is are highly conserved lysine and threonine residues which bind to the γ-phosphate of the nucleotide (Ramakrishnan et al 2002). The Walker B motif has consensus sequence of DEAD/DEVH. The first aspartic acid residue of the consensus sequence coordinates Mg^{2+}, which is required for ATP binding and the glutamate is essential for ATP hydrolysis (Caruthers & McKay 2002). In helicases, this ATPase motif is responsible for coupling of ATP hydrolysis and helicase activity. Mutations within these motifs abolish ATPase and helicase activity.

A wealth of biochemical information is available for Walker A and Walker B motif mutations in motor proteins such as helicases. The biochemical effect of these mutations in archaeal Hel308 helicases has already been studied; in *Sulfolobus solfataricus* and *Methanothermobacter thermautotrophicus* it was shown that mutation of the conserved lysine residue of the Walker A motif resulted in helicase-defective proteins (Richards et al 2008, Woodman et al 2011). As mentioned previously, biochemical analysis of proteins expressed from halophilic organisms such as *H. volcanii* is problematic. Therefore a genetic analysis of Walker A and Walker B mutants was carried out. In order to analyse these mutations, the strains H1554 (*hel308-K53R*) and H1555 (*hel308-D145N*) were generated containing Walker A and Walker B motif mutations respectively. Details of the position of these mutations and amino acid substitutions can be seen in Figure 5.3.

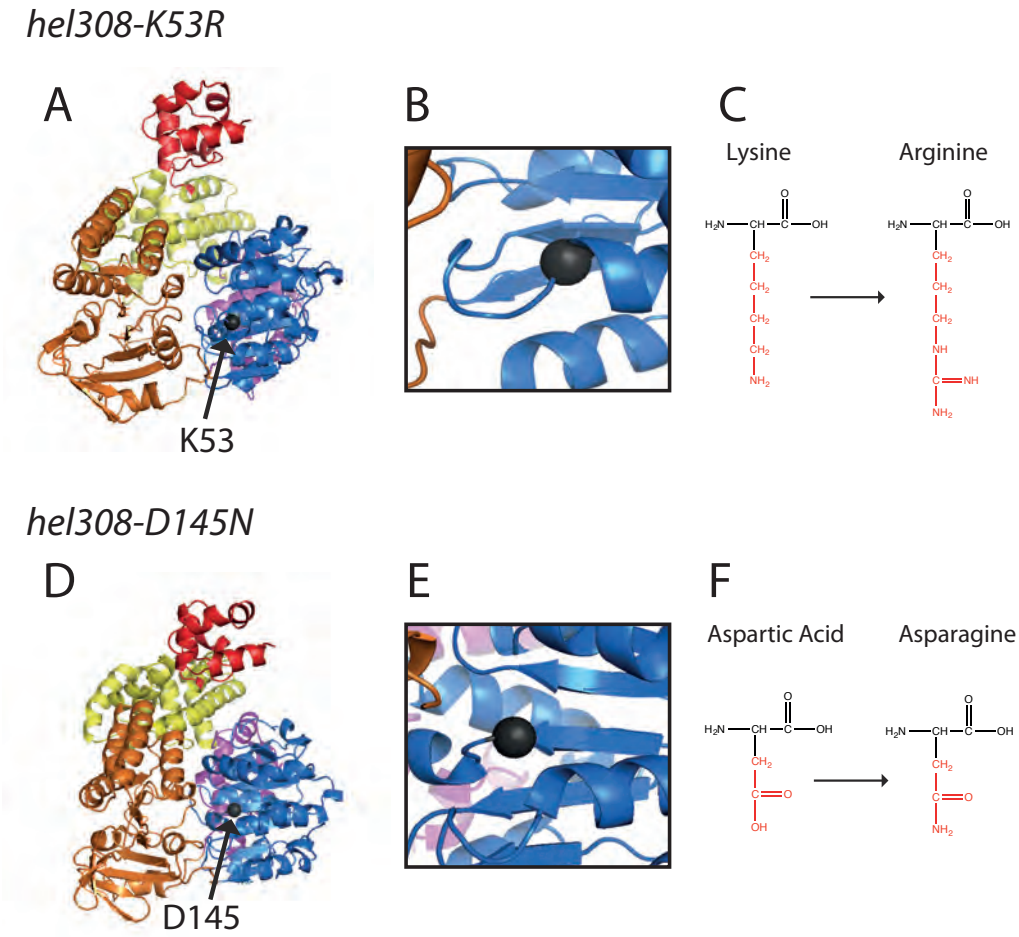


Figure 5.3: **Location of K53 and D145 residues in the structure of Hel308.** The structure of Hel308 from *H. volcanii* was predicted from the amino acid sequence using Phyre2 (Kelley et al 2015). **(A)** Predicted structure of Hel308, position of K53 is located in the RecA fold of domain 1, indicated with black sphere. **(B)** K53 is in the Walker A motif, typified by a looped structure. **(C)** Amino acid substitution of lysine with arginine, R group is shown in red. **(D)** Predicted structure of Hel308, position of D145 is located in the RecA fold of domain 1, indicated with black sphere. **(E)** D145 is in the Wlaker B motif, which is found in a loop structure stacked above the Walker A motif. **(F)** Amino acid substitution of aspartic acid with asparagine, R group is shown in red. Structure of of Hel308 from *H. volcanii,* domain 1 (blue), domain 2 (orange), domain 3 (purple), domain 4 (yellow) and domain 5 (red).

In H1554 (*hel308-K53R*) the positively charged lysine (K) was mutated to arginine (R), in this mutation the positive charge of the side chain is maintained but the bulky side chain of arginine sterically hinders the binding of ATP (Hishida et al 1999). In H1555 (*hel308-D145N*) the negatively charged aspartic acid (D) was mutated to asparagine (N) that has a polar uncharged side chain. This change in charge interferes with the coordination of Mg^{2+}, which results in a loss of affinity for ATP.

Survival Following Treatment with DNA-damaging Agents

Strains deleted for *hel308* do not appear to have a growth defect after treatment with the ss/dsDNA break inducing agent UV but do have an inscreased sentisitvity to DNA crosslinking agents such as MMC (Adelman et al 2013, Guy & Bolt 2005, Takata et al 2013). In order to study the effects of *hel308* Walker A and Walker B mutants on repair of DNA breaks and crosslinks, cells were treated with UV and MMC respectively, as described previously (Chapter 4: Genetic chatacterisation of *hel308*). The survival fraction of H1554 (*hel308-K53R*) and H1555 (*hel308-D145N*) was compared to that of H1392 (*Δhel308*) and H26 (*hel308+*), Figure 5.4.

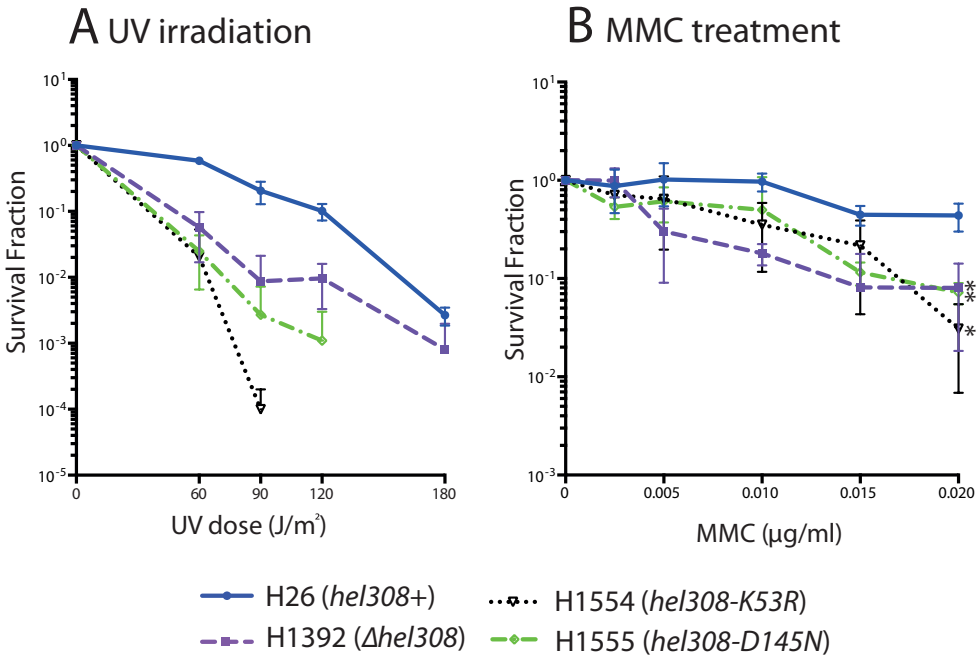


Figure 5.4: **Survival frequency of *hel308-K53R* and *hel308-D145N* following treatment with DNA-damaging agents. (A)** Survival following treatment with UV irradiation. Both H1554 (*hel308-K53R*) and H1555 (*hel308-D145N*) have a poor survival whereas the survival of H1392 (*Δhel308*) is comparable to wild type H26 (*hel308+*). **(B)** Survival following treatment with MMC. Both H1554 (*hel308-K53R*) and H1555 (*hel308-D145N*) show a survival fraction comparable to H1392 (*Δhel308*). Survival fraction is calculated relative to un-treated control. Each data point is generated as an average of at least 3 independent trials. Standard error is shown. Asterisk (*) indicates that the highest dose of UV (180 J/m^2) or MMC (0.02 μg/ml) is significantly different to H26 (wild type) with $P < 0.05$. P-value calculated from two-tailed *t*-test in mutated strains compared to H26 (wild-type).

Unlike H1392 (Δ*hel308*), H1554 (*hel308-K53R*) and H1555 (*hel308-D145N*) both have poor survival fractions after treatment with UV irradiation, no growth was seen after 90 -120 J/m^2. Since Hel308 is not known to be involved in repairing DNA damage induced by UV, this phenotype could be due to these ATPase defective helicases persisting on damaged DNA and therefore interfering with repair of UV DNA damage. After treatment with MMC, H1554 (*hel308-K53R*) and H1555 (*hel308-D145N*) have a survival fraction significantly lower than that of wild type (P-values < 0.05, calculated from a two-tailed *t*-test). But not significantly different to H1392 (Δ*hel308*) (P-values < 0.05, calculated from a two-tailed *t*-test). This is expected as Walker A and Walker B mutations are known to abolish the ATPase and helicase activity of helicases.

Growth Rate

In order to visually compare the growth rates, H1554 (*hel308-K53R*) and H1555 (*hel308-D145N*) were streaked onto complete media alongside H1392 (Δ*hel308*) and H26 (*hel308+*), Figure 5.5.

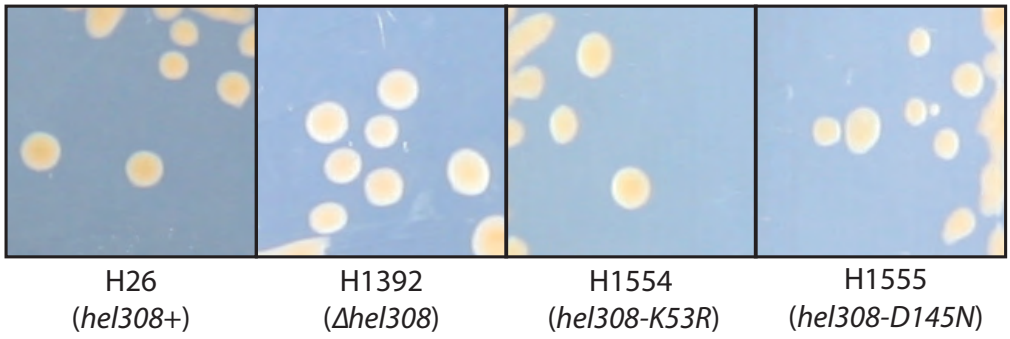


Figure 5.5: **Growth of *hel308-K53R* and *hel308-D145N* strains.** No difference in single colony size can be observed between H1554 (*hel308-K53R*), H1555 (*hel308-D145N*), H1392 (Δ*hel308*) and H26 (*hel308+*). All plate images are to the same scale.

The colony sizes of H1554 (*hel308-K53R*) and H1555 (*hel308-D145N*) were no different to H1392 (Δ*hel308*) or H26 (*hel308+*). A limitation of using growth on solid media is that only large variations can be readily observed, strains with similar growth rates cannot be distinguished by this qualitative method. For these reasons, growth of these strains in liquid culture was

measured as described previously (Chapter 4: Genetic chatacterisation of *hel308*), Figure 5.6.

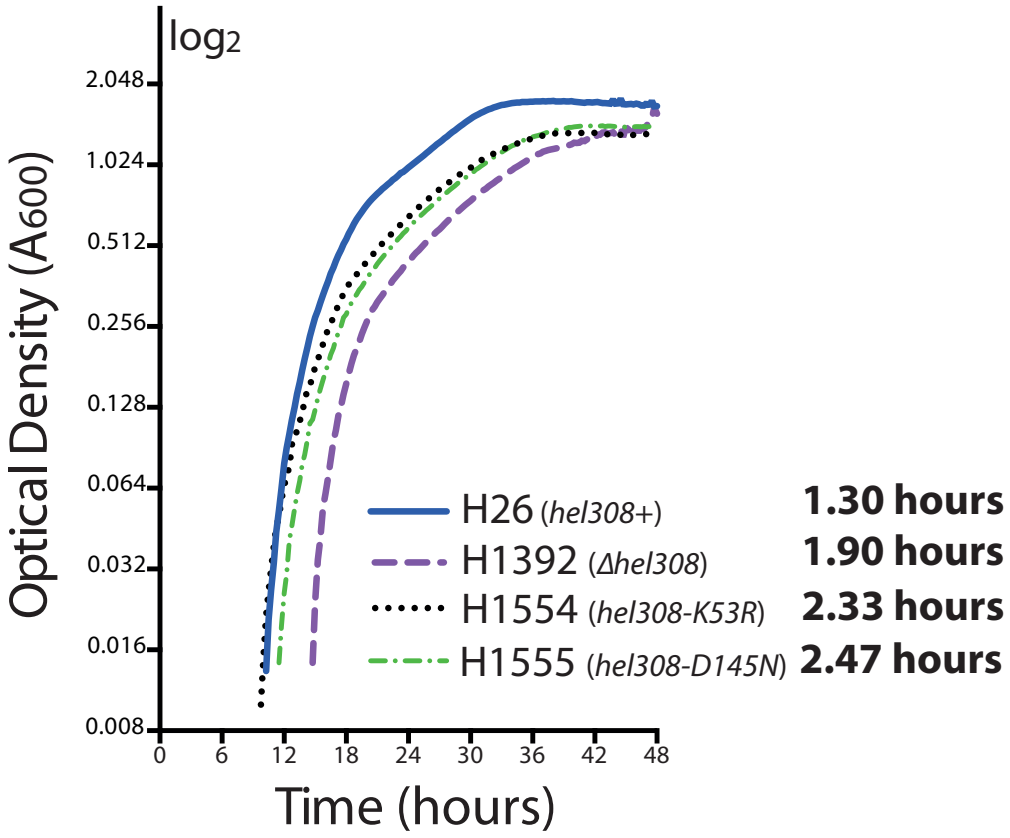


Figure 5.6: **Exponential growth rate of *hel308-K53R* and *hel308-D145N* strains.** Growth was measured by A_{600}. Generation time is indicated at the side of the strain name. H1554 (*hel308-K53R*), H1555 (*hel308-D145N*) have a slower growth rate than H1392 (Δ*hel308*). Graph plotted on a $\log_2$ scale. 3 repeats were carried out for each strain. Generation time calculated from linear growth phase. All strains and repeats were incubated on the same 48 well plate and measured simultaneously using an Epoch 2 Microplate Spectrophotometer (BioTek).

H1554 (*hel308-K53R*), H1555 (*hel308-D145N*) have a slower growth rate than H1392 (Δ*hel308*). These ATPase defective helicases could be persisting on the DNA and therefore interfering with DNA replication, DNA repair and transcription, this could account for the slower growth compared to H1392 (Δ*hel308*).

5.3.2 Domain 2 Mutations F316A, H317G and E330G

In the Hel308 DNA co-crystal from *Archaeoglobus fulgidus*, single stranded DNA threads over two RecA folds present in domains 1 and 2 and passes

through the central cavity of the protein (Woodman & Bolt 2011) Figure 5.7A. Domain 2 of Hel038 binds and separates the strands of dsDNA via a prominent β-hairpin loop that melts two base pairs of the duplex DNA.

Domain 2 makes several contacts with DNA; the backbone of the unwound 3' binds at motifs IVa and IVb and upon ATP binding, motif IV is thought to push DNA towards domain 1 as Hel308 translocates. Furthermore, a ratchet helix from domain 4 interacts with motif IVa in domain 2. It is proposed that the position of the ratchet helix is modulated by an ATP dependent movement in domain 2 which could directly alter the geometry of the single strand DNA binding site between domains 1 and 4 (Buttner et al 2007).

Motif IV and the highly conserved motif IVa (HHAGL) in particular, is important in binding the 3' tail of unwound DNA passing through the central pore of Hel308 and helicase traslocation. Therefore, it is of interest to study residues in this area for this reason the strains H2396 (*hel308-F316A*), H2097 (*hel308-H317G*) and H2078 (*hel308-E330G*) were constructed. In the *A. fulgidus* Hel308-DNA co-crystal, residues F300 and H301 (H316 and H317 in *H. volcanii* respectively) are positioned in the surface of the central pore of the helicase and next to the 3' tail of unwound DNA, Figure 5.7B. The *hel308-E330G* mutation (H314 in *A. fulgidus)* is located in α- helix 14 between motifs IVa and IVb and faces away from the central pore of the helicase and towards the 'back' of the protein and domain 1, Figure 5.7B. A schematic of this region and these mutations can be seen in Figure 5.7C.

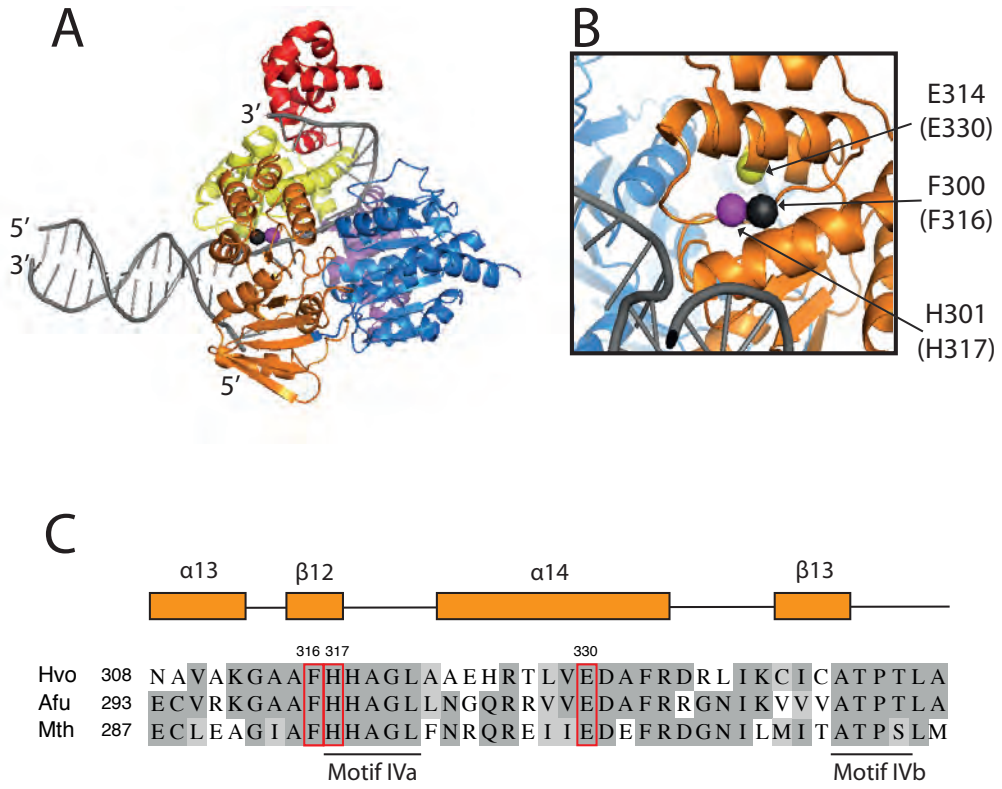


Figure 5.7: **Hel308-DNA co-crystal structure from *A. fulgidus*. (A)** Hel308-DNA co-crystal; DNA (grey), domain 1 (blue), domain 2 (orange), domain 3 (purple), domain 4 (yellow) and domain 5 (red). Single stranded DNA passes through the central pore of the helicase and interacts with domain 2**. (B)** A close up of domain 2, H301 (purple sphere) in motif Iva and F300 (black sphere) are positioned facing into the central pore of the helicase next to unwound 3' DNA. E314 (yellow sphere) is located at the 'back' of domain 2 facing away from the central pore, equivalent *H. volcanii* residues are shown in brackets. Domain 4 has been removed from this image for clarity. **(C)** Alignment of domain 2 from *H. volcanii, A. fulgidus* and *M. thermautotrophicus,* residues to be mutated are shown in red boxes and residue numbers above correspond to *H. volcanii* Hel308.

In H2396 (*hel308-F316A*) the bulky hydrophobic phenylalanine (F) was mutated to the smaller hydrophobic amino acid alanine (A). The change in amino acid size could interfere with inter-domain and DNA interactions. In H2097 (*hel308-H317G*) the bulky positively charged histidine (H) was mutated to the small non-polar amino acid glycine (G). Amino acid 317 is in the conserved motif Iva which interacts with negativly charged DNA backbone, removal of the postiviely charged H could result in a loss or reduction of DNA binding. In H2078 (*hel308-E330G*) the negatively charged glutamic acid (E) was mutated to the small non-polar amino acid glycine (G), this change of charge could interfere with inter-domain and DNA interactions within domain

2. The amino acid substitutions and the position of these mutations in a Hel308 predicted structure from *H. volcanii* are detailed in Figure 5.8.

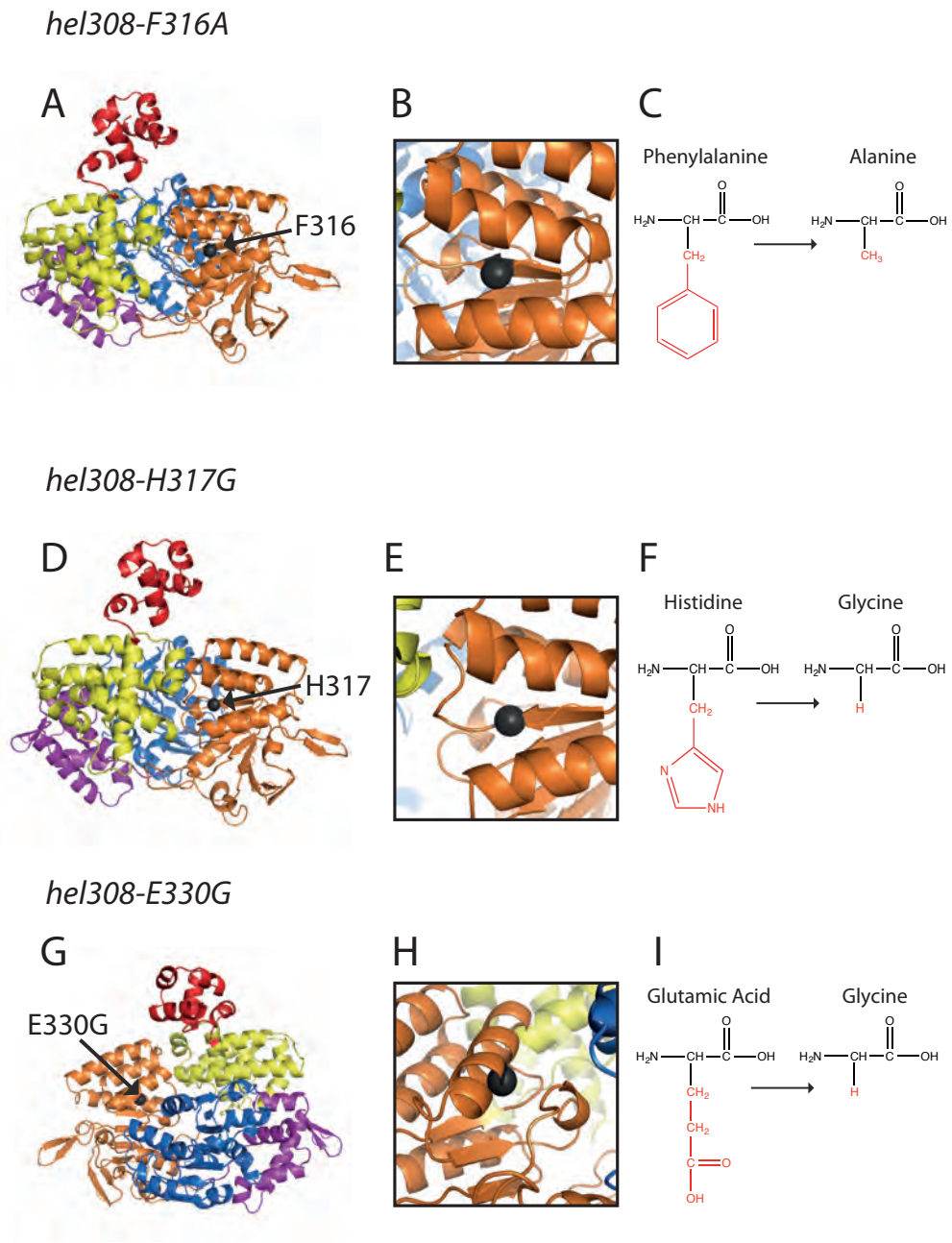


Figure 5.8: **Location of F316, H317 and E330 residues in the structure of Hel308.** The structure of Hel308 from *H. volcanii* was predicted from the amino acid sequence using Phyre2 (Kelley et al 2015). **(A)** Predicted structure of Hel308, position of H316 is located adjacent to motif IVa in domain 2, indicated with black sphere. **(B)** H316 is angled towards the central pore of the helicase and ssDNA.**(C)** Amino acid substitution of phenylalanine with alanine, R group is shown in red. **(D)** Predicted structure of Hel308, H317 is motif IVa in domain 2, indicated with black sphere. **(E)** H317 is angled towards the central pore of the helicase and ssDNA. **(F)** Amino acid substitution of histidine with glycine, R group is shown in red. **(G)** Predicted structure of Hel308, position of H330 is located between motifs Iva and IVb in domain 2, indicated with black sphere. **(B)** H330 is angled towards the 'back' of the helicase and twards domain 1. **(C)** Amino acid substitution of glutamic acid with glycine, R group is shown in red. Domain 1 (blue), domain 2 (orange), domain 3 (purple), domain 4 (yellow) and domain 5 (red).

Survival Following Treatment with DNA-damaging Agents

In order to study the effects of the *hel308* domain 2 mutants, H2078 (*hel308-E330G*), H2097 (*hel308-H317G*) and H2396 (*hel308-F316A*) on repair of DNA breaks and crosslinks, cells were treated with UV and MMC respectively, as described previously, Figure 5.9.

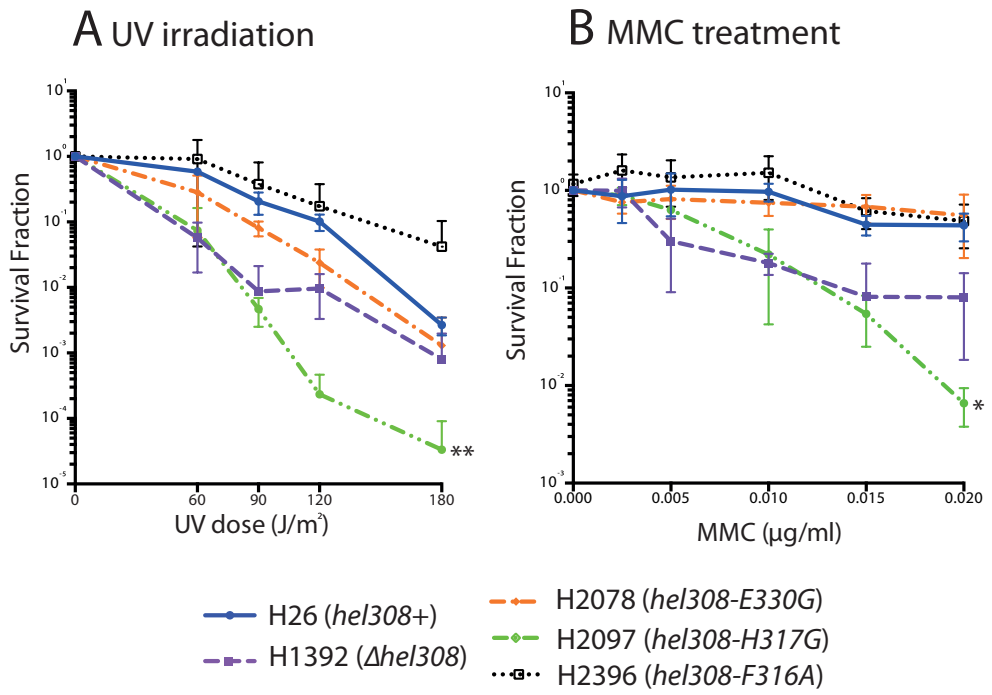


Figure 5.9: **Survival frequency of H2078 *(hel308-E330G)*, H2097 *(hel308-H317G)* and H2396 *(hel308-F316A)* following treatment with DNA-damaging agents. (A)** Survival following treatment with UV irradiation. H2078 *(hel308-E330G)* shows survival comparable to H1392 (Δ*hel308*) and wild type H26 (*hel308*+). H2396 *(hel308-F316A)* shows an improvement in survival and H2097 *(hel308-H317G)* has a low survival fraction compared to H1392 (Δ*hel308*) and wild type H26 (*hel308*+). **(B)** Survival following treatment with MMC. H2078 *(hel308-E330G)* and H2396 *(hel308-F316A)* have survival fractions similar to wild type H26 (*hel308*+) levels, whereas H2097 *(hel308-H317G)* has a survival fractions worse than that of H1392 (Δ*hel308*). Survival fraction is calculated relative to un-treated control. Each data point is generated as an average of at least 3 independent trials. Standard error is shown. Asterisk (*) indicates that the highest dose of UV (180 J/m^2) or MMC (0.02 μg/ml) is significantly different to H26 (wild type) with $P < 0.05$, ** indicates $P < 0.01$. P-value calculated from two-tailed *t*-test in mutated strains compared to H26 (wild-type).

H2078 *(hel308-E330G)* and H2396 *(hel308-F316A)* have survival fractions with no significant difference to wild type H26 (*hel308*+) after treatment with either UV irradiation or MMC, (P-values > 0.05, calculated from a two-tailed *t*-tests). Indicating that these residues are not essential to the function of Hel308 or that these mutations do not cause a significant changes to affect the

function of Hel308. H2097 *(hel308-H317G)* has a significantly lower survival fraction in comparison with wild type (H26) after treatment with UV and MMC (P values of 0.0043 and 0.0322 respectively), indicating that this residue is essential in the functioning of Hel308.

Growth Rate

In order to visually compare the growth rates, H2078 (*hel308-E330G*), H2097 (*hel308-H317G*) and H2396 (*hel308-F316A*) were streaked onto complete media alongside H1392 (Δ*hel308*) and H26 (*hel308+*), Figure 5.10.

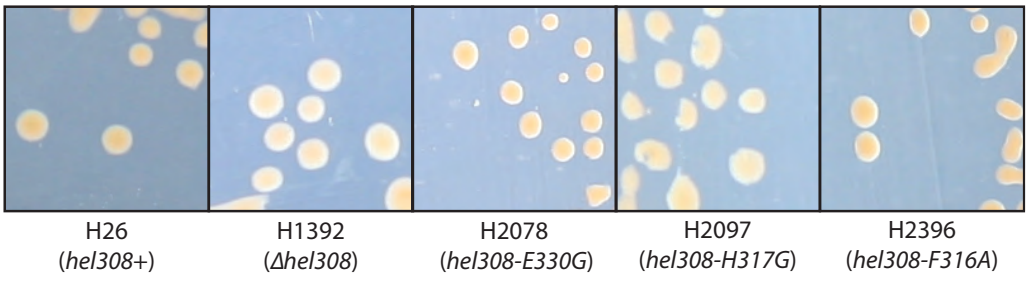


Figure 5.10: **Growth of *hel308-E330G, hel308-H317G* and *hel308-F316A* strains.** No significant difference in single colony size can be observed between H2078 (*hel308-E330G*), H2097 (*hel308-H317G*) and H2396 (*hel308-F316A*), H1392 (Δ*hel308*) and H26 (*hel308+*). All plate images are to the same scale.

H2097 (*hel308-H317G*) and H2396 (*hel308-F316A*) have growth no different to H1392 (Δ*hel308*) and wild type H26 (*hel308+*), where as H2078 (*hel308-E330G*), are slightly smaller than H1392 (Δ*hel308*) and wild type H26 (*hel308+*). A limitation of using growth on solid media is that only large variations can be readily observed and so the growth rate of H2078 (*hel308-E330G*), H2097 (*hel308-H317G*) and H2396 (*hel308-F316A*) was measured in liquid culture, Figure 5.11.

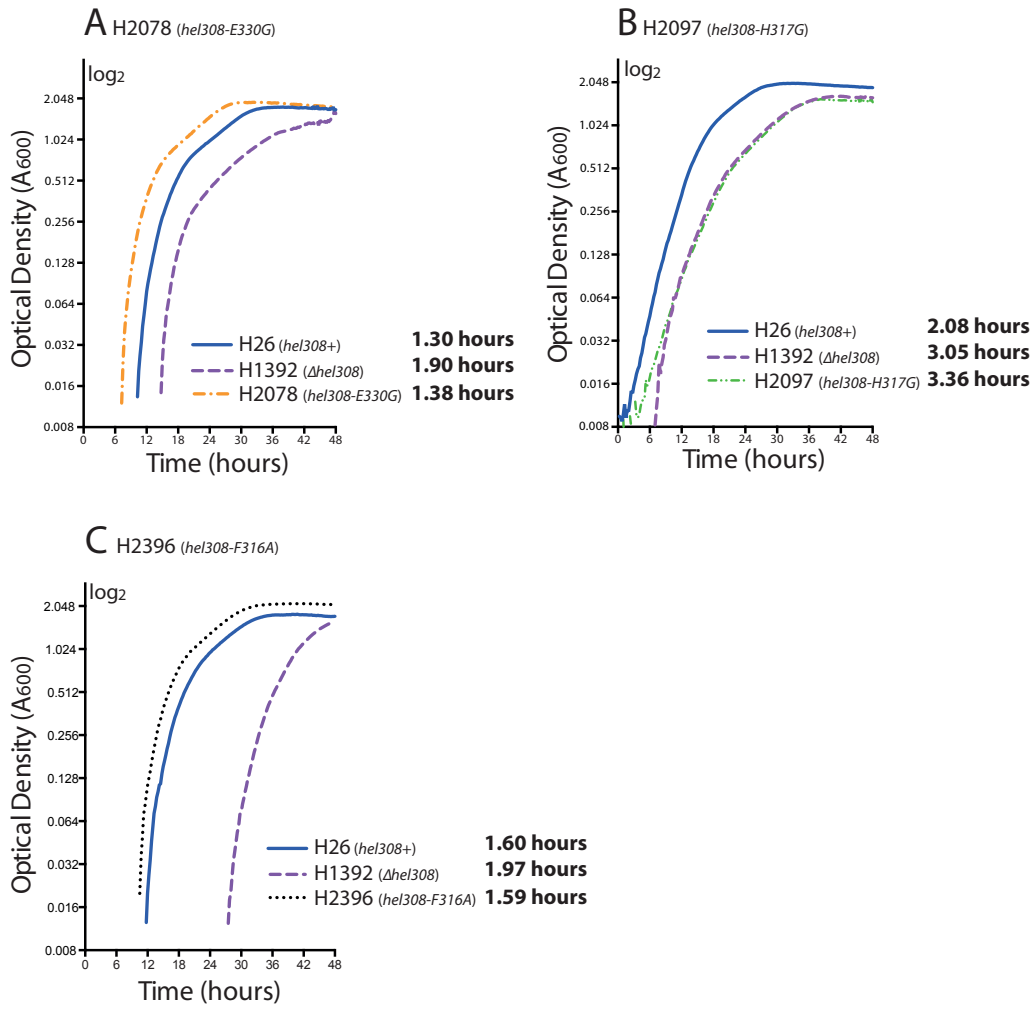


Figure 5.11: **Exponential growth rate of *hel308-E330G, hel308-H317G* and *hel308-F316A* strains.** Growth was measured by A_{600}. Generation time is indicated at the side of the strain name. **(A)** H2078 (*hel308-E330G*) has a generation time similar to wild type H26 (*hel308+*). **(B)** H2097 (*hel308-H317G*) has a generation time slower than that of H1392 (*Δhel308*). **(C)** H2396 (*hel308-F316A*) has a generation time similar to wild type H26 (*hel308+*). Graph plotted on a $\log_2$ scale. 3 repeats were carried out for each strain. Generation time calculated from linear growth phase. All strains and repeats were incubated on the same 48 well plate and measured simultaneously using an Epoch 2 Microplate Spectrophotometer (BioTek).

As discussed previously in Chapter 4: Genetic chatacterisation of *hel308*, the generation time of H26 (*hel308+*) and H1392 (*Δhel308*) is not consistent from experiment to experiment when measuring the growth using the Epoch 2 Microplate Spectrophotometer (BioTek). However, the relationship between these strains within a single experiment is always consistent; i.e. that H1392 (*Δhel308*) always has a generation time slower than H26 (*hel308+*). The generation times stated are not absolute. Therefore, in this study growth curves generated by this method are used to illustrate the differences in generation

times between a set of strains in the same experiment rather than an exact determination of each generation time.

H2078 (*hel308-E330G*) has a generation time comparable to wild type H26 (*hel308+*) of 1.38 hours and 1.30 hours respectively. Similarly H2396 (*hel308-F316A*) has a generation time comparable to wild type H26 (*hel308+*) of 1.59 hours and 1.60 hours respectively. This indicates that these residues are either not essential to the function of Hel308 or that these mutations do not cause significant changes in the protein to affect the function of Hel308. By contrast, H2097 (*hel308-H317G*) has a generation time slower than that of H1392 (*Δhel308*) of 3.36 hours and 3.05 hours respectively, indicating that this residue is essential to the function of Hel308.

Recombination Frequency

The improvement in growth and suvival after treatment can be seen in the H1844 (*Δhel308 ΔradB*) strain compared to the H1844 (*Δhel308 ΔradB*) (in Chapter 4: Genetic chatacterisation of *hel308*) suggests that Hel308 is implicated in homologous recombination. It is proposed that Hel308 could act as an anti-recombinase. Therefore it is of interest to analyse the recombination frequence of *hel308* point mutants, Table 5.1

Table 5.1: **Recombination frequencies of *hel308-E330G*, *hel308-H317G* and *hel308-F316A* strains.**

Strain	**H164** *hel308+*	**H2117** *Δhel308*	**H2397** *hel308-F316A*	**H2261** *hel308-H317G*	**H2263** *hel308-E330G*
Recombination Frequency (RF)	4.94×10^{-5} ($+/- 3.01\times10^{-5}$)	3.23×10^{-5} ($+/- 1.17\times10^{-5}$)	1.53×10^{-4} ($+/- 2.98\times10^{-2}$)	2.04×10^{-4} ($+/- 1.24\times10^{-4}$)	1.53×10^{-4} ($+/- 1.28\times10^{-4}$)
Transformation Efficiency (TE)	1.07×10^{-5} ($+/- 3.25\times10^{-6}$)	3.00×10^{-5} ($+/- 0.00$)	1.43×10^{-4} ($+/- 1.24\times10^{-4}$)	9.75×10^{-6} ($+/- 1.86\times10^{-6}$)	8.02×10^{-6} ($+/- 1.21\times10^{-6}$)
Relative recombination frequency (normalised by TE)	4.62×10^{0}	1.08×10^{0}	4.77×10^{-2}	2.09×10^{-1}	1.91×10^{-1}
	1×	**0.23×**	**103.27×**	**4.53×**	**4.14×**
Crossover fraction	13.49% (126)	8.75% (120)	0.00% (120)	17.5% (80)	11.25% (160)
Non-crossover fraction	86.51% (126)	91.25% (120)	100.00% (120)	82.50% (80)	88.75% (160)

Values in bold indicate the amount of recombination, crossover or non-crossover events compared to wild-type H164 (*hel308+*). Values are generated as an average of at least 3 independent trials, +/- standard error is shown in brackets. Cells are shaded blue to indicate recombination defect and red to indicate hyper-recombination. Fraction of crossover and non-crossover events represented as a percentage, cells are shaded where values differ significantly from the wild type (P =0.05), blue indicates a decrease, red indicates an increase. Number of colonies assayed for crossover and non-crossover is indicated in brackets underneeth the percentages.

H2397 (*hel308-F316A*) has an extremely hyper-recombinogenic phenotype with recombination levels 103.27× that of wild type H164 (*hel308+*). No crossover recombination events were observed for this strain and non-crossover events were significantly higher than wild type with two degrees of freedom with a chi-squared test). Both H2261 (*hel308-H317G)* and H2263 (*hel308-E330G)* show recombination levels around four times that of wild type, with recombination levels of 4.53× and 4.14× respectively. These results suggest that Hel308 and in particular domain 2 plays a significant role in the regulation of homologous recombination.

5.3.3 Domain 2-3 linker Mutations D420A and E422G

Domain 3 contains a winged helix (WH), which are commonly seen in nucleic acid binding proteins. Extensive DNA interactions occur across domain 3 and this domain is important for inter-domain communication during helicase unwinding. This WH domain may help maintain the structure of the enclosed Hel308 ring around ssDNA (Woodman & Bolt 2011). It is possible that

Hel308 may undergo some conformational changes during helicase unwinding and loading onto ssDNA. The conserved extended linker region that connects domains 2 and 3 could act as a hinge that allows communication between the RecA domains 1 and 2 and the ratchet in domain 4 (Richards et al 2008).

Mutations in this region could have profound effects on the movement and conformation of domains 1-4 relative to each other and therefore on helicase function as a whole. To study the effects of disruptions in this region the strains H2077 (*hel308-D420A*) and H2076 (*hel308-E422G*) were constructed, the mutations in both of these strains are found in the centre of the 2-3 linker region. The amino acid substitutions and the position of these mutations are detailed in Figure 5.12.

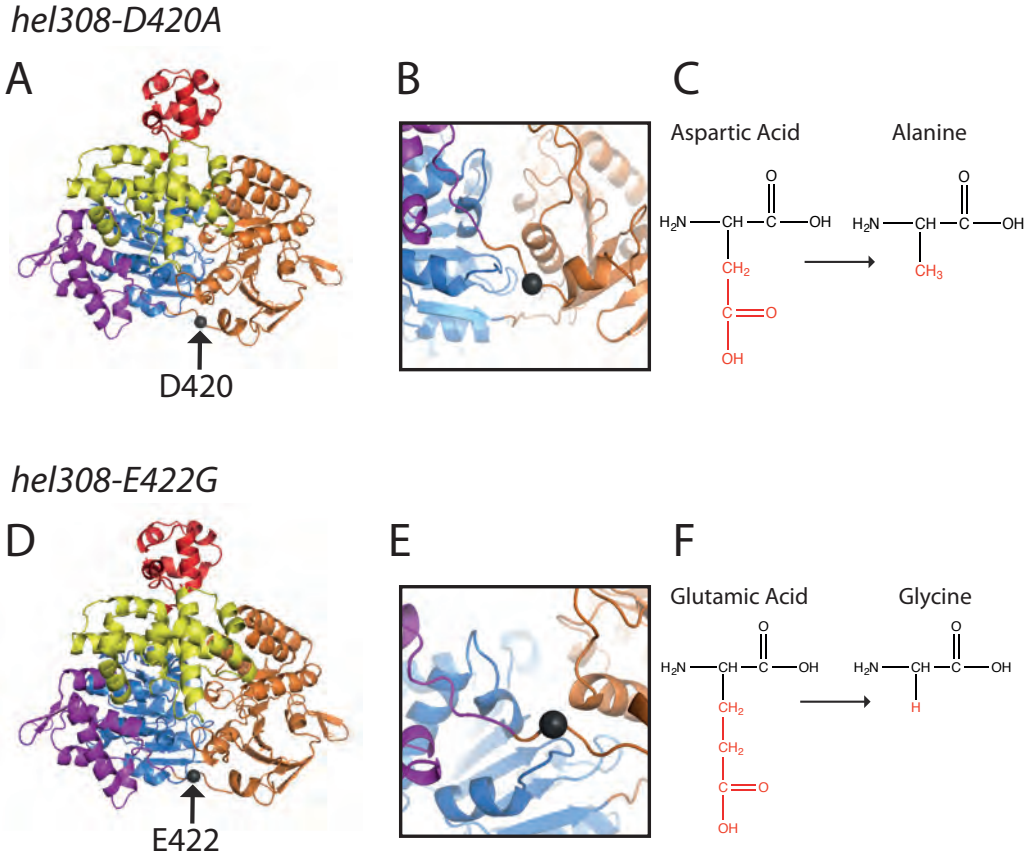


Figure 5.12: **Location of D420 and E422 residues in the structure of Hel308.** The structure of Hel308 from *H. volcanii* was predicted from the amino acid sequence using Phyre2 (Kelley et al 2015). **(A)** Predicted structure of Hel308, position of D420 is located in the centre of the domain 2-3 linker region, indicated with black sphere. **(B)** The position of D420 seen from 'underneath' Hel308. **(C)** Amino acid substitution of aspartic acid with alanine, R group is shown in red. **(D)** Predicted structure of Hel308, position of E422 is located in the centre of the domain 2-3 linker region, indicated with black sphere. **(E)** The position of E422 seen from 'underneath' Hel308. **(F)** Amino acid substitution of glutamic acid with glycine, R group is shown in red. Domain 1 (blue), domain 2 (orange), domain 3 (purple), domain 4 (yellow) and domain 5 (red).

In H2077 (*hel308-D420A*) the negatively charged aspartic acid (D) was substituted with the smaller hydrophobic amino acid alanine (A) and in H2076 (*hel308-E422G*) the negatively charged glutamic acid (E) was substituted with the small non-polar amino acid glycine (G). The loss of negative charge in this region could disturb inter-domain interactions. Furthermore, the substitution of D and E to smaller amino acids could alter the conformation of this linker region.

Survival Following Treatment with DNA-damaging Agents

In order to study the effects of the *hel308* domain 2/3 mutants, H2076 (*hel308-E422G*) and H2077 (*hel308-D420A*) on repair of DNA breaks and crosslinks, cells were treated with UV and MMC respectively, as described previously, Figure 5.13.

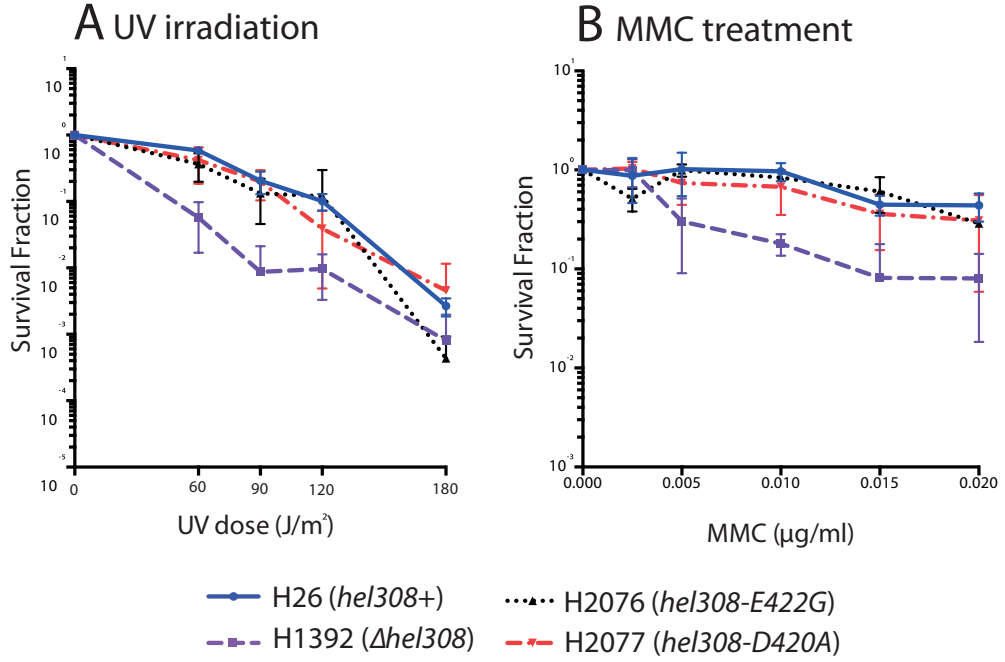


Figure 5.13: **Survival frequency of *hel308-E422G* and *hel308-D420A* following treatment with DNA-damaging agents. (A)** Survival following treatment with UV irradiation. No difference in survival of either strain compared to H1392 (Δ*hel308*) and H26 (*hel308*+). **(B)** Survival following treatment with MMC. No difference in survival of either strain compared to H26 (*hel308*+). Survival fraction is calculated relative to un-treated control. Each data point is generated as an average of at least 3 independent trials. Standard error is shown.

No significant difference in survival fractions was observed for H2076 (*hel308-E422G*) and H2077 (*hel308-D420A*) after treatment with UV irradiation or MMC compared to H26 (*hel308*+) (P-values > 0.05, calculated

from a two-tailed *t*-tests). Indicating that these residues are not essential to the function of Hel308 or that these mutations do not cause significant changes in the protein to affect the function of Hel308.

Growth Rate

In order to visually compare the growth rates, H2076 (*hel308-E422G*) and H2077 (*hel308-D420A*) were streaked onto complete media alongside H1392 (Δ*hel308*) and H26 (*hel308*+), Figure 5.14.

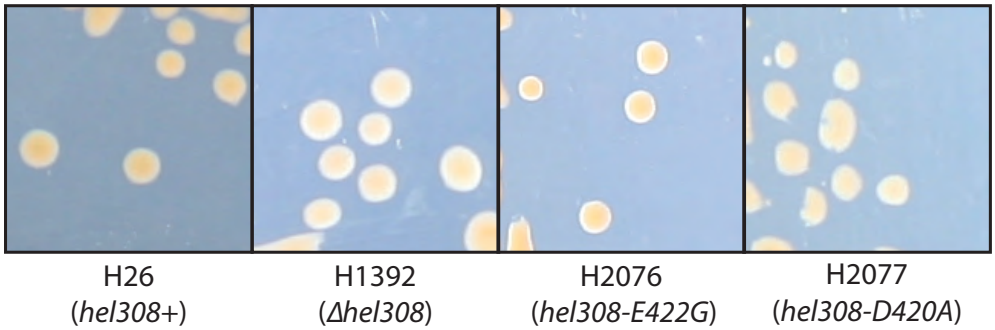


Figure 5.14: **Growth of *hel308-E422G* and *hel308-D420A* strains.** No significant difference in single colony size can be observed between *H2076 (hel308-E422G), H2077 (hel308-D420A),* H1392 (Δ*hel308*) and H26 (*hel308*+). All plate images are to the same scale.

H2076 (*hel308-E422G*) and H2077 (*hel308-D420A*) have growth no different to H1392 (Δ*hel308*) and wild type H26 (*hel308*+). A limitation of using growth on solid media is that only large variations can be readily observed and so the growth rates of these *hel308* mutants was measured in liquid culture, Figure 5.15.

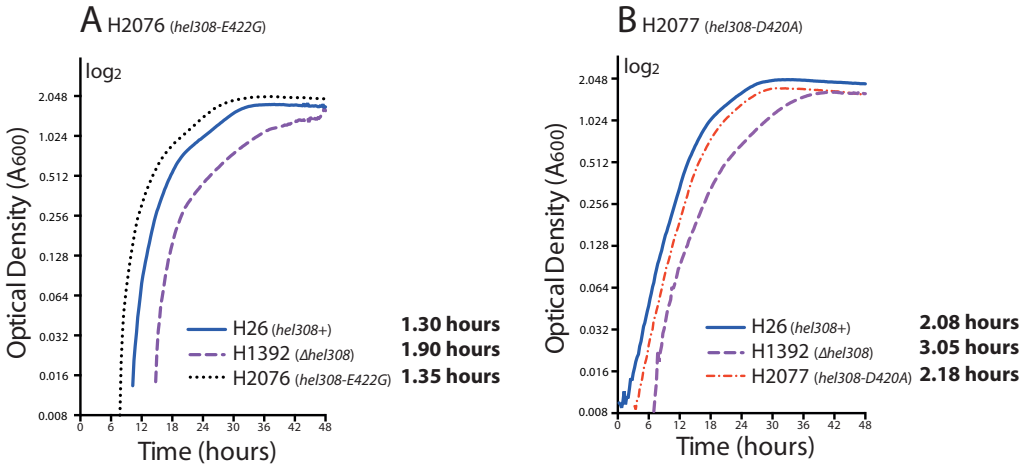


Figure 5.15: **Exponential growth rate of *hel308-E422G* and *hel308-D420A* strains.** Growth was measured by A_{600}. Generation time is indicated at the side of the strain name. **(A)** H2076 *(hel308-E422G)* has a generation time similar to wild type H26 (*hel308+*). **(B)** H2077 (*hel308-D420A*) has a generation time similar to wild type H26 (*hel308+*). Graph plotted on a $\log_2$ scale. 3 repeats were carried out for each strain. Generation time calculated from linear growth phase. All strains and repeats were incubated on the same 48 well plate and measured simultaneously using an Epoch 2 Microplate Spectrophotometer (BioTek).

H2076 (*hel308-E422G*) has growth similar to wild type H26 (*hel308+*) with generation times of 1.35 hours and 1.30 hours respectively. Likewise, the generation time of H2077 (*hel308-D420A*) is comparable to wild type of 2.18 hours and 2.08 hours respectively. Indicating that these residues are not essential to the function of Hel308 or that these mutations do not cause significant changes in the protein to affect the function of Hel308.

Recombination Frequency

Hel308 is implicated in homologous recombination and is proposed that Hel308 could even act as an anti-recombinase. Therefore it is of interest to analyse the recombination frequence of *hel308* point mutants, Table 5.2.

Table 5.2: **Recombination frequencies of *hel308-E422G* and *hel308-D420A* strains.**

Strain	**H164** *hel308+*	**H2117** *Δhel308*	**H2259** *hel308-D420A*	**H2257** *hel308-E422G*
Recombination Frequency (RF)	4.94×10^{-5} ($+/- 3.01\times10^{-5}$)	3.23×10^{-5} ($+/- 1.17\times10^{-5}$)	4.95×10^{-5} ($+/- 4.26\times10^{-5}$)	6.48×10^{-5} ($+/- 2.73\times10^{-5}$)
Transformation Efficiency (TE)	1.07×10^{-5} ($+/- 3.25\times10^{-6}$)	3.00×10^{-5} ($+/- 0.00$)	1.93×10^{-5} ($+/- 1.28\times10^{-5}$)	6.25×10^{-6} ($+/- 1.12\times10^{-6}$)
Relative recombination frequency (normalised by TE)	4.62×10^{0}	1.08×10^{0}	2.57×10^{0}	1.04×10^{-1}
	1×	**0.23×**	**0.55×**	**2.24×**
Crossover fraction	13.49% (126)	8.75% (120)	8.70% (115)	8.75% (160)
Non-crossover fraction	86.51% (126)	91.25% (120)	91.30% (115)	91.25% (160)

Values in bold indicate the amount of recombination, crossover or non-crossover events compared to wild-type H164 (*hel308+*). Values are generated as an average of at least 3 independent trials, +/- standard error is shown in brackets. Cells are shaded blue to indicate recombination defect and red to indicate hyper-recombination. Fraction of crossover and non-crossover events represented as a percentage, cells are shaded where values differ significantly from the wild type (P =0.05), blue indicates a decrease, red indicates an increase. Number of colonies assayed for crossover and non-crossover is indicated in brackets underneeth the percentages.

H2259 (*hel308- D420A)* displays a reduced recombination frequency of 0.55× that of wild type H164 (*hel308+*). H2257 (*hel308- E422G)* has a moderate hyper-recombinogenic phenotype with recombination frequency 2.24× that of wild type. For both mutations the crossover and non-crossover fractions are not significantly different to wildtype (with two degrees of freedom with a chi-squared test).

5.3.4 Domain 5 mutation R743A

Domain 5 is an auto-inhibitory domain that couples ATP hydrolysis to DNA unwinding. After emerging from the central pore of Hel308 the 3' tail of unwound DNA bends around domain 4 and binds a helix-link-helix (HLH) structure in domain 5 via the phosphate-sugar backbone (Buttner et al 2007). This interaction could be resposible for Hel308 recognising replication forks. Branched DNA structures are the prefered substrates for Hel308 *in vitro* (Fujikane et al 2005, Guy & Bolt 2005).

Archaeal Hel308 helicases contain a highly conserved triplet of arginine residues, RXRAR (where X can be any residue). In the *A. fulgidus* Hel308-DNA co-crystal the central arginine of this motif (Arg644) contacts the phosphate backbone of the ssDNA emerging from the central pore of the helicase. The other two arginine residues (Arg642 and Arg646) form contacts between domain V and the helicase ratchet domain IV (Woodman et al 2007).

In *M. thermautotrophicus* Hel308, the mutation of the first arginine in the RXRAR motif to an alanine (R647A) resulted in increased ATPase activity, implying that this residue possibly acts as part of a braking mechanism to stop ATP hyrolysis until DNA is engaged. Additionally, R647A and mutation of the final arginine to an alanine (R651A) abolishes helicase actvity *in vitro*. A mutation of the central arginine to an alanine (R649A) resulted in DNA binding defects and reduced DNA-stimulated ATP hydrolysis (Woodman et al 2011, Woodman et al 2007).

However, in *S. solfataricus* the mutation of the central arginine to an alanine (R662A) or the deletion of domain 5 altogether resulted in ssDNA binding affinities comparable with or only slightly weaker than the wild-type enzyme, but an increase of helicase activity (Richards et al 2008).

Mutations in the RXRAR motif have profound effects on DNA binding, ATPase and helicase activity of Hel308. To study the genetic effects of a mutation in this region, the strain H2398 (*hel308-R743A*) was constructed in *H. volcanii*. In this strain the central arginine (R) in the RXRAR motif was substituded to an alanine (A), this mutation results in a loss of positive charge which will interfere with the binding of DNA. The amino acid substitution and the position of this mutation are detailed in Figure 5.16.

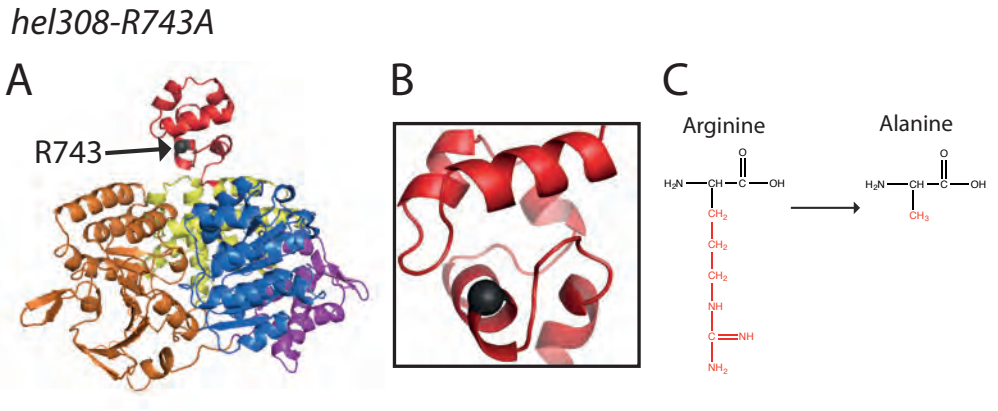


Figure 5.16: **Location of R743 in the structure of Hel308.** The structure of Hel308 from *H. volcanii* was predicted from the amino acid sequence using Phyre2 (Kelley et al 2015). **(A)** Predicted structure of Hel308, R743 is located at the base of domain 5 in the region where the 3' tail of unwound DNA emerges from the central pore of Hel308, indicated with black sphere. **(B)** View of R743 from 'underneath' from the angle at which ssDNA emerges from Hel308. **(C)** Amino acid substitution of arginine with alanine, R group is shown in red. Domain 1 (blue), domain 2 (orange), domain 3 (purple), domain 4 (yellow) and domain 5 (red).

Survival Following Treatment with DNA-damaging Agents

In order to study the effects of a *hel308* domain 5 mutant, H2398 (*hel308-R743A*) on repair of DNA breaks and crosslinks, cells were treated with UV and MMC respectively, as described previously, Figure 5.17.

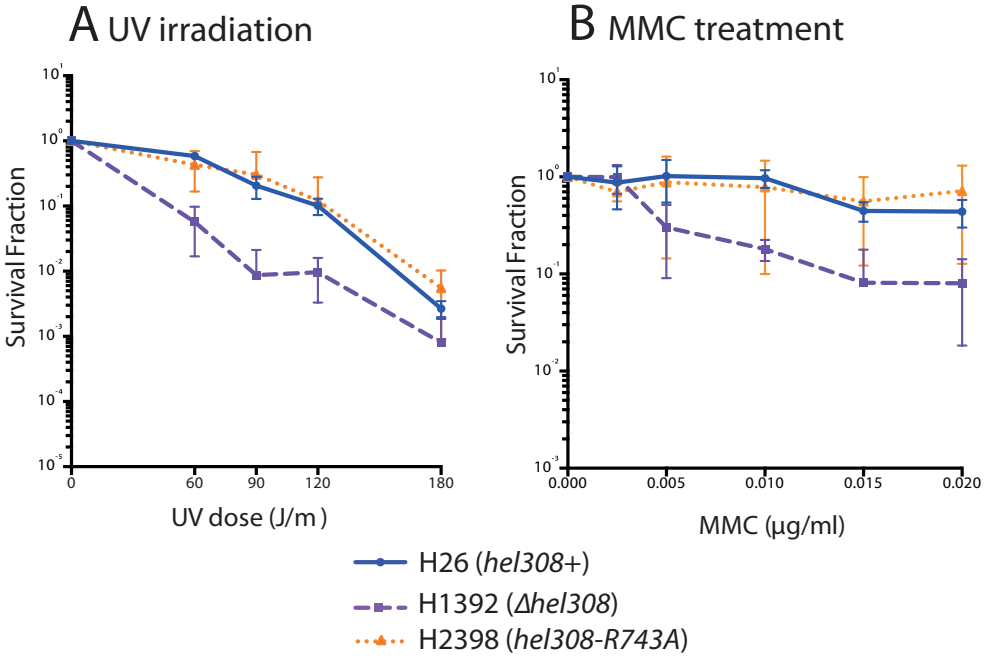


Figure 5.17: **Survival frequency of *hel308-R743A* following treatment with DNA-damaging agents. (A)** Survival following treatment with UV irradiation. No difference in survival of H2398 (*hel308-R743A*) compared to H1392 (Δ*hel308*) and H26 (*hel308+*). **(B)** Survival following treatment with MMC. No difference in survival of H2398 (*hel308-R743A*) compared to H26 (*hel308+*). Survival fraction is calculated relative to un-treated control. Each data point is generated as an average of at least 3 independent trials. Standard error is shown.

No significant difference in survival fraction was observed for H2398 (*hel308-R743A*) after treatment with UV irradiation or MMC compared to H26 (*hel308+*) (P-values > 0.05, calculated from a two-tailed *t*-tests). Indicating that these residues are not essential to the function of Hel308 or that these mutations do not cause significant changes in the protein to affect the function of Hel308.

Growth Rate

In order to visually compare the growth rates, H2398 (*hel308-R743A*) was streaked onto complete media alongside H1392 (Δ*hel308*) and H26 (*hel308+*), Figure 5.18.

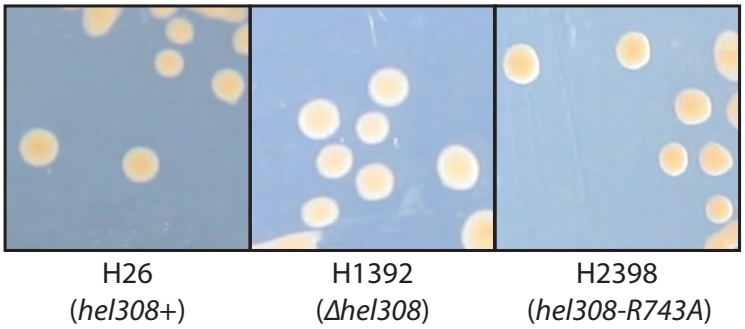


Figure 5.18: **Growth of a *hel308-R743A* strain.** No significant difference in single colony size can be observed between of H2398 *(hel308-R743A),* H1392 (Δ*hel308*) and H26 (*hel308+*). All plate images are to the same scale.

H2398 (*hel308-R743A*) has growth no different to H1392 (Δ*hel308*) and wild type H26 (*hel308+*). A limitation of using growth on solid media is that only large variations can be readily observed and so the growth rates of these *hel308* mutants was measured in liquid culture, Figure 5.19.

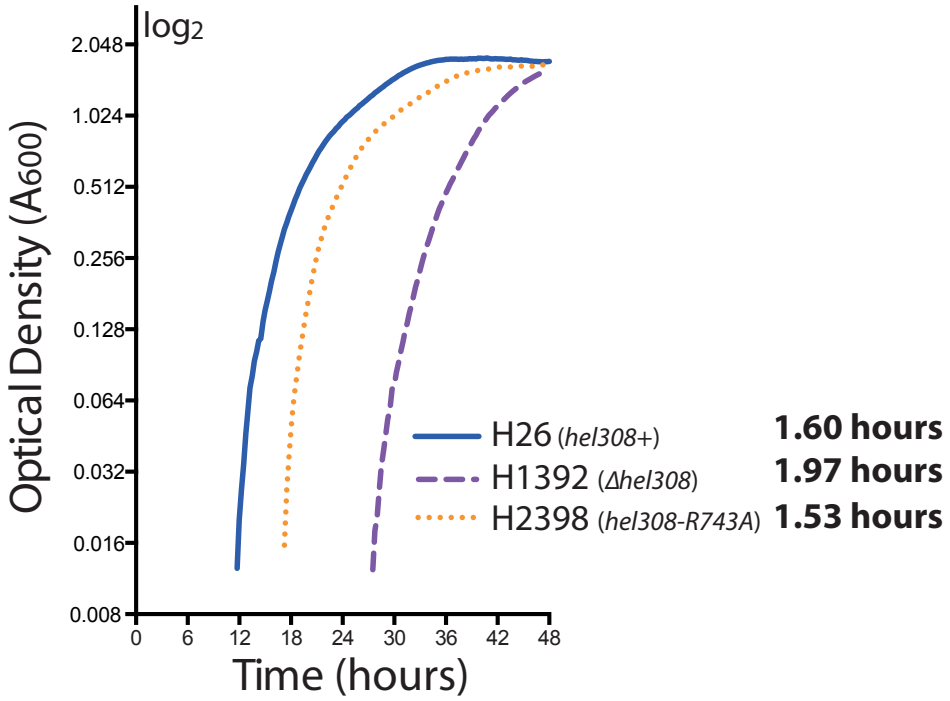


Figure 5.19: **Exponential growth rate of a *hel308-R743A* strain.** Growth was measured by A_{600}. Generation time is indicated at the side of the strain name. H2398 *(hel308-R743A)* has a generation time similar to wild type H26 (*hel308+*). Graph plotted on a $\log_2$ scale. 3 repeats were carried out for each strain. Generation time calculated from linear growth phase. All strains and repeats were incubated on the same 48 well plate and measured simultaneously using an Epoch 2 Microplate Spectrophotometer (BioTek).

H2398 (*hel308-R743A*) has growth similar to wild type H26 (*hel308+*) with generation times of 1.53 hours and 1.60 hours respectively. Indicating that this residue is not essential to the function of Hel308 or that the mutation does not cause significant changes in the protein to affect the function of Hel308.

Recombination Frequency

Hel308 is implicated in homologous recombination and is proposed that Hel308 could even act as an anti-recombinase. Therefore it is of interest to analyse the recombination frequence of *hel308* point mutants, Table 5.3.

Table 5.3: **Recombination frequency of a *hel308-R743A* strain.**

Strain	H164	H2117	H2398
	hel308+	*Δhel308*	*hel308-R743A*
Recombination Frequency (RF)	4.94×10^{-5} ($+/- 3.01\times10^{-5}$)	3.23×10^{-5} ($+/- 1.17\times10^{-5}$)	3.26×10^{-1} ($+/- 2.06\times10^{-1}$)
Transformation Efficiency (TE)	1.07×10^{-5} ($+/- 3.25\times10^{-6}$)	3.00×10^{-5} ($+/- 0.00$)	1.06×10^{-3} ($+/- 8.50\times10^{-4}$)
Relative recombination frequency (normalised by TE)	4.62×10^{0}	1.08×10^{0}	3.08×10^{-2}
	1×	**0.23×**	**66.64×**
Crossover fraction	13.49% (126)	8.75% (120)	0.00% (120)
Non-crossover fraction	86.51% (126)	91.25% (120)	100.00% (120)

Values in bold indicate the amount of recombination, crossover or non-crossover events compared to wild-type H164 (*hel308+*). Values are generated as an average of at least 3 independent trials, +/- standard error is shown in brackets. Cells are shaded blue to indicate recombination defect and red to indicate hyper-recombination. Fraction of crossover and non-crossover events represented as a percentage, cells are shaded where values differ significantly from the wild type (P =0.05), blue indicates a decrease, red indicates an increase. Number of colonies assayed for crossover and non-crossover is indicated in brackets underneeth the percentages.

H2398 (*hel308-R743A*) has an extremely hyper-recombinogenic phenotype with recombination levels 66.64× that of wild type H164 (*hel308+*). No crossover recombination events were observed for this strain and non-crossover events were significantly higher than wild type with two degrees of freedom with a chi-squared test. This result suggests that Hel308 and in particular domain 5 plays a significant role in the regulation of homologous recombination.

5.3.5 H1391 (*Δhel308*) and H1392 (*Δhel308*) Comparison

In this chapter the *Δhel308* strain H1392 was used as a control, whereas in Chapter 4: Genetic chatacterisation of *hel308*, the *Δhel308* strain used was H1391. H1391 and H1392 are two different clones from the same pop-out of pTA1276 (*Δhel308 pyrE2+)* from *hel308* locus in the parental strain H26. To ensure that these strains were phenotypically similar, the survival following treatment with DNA damaging agents and growth rates were compared, Figure 5.20.

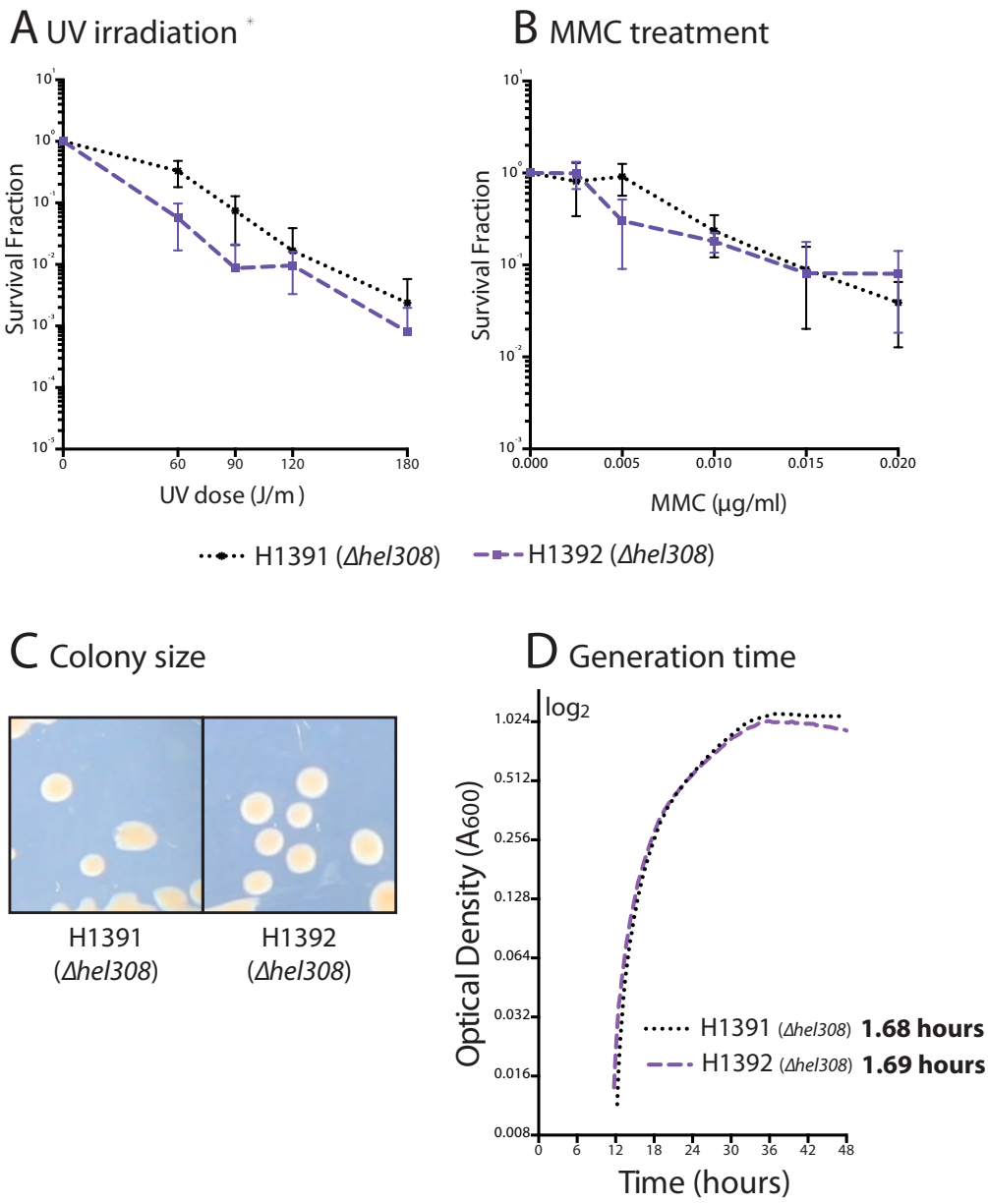


Figure 5.20: **Survival frequency following treatment with DNA-damaging agents and growth rates of H1391 (Δ*hel308*) and H1392 (Δ*hel308*). (A)** Survival following treatment with UV irradiation. No difference in survival between H1391 (Δ*hel308*) and H1392 (Δ*hel308*). Survival fraction is calculated relative to un-treated control. Each data point is generated as an average of at least 3 independent trials. Standard error is shown. **(B)** Survival following treatment with MMC. No difference in survival between H1391 (Δ*hel308*) and H1392 (Δ*hel308*). **(C)** No difference in colony sizes. Both images are to the same scale. **(D)** Growth was measured by A_{600}. Generation time is indicated at the side of the strain name. H1391 (Δ*hel308*) and H1392 (Δ*hel308*) have similar generation times. Graph plotted on a $\log_2$ scale. 3 repeats were carried out for each strain. Generation time calculated from linear growth phase. All strains and repeats were incubated on the same 48 well plate and measured simultaneously using an Epoch 2 Microplate Spectrophotometer (BioTek).

In all assays shown, H1391 (Δ*hel308*) and H1392 (Δ*hel308*) are phenotypically near-identical. No significant difference in survival frequencies following UV irradiation and MMC are observed (P-values > 0.05, calculated from a two-

tailed *t*-tests) and both strains have near identical generation times of 1.68 hours and 1.69 hours for H1391 and H1392 respectively. This demonstrates that strains in this chapter that are derived from H1392 (Δ*hel308*) and strains derived from H1391 (Δ*hel308*) in Chapter 4: Genetic chatacterisation of *hel308*, can be directly compared.

5.4 Discussion

A summary of the DNA damage, growth and recombination phenotypes of all Hel308 point mutations covered in this chapter is given in Table 5.4. Δ*hel308* is included for reference.

Table 5.4: **Summary of phenotypes of Hel308 point mutants**

Mutation	UV	MMC	Generation time	Recombination frequency	CO	NCO
Δ*hel308*	WT	↓ WT	↓ WT	↓ WT	↓ WT	WT
K53R (1)	↓ WT	↓ WT (=Δ*hel308*)	↓ WT (=Δ*hel308*)	-	-	-
D145N (1)	↓ WT	↓ WT (=Δ*hel308*)	↓ WT (=Δ*hel308*)	-	-	-
F316A (2)	WT	WT	WT	↑↑↑ WT	↓↓ WT (0%)	↑↑ WT (100%)
H317G (2)	↓ WT	↓ Δ*hel308*	↓ Δ*hel308*	↑ WT	WT	WT
E330G (2)	WT	WT	WT	↑ WT	WT	WT
D420A (2-3 linker)	WT	WT	WT	↓ WT	WT	WT
E422G (2-3 linker)	WT	WT	WT	↑ WT	↓ WT	WT
R743A (5)	WT	WT	WT	↑↑↑ WT	↓↓ WT (0%)	↑↑WT (100%)

Cells are shaded blue to indicate a defect in survival following DNA damage, generation time or recombination. Defects equivalent to or more severe than Δ*hel308* strains are indicated. Cells shaded red indicate hyper-recombination. Domain location of each mutant is indicated in brackets.

K53R and D145N

The K53R and D145N mutations lie in the Walker A and Walker B motifs that are responsible for the binding and hydrolysis of ATP, respectively. Mutations in these motifs result in helicase-defective proteins, and this study found that survival rates after treatment with the interstrand crosslinking agent MMC and generation times are similar to those of a Δ*hel308* strain. These are expected phenotypes for these mutations and have been documented for Hel308 proteins from *Sulfolobus solfataricus* and *Methanothermobacter thermautotrophicus* (Richards et al 2008, Woodman et al 2011). However, an interesting result was the reduction in survival following treatment with UV irradiation, since

Δhel308 does not share this phenotype and Hel308 is not known to be involved in the repair of UV-induced lesions such as double-strand DNA breaks.

Initial DNA binding of Hel308 is ATP-independent and Hel308 is able to melt 2 bp of DNA in the absence of ATP but then requires ATP for further translocation (Buttner et al 2007, Woodman et al 2007). Therefore, the UV sensitivity phenotype could be due to the Walker A and B mutants being able to bind the DNA in an ATP-independent fashion, and then requiring the binding and hydrolysis of ATP to translocate or release from DNA. These mutants, unable to act as an ATPase, may become 'locked' onto the ssDNA at sites of double strand breaks (DSBs) and interfere with DSB repair, thus resulting in a reduced survival fraction.

Due to time constraints, recombination assays were not performed with these mutants. However, it would be of interest to measure the levels of recombination in these strains since other helicase-defective Hel308 mutations such as *Δhel308* and *hel308- H317G* display altered recombination levels.

F316A

The F316A mutation resides in domain 2 and faces towards the inner channel of Hel308, through which the 3' tail of unwound DNA passes (Woodman & Bolt 2011). This mutation had no effect on the generation time or the survival rate of the mutant strain following treatment with UV and MMC.

However, a striking increase of recombination frequency of around 100 times that of a wild type strain was observed. This suggests that this residue is involved in the downregulation of recombination and upon mutation this regulatory role of Hel308 has been impaired. The mutation of phenylalanine (F) to alanine (A) results in a change from a large hydrophobic residue to a smaller residue that less hydrophobic. Phenylalanine can interact with other aromatic side chains, and could possibly play a role in inter-domain interactions and substrate binding within Hel308 (Betts & Russell 2003).

Mutation to alanine may have perturbed key interactions within Hel308 that allow for correct function in the regulation of homologous recombination.

A second striking result is that the entirety of recombination events assayed in this strain result in only non-crossover (NCO) products (where genetic exchange has not occurred). This suggests that in this strain, homologous recombination is carried out only by the synthesis dependent strand-annealing (SDSA) pathway, which results only in non-crossover products, pink box Figure 5.21 (San Filippo et al 2008). If homologous recombination were to follow other pathways such as double strand break repair (DSBR), where crossovers account for 50% of the outcomes, the proportion of crossover products would be higher. During the initiation of SDSA, the D-loop that is generated during synapsis and extended by DNA polymerases can be reversed by the action of RecQ helicases such as Srs2 in yeast and RecQ5, BLM and FANCJ in mammals (Bugreev et al 2007, Hu et al 2007, Ira et al 2003, Sommers et al 2009). Therefore, RecQ helicases are known to promote SDSA and are regarded as anti-recombinases. The recombination data generated from this mutation suggests that not only can Hel308 regulate the level of recombination, but also plays a role in which recombination pathway is used.

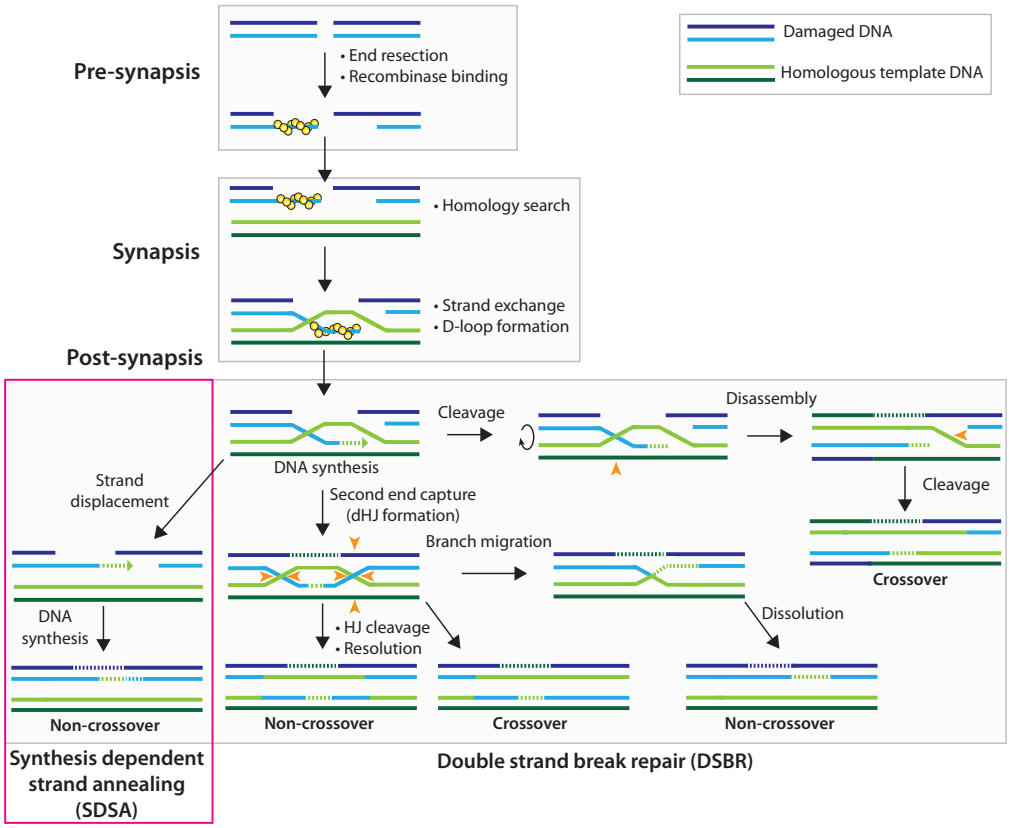


Figure 5.21: **Homologous recombination pathways.** A schematic of homologous recombination using the repair of double strand breaks as an example. Recombination occurs in three stages: pre-synapsis, synapsis and post-synapsis. Synthesis dependent strand annealing (SDSA) results in only non-crossover products (pink box) whereas double strand break repair (DSBR) results in both crossover and non-crossover products. Yellow spheres indicate recombinases, dashed lines indicate newly synthesised DNA and orange arrowheads indicate cleavage events. SDSA results

H317G

The H317G mutation also resides in domain 2 and faces towards the inner channel of Hel308, through which the 3' tail of unwound DNA passess (Woodman & Bolt 2011). H317 is the first histidine in the highly conserved IVa motif (HHAGL), which interacts with the negativly-charged DNA backbone (Buttner et al 2007). This mutation results in a defect in survival following treatment with UV and MMC, as well as a growth defect that is slightly worse than that seen with a *Δhel308* strain. These results suggest that this residue is essential for the correct DNA binding of Hel308 and its mutation has led to a loss in Hel308 function. An explanation for the growth defect and the survival following treatment with DNA damaging agents being worse than

a *Δhel308* strain could be that this mutation is causing Hel308 to persist on the DNA substrate and thereby block other repair pathways.

In this strain, recombination levels are elevated to around 4.5 times that of wildtype cells. This suggests that the usual activity of Hel308 is to limit the level of homologous recombination within the cell, and that DNA contacts made by the IVa motif is important for this activity.

E330G

In *A. fulgidus*, the equivalent residue to E330 is located in α-helix between the DNA binding motifs IVa and IVb (Buttner et al 2007). This residue faces away from the central pore of the helicase and towards the 'back' of the protein. The E330G mutation has no effect on growth or the survival rate of the strain following treatment with UV and MMC. This mutant exhibited an increase in recombination levels to around 4.1 times that of a wild type strain. Glutamic acid (E) is negatively charged and more rigid than glycine (G), which is small and non-polar (Betts & Russell 2003). The change in charge could have interfered with inter-domain interactions within Hel308. Furthermore, glycine is a small amino acid and so could have introduced conformational flexibility within the domain, leading to the increased recombination frequency of the strain.

D420A

The D420A mutation is found in the extended linker region that connects domains 2 and 3, and is thought to act as a hinge to allow movement of the domains 1-4 within Hel308 (Buttner et al 2007, Richards et al 2008). This mutation has no effect on growth or the survival rate of the strain following treatment with UV and MMC. However, a reduction in recombination frequecy was seen to around 50% of that observed in wild type strains. Aspartic acid (D) is a negatively charged amino acid that is able to form salt bridges with other amino acids (Betts & Russell 2003). As well as a loss of charge, the mutation of aspartic acid to the small amino acid alanine (A) could cause a change in the

conformation of the linker region, which in turn could disrupt domain interactions in Hel308. This mutation suggests that the correct conformation of Hel308 is necessary for the proper regulation of recombination.

E422G

The E422G mutation is also found in the domain 2-3 linker region of Hel308. Again, this mutation has no effect on growth or the survival rate of the strain following treatment with UV and MMC, but an effect was seen on the level of recombination. This mutation stimulated an increase in recombination to around 2 times that found in wild type strains. The mutation from a glutamic acid (E) to glycine (G) would potentially increase the flexibility of this linker region and therefore Hel308 as a whole, since glycine has the smallest R group of all the amino acids consisting of only a hydrogen atom. The added flexibility in this region appears to be promoting recombination, again suggesting that the conformation of Hel308 is important in the regulation of homologous recombination.

R743A

Mutations within domain 5 and in particular the RXRAR motif are well documented to interfere with the ATPase and helicase activity of Hel308 (Richards et al 2008, Woodman et al 2007). However the mutation R743A, which is located in the central arginine of the RXRAR motif, has no effect on growth or the survival rate of the strain following treatment with UV and MMC. This is surpising as in *M. thermautotrophicus*, mutation of the central arginine within the RXRAR motif to an alanine resulted in DNA binding defects and reduced DNA-stimulated ATP hydrolysis (Woodman et al 2007). Based on this observation, a defect in helicase ability would be expected in the *Haloferax volcanii* Hel308-R743A mutant, but the MMC and growth phenotypes suggest otherwise. Perhaps, the effect of this mutation is more similar to results seen in *S. solfataricus*, where the mutation of the central arginine to an alanine resulted in ssDNA binding affinities comparable to the wild-type enzyme, as well as an increase in helicase activity (Richards et al

2008). A similar increase in helicase activity of the *Haloferax volcanii* enzyme could account for the remarkable increase in recombination frequency to around 67 times that of a wild type strain. Similar to the F316A mutation, the dramatic increase in recombination levels is partnered with only non-crossover (NCO) products being detected. As with the F316A mutation, this could suggest that the R743A mutation is leading to preferential use of the SDSA pathway, pink box Figure 5.21, again suggesting that not only is Hel308 involved in regulating the levels of homologous recombination in the cell but also in which pathway is used.

Summary

From this study, three residues have been shown to be essential for the function of Hel308 as a helicase. These are K53, D145 and H317, which are located in the Walker A, Walker B and IVa motifs respectively; mutations of these residues lead to DNA damage and growth phenotypes comparable to those of a *Δhel308* strain.

The most striking observations in this study are the dramatic increase in recombination levels from mutations in residues F316 and R743. The increased recombination exclusively results in non-crossover products, suggesting that excess recombination is being routed down the SDSA pathway only. These observations suggest that not only can Hel308 modulate the levels of recombination within the cell, but has some control over which pathway of homologous recombination is used.

It is notable that mutations that showed no change in the growth rate or response to DNA damaging agents (F316A, E330G, D420A, E422G and R743A) showed perturbations in the levels of recombination. This is an unexpected observation and suggests that Hel308 could be playing two roles in the cell. The first role of Hel308 is as a helicase that unwinds replication fork-like structures at sites of DNA damage. Hel308 is able to unwind lagging strand-like structures from replication fork substrates *in vitro* (Guy & Bolt 2005). Hel308 also co-localises at stalled replication forks and interacts with

replication fork proteins such as RPA, FANCD2 and Rad51 (Tafel et al 2011). This role is largely unaffected by the mutational changes to Hel308 made in this study, since growth and survival after treatment with DNA damaging agents are at wild type levels, and suggests that the helicase activity of Hel308 is robust. The second role of Hel308 is specific to the regulation of homologous recombination and this role is dependent on the correct conformation of the protein. Mutations that potentially change the structure of Hel308 resulted in distinct changes in the levels of homologous recombination.

5.5 Future Perspectives

Further point mutants could be made that would aid in fully understanding the function of Hel308 in the cell.

Due to the striking increase in recombination upon mutation of the central arginine residue of the RXRAR motif (R743A), it would be of interest to mutate the other key arginine residues within this motif (for example, R745A and R747A). The genetic analysis of these point mutants would complement the biochemical data from *S. solfataricus* and *M. thermautotrophicus,* together giving a robust insight into the ATPase and helicase functions of Hel308. Again, due to the striking recombination phenotype seen with the F316A mutation, it would be interesting to generate further mutations in this region of Hel308, which would pinpoint the specific region of domain 2 that is involved in the regulation of homologous recombination.

Due to time constraints, it was not possible to generate point mutants in domain 4. Mutations in this domain would give insight into the function of the winged helix (WH) of Hel308. The WH domain in Hel308 is non-canonical and in *M. thermautotrophicus,* mutagenesis of the conserved aromatic residues in the WH domain have a profound effect on helicase unwinding; interestingly, mutations in the recognition helix have little effect (Woodman & Bolt 2011). It would be of interest to carry out genetic analysis of corresponding mutations in Hel308 from *H. volcanii* (for example, E494G and Q497G). Furthermore, biochemical analysis such as ATPase and helicase unwinding assays in *M.*

thermautotrophicus using mutations in Hel308 that are equivalent to those generated during this study may give further insights into the precise mechanistic action of Hel308.

For strains that display significant increases or decreases in levels of recombination, it would be interesting to couple these mutations with a *radB* deletion. In *H. volcanii*, a low level of recombination still occurs within *radB* deleted strains, at approximately 5% of wild-type levels (Haldenby 2007). Therefore, introducing Hel308 point mutations in this background could indicate whether Hel308 is interacting directly with RadB to regulate homologous recombination or is acting by other means.

The Hel308 point mutants that were seen to elicit striking increases of recombination showed little or no effect on cell growth. Perhaps the differences in generation time are too subtle to be detected using the spectrophotometric growth assay employed in this study. Pairwise growth competition assays as described by Delmas and colleagues use the *bgaHa* marker to monitor the relative growth of pairs of strains in exponential phase over the course of up to 8 days, allowing for the detection of very small differences in growth (Delmas et al 2009). It would be interesting to see if any growth defects could be observed by this method and if they correlate with other phenotypes.

Lastly, employing protein pull-down assays to investigate protein:protein interactions of Hel308 point mutants in *H. volcanii* could give insights and perhaps a mechanism of how Hel308 interacts with other proteins, in particular those involved in homologous recombination.

5.6 Conclusions

The mutations generated in this study suggest a multifaceted role for Hel308 in the repair of DNA damage and the regulation of homologous recombination. This insight requires further investigation to define the function of Hel308 in the cell.

Chapter 6: *in vitro* Analysis of Hel308

6.1 Background

6.1.1 Halophilic proteins

There are two alternative strategies employed by halophilic organisms for maintaining an osmotic balance between intracellular and extracellular salt concentrations. The first is a 'salt out' strategy used predominantly by halophilic bacteria and eukaryotes, where salts are actively pumped out from the cell and the cytoplasm is packed with organic solutes such as glycerol or glycine betaine to maintain the osmotic balance (Christian & Waltho 1962, Oren 1999, Oren 2008). Halophilic archaea and a few bacteria maintain osmotic balance by accumulating high levels of salt in the cytoplasm, this is termed a 'salt-in' approach (Oren et al 2002).

Proteins in halophilic archaea have adapted to function in high salt and low water conditions by several different strategies (Mevarech et al 2000). Halophilic proteins tend to feature a large number of acidic residues on their surface such as aspartic acid and glutamic acid, but have a small number of nonpolar residues. This generates an overall low isoelectric point (pI) and high density of negative charges on the surface of the protein that will co-ordinate a network of hydrated cations, allowing the protein to stay soluble in solution (Lanyi 1974). Halophilic proteins have a reduced surface hydrophobicity by replacing large hydrophobic side groups with small hydrophilic ones. Some halophilic proteins have extra domains or peptide insertions that are extremely rich in acidic residues, which are essential for correct protein folding (Graziano & Merlino 2014).

6.1.2 Protein purification

To maintain activity and structural stability, halophilic proteins require high concentrations of salt and low water availability. Therefore, expression of

halophilic proteins in mesophilic hosts such as *E. coli* can be problematic because halophilic proteins tend to mis-fold and aggregate in low ionic conditions (Ishibashi et al 2003, Jaenicke 2000).

Halophilic proteins expressed in *E. coli* are often insoluble and require recovery from inclusion bodies, followed by denaturation and refolding in hypersaline solutions. For example, Connaris and colleagues expressed *H. volcanii* dihydrolipoamide dehydrogenase (a flavoprotein enzyme involved in energy metabolism) in *E. coli* as inclusion bodies and purified it using copper metal ion affinity chromatography in the presence of 2 M KCl. The protein was refolded by solubilisation in 8 M urea followed by dilution into a buffer containing 2 M KCl, maximal activity was obtained after 3 days incubation at 4ºC (Connaris et al 1999). Protein recovery is only reliable with well-characterised proteins in which correct folding can be confirmed by a functionality assay. Furthermore, proteins expressed in exogenous hosts such as *E. coli* may not undergo the necessary post-translational modifications such as acetylation or ubiquitination that are carried out in the native host (Altman-Price & Mevarech 2009, Humbard et al 2010).

Purification of halophilic proteins expressed in *H. volcanii*

Halophilic proteins expressed natively in *H. volcanii* do not require refolding. Several protein purification techniques are available for recombinant proteins expressed in *H. volcanii*.

For example, Jolley and colleagues expressed dihydrolipoamide dehydrogenase in *H. volcanii* and the protein was purified using a specifically adapted hydroxylapatite chromatography technique (Jolley et al 1996). Hydroxylapatite resin $(Ca_5(PO_4)_3OH)_2$ contains positively charged calcium ions and negatively charged phosphate ions that interact carboxylate residues at the protein surface and basic protein residues respectively. Depending on specific protein properties, proteins will be eluted at different points across a phosphate gradient (Schroder et al 2003).

Additionally, Humbard and colleagues employed tandem affinity tagging to purify 20S proteasome from *H. volcanii*. The α1-subunit was C-terminally His_6-tagged and the ß-subunit was StrepII-tagged and purification was achieved by metal affinity chromatography followed by application of the protein complex to a StrepTactin (streptavidin variant) column (Humbard et al 2009). This approach of purification could be suitable for developing generic methods of expressing and purifying proteins from *H. volcanii* due to compatibility of these tags with high salt.

However, the protein expression constructs used in these studies were custom made and tailored to the protein of interest in question. The development of generic plasmid vectors and host strains for conditional overexpression of halophilic proteins in *H. volcanii* by Allers and colleagues are described in further detail below.

Plasmid constructs

Prior to the commencement of this study, Allers and colleagues had developed generic episomal overexpression constructs that allow for native conditional over-expression of N-terminally His_6-tagged *H. volcanii* proteins (Allers et al 2010).

pTA963 and pTA1228 feature a *his_6-tag* flanked by restriction sites that allow for in-frame insertion of a gene of interest (Figure 6.1). pTA963 contains the restriction sites *Pci*I, *Nco*I and *Bsp*HI to allow for insertion of genes that have the second codon of comprising of a T, G and A respectively. The improved vector pTA1228 contains an additional *Nsp*I restriction site, which is compatible with *Sph*I and allows for insertion of genes where the second codon starts with a C.

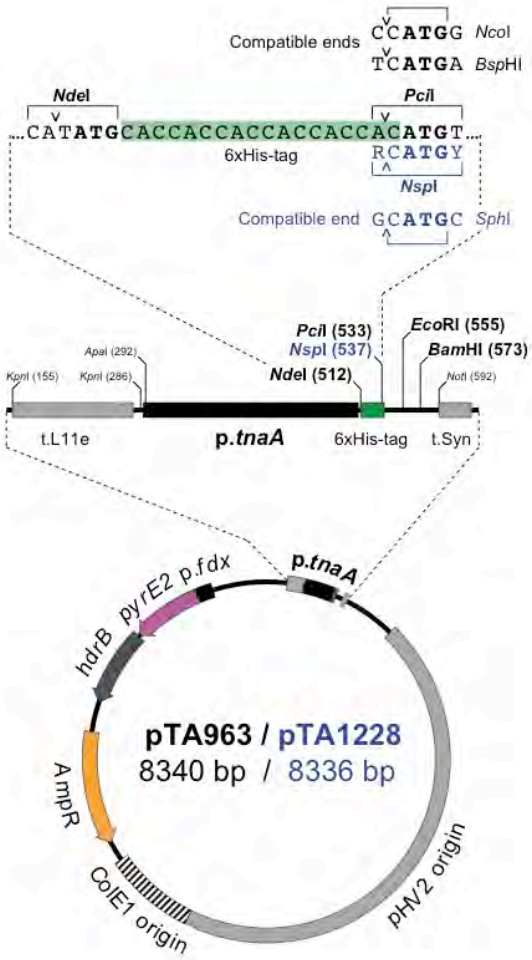


Figure 6.1: **Plasmids for conditional overexpression of His_6-tagged proteins.** pTA1228 differs from pTA963 by having additional sites available (blue). Plasmids contain a *his_6-tag* downstream of a tryptophan inducible promoter (*p.tnaA*) and *pyrE2* and *hdrB* markers for selection in *H. volcanii*. hown on the schematic are restriction sites available for inserting genes of interest (and available compatible enzymes). Abbreviations: *ampR* (ampicillin resistance gene, *E. coli*), *colE1 ori* (*E. coli* origin of replication), *f1(+)ori* (*E.coli* origin of replication), *hdrB* (thymidine biosynthesis, *H. volcanii*), *his_6-tag* (hexahistidine tag), *pHV2* (*H. volcanii* origin of replication), *p.tnaA* (tryptophan inducible promoter), *p.fdx* (ferredoxin promoter), *pyrE2* (uracil biosynthesis, *H. volcanii*), *t.L11e* (terminator), *t.Syn* (terminator). In the restriction sites, Y = C or T and R = A or G. Adapted from (Allers et al 2010)

The *p.tnaA* promoter used to express proteins from pTA963 and pTA1228 is leaky in *E. coli* (Allers et al 2010, Large et al 2007). Therefore, plasmids containing genes encoding for halophilic proteins that are toxic to *E.coli* must be constructed directly in *H. volcanii*.

Overexpression Strains

Previous to this study, the *H. volcanii* strain H1424 (Δ*pyrE2*, Δ*hdrB*, Δ*nrr*, *Nph-pitA*, *cdc48d-Ct*) was available for protein overexpression (Allers et al

2010, Stroud et al 2012). The Δ*mrr*, *Nph-pitA* and *cdc48d-Ct* alleles were all generated specifically for this technique, and will be discussed below.

H. volcanii encodes a restriction enzyme, Mrr, which cuts methylated 5′-GATC-3′ sequences. *E. coli* contains a Dam methylase that methylates 5′-GATC-3′ sequences. For efficient transformation of plasmids into *H. volcanii*, the plasmid DNA must be passaged through a *dam-* strain of *E. coli* prior to transformation as the efficiency of transforming dam+ DNA into a *mrr+* strain is 50x lower than transforming with dam- DNA or using a Δ*mrr* strain (Holmes et al 1991). The *mrr* gene was deleted from *H. volcanii* to allow for efficient transformation of plasmids directly into *H. volcanii* (>10^7 transformants/μg DNA) without needing to be passaged first through a *dam-* strain of *E. coli* (Allers 2010).

The *pitA* protein of *H. volcanii* contains a histidine rich linker region and is therefore a major co-contaminant when using metal affinity chromatography to purify His-tagged proteins (Bab-Dinitz et al 2006, Humbard et al 2009). To overcome this problem, the histidine rich linker in *H. volcanii* was replaced with a non-histidine rich orthologue from *Natronomonas* pharaonis (Allers et al 2010).

The elimination of the PitA contaminant revealed another histidine rich contaminant at the same size as PitA: Cdc48d (HVO_1907). *cdc48d* contains a histidine rich C-terminal domain, a truncated version of this gene was generated to reduce the problem of co-contamination during metal affinity chromatography. Both PitA and Cdc48d are essential genes in *H. volcanii*.

However, even after removal of the two major histidine rich contaminants PitA and Cdc48d, many other histidine rich contaminant proteins still persist during metal affinity chromatography (Stroud et al 2012), Figure 6.2. The abundance of histidine rich proteins is characteristic of halophilic organisms.

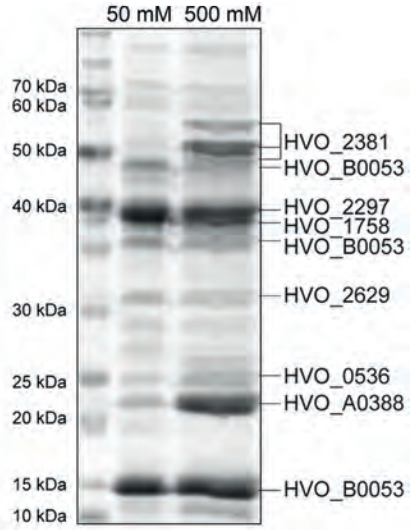


Figure 6.2: **Histidine rich contaminant proteins in metal affinity chromatography.** Empty vector protein overexpression plasmid pTA963 was used in a mock protein overexpression. Histidine rich cellular proteins were purified from the soluble lysate fraction by affinity chromatography on a Ni2+ chelating column, bound proteins were eluted using 50 and 500 mM imidazole. Proteins were precipitated using using trichloroacetic acid and deoxycholate to enhance visualisation, Proteins were identified by mass spectrometry, *H. volcanii* gene numbers indicated. Figure adapted from (Stroud et al 2012)

The naturally occurring histidine rich contaminants that are observed during metal affinity chromatography hinder protein purification in *H. volcanii* as well as the detection of protein:protein interactions by protein pull down assays.

6.2 Aims

The development of the generic episomal overexpression constructs pTA963 and pTA1228 and the expression strain H1424 (Δ*pyrE2*, Δ*hdrB*, Δ*mrr*, *Nph-pitA*, *cdc48d-Ct*) has allowed for the expression and purification of His_6-tagged recombinant proteins from *H. volcanii*. However, *H. volcanii* contains many histidine rich proteins which cause contamination during metal affinity chromatography, therefore it is of interest to develop new methods for protein expression and purification in *H. volcanii*. The aims of this chapter are to:

- Develop generic strains, plasmids and techniques for overexpression and purification of His_6-tagged and StrepII-tagged recombinant proteins

natively in *H. volcanii*.

- Screen for *in vivo* protein:protein interactions of Hel308 by expressing His_6-tagged and StrepII-tagged Hel308 natively in *H. volcanii* and analysing co-purifying proteins. This may give further insight into the cellular role of Hel308.

6.3 Results

6.3.1 Development of improved strains for protein overexpression

In *H. volcanii,* the overexpression of proteins on episomal plasmids is induced by the addition of tryptophan (Large et al 2007). Overexpression plasmids contain the gene of interest downstream of the tryptophan inducible promoter p.*tnaA*. A final concentration of ~4.5 mM tryptophan is added to *H. volcanii* culture to induce protein expression.

To improve strains for protein expression and purification the *tnaA* gene was deleted. *tnaA* encodes for tryptophanase, an enzyme which degrades tryptophan. Deletion of *tnaA* will eliminate the breaking down of tryptophan by tryptophanase therefore all of the tryptophan added will be used for the induction of protein expression. A second reason for deleting *tnaA* is that in *Escherichia coli*, the enzyme tryptophanase (*tnaA*) produces indole from tryptophan (Li & Young 2013). Indole is a signalling molecule that that is involved in the regulation of many processes for example motility, quorum sensing, biofilm formation and quiescence (Bansal et al 2007, Chen et al 2015, Hu et al 2010). Indole has been shown to promote quiescence by blocking cell division and growth at concentrations of 3 to 5 mM. Cell division is blocked as the presence of indole prevents the formation of the FtsZ ring; FtsZ polymerises into a ring structure at mid-cell and defines the site of cell division. Without FtsZ cytokinesis is unable to occur (Chen et al 2015, Chimerel et al 2012, de Boer et al 1992).The addition of tryptophan to *H. volcanii* cultures is required for protein expression. However, this very same

tryptophan could be converted to indole by the tryptophanase gene *tnaA* and therefore preventing FtsZ ring formation in *H. volcanii* resulting in quiescence (Wang & Lutkenhaus 1996). For this reason *tnaA* (Hvo_0009) was deleted from *H. volcanii* to generate an improved strain for efficient protein overexpression by tryptophan induction.

Hvo_0009 (*tnaA*) was deleted from the overexpression and purification strain H1424 and H1895, which is H1424 with an additional deletion of *pilB3C3*. Details of these strains are described below.

H1424 (*ΔpyrE2, Nph-pitA, Δmrr, cdc48d-ct, ΔhdrB)* is a protein overexpression and purification strain in *H. volcanii* (Allers et al 2010, Stroud et al 2012). This strain was developed in order to reduce the co-purification of naturally histidine rich proteins with *his$_6$-tagged* recombinant proteins.

H1895 (*ΔpyrE2, Nph-pitA, Δmrr, cdc48d-ct, ΔhdrB, ΔpilB3C3)* is derived from H1424 and is deleted for Hvo_1033 (*pilC3)* and Hvo_1034 (*pilB3*), which encode an ATPase and an integral membrane protein of the pilus assembly system respectively (Strillinger et al 2016). These proteins are necessary for biofilm formation and deletion of these genes eliminates the attachment of *H. volcanii* to surfaces of any kind while leaving cells motile, therefore preventing biofilm formation during culture within a fermenter.

Hvo_0009 (*tnaA*) was deleted from H1424 and H1895 using the *ΔtnaA::hdrB+* deletion construct pTA1615 to generate the pop-outs H2167 and H2169 respectively. *tnaA* deletions were confirmed by colony hybridisation using a 1488 bp radiolabelled probe of *tnaA* digested from pTA875 with *Age*I, Figure 6.3. *tnaA* deletions were also confirmed by the ability of the strains to grow on media lacking Thy.

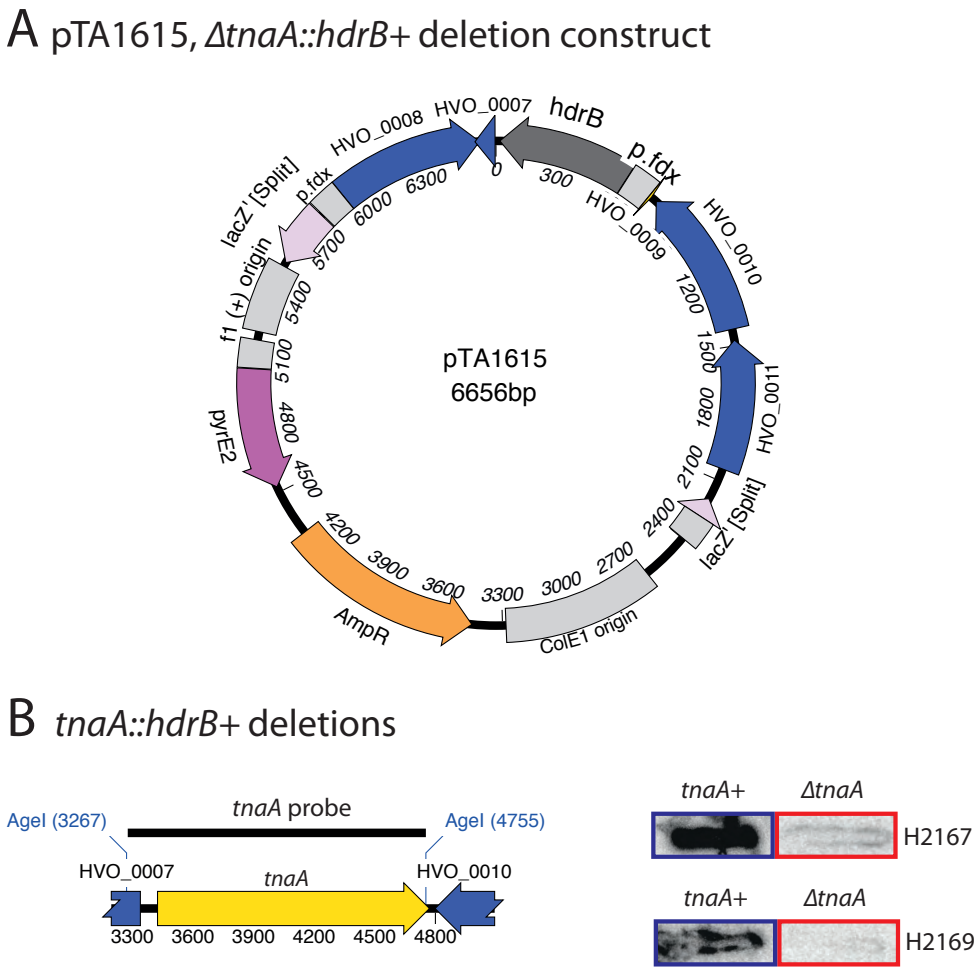


Figure 6.3: **Deletion of Hvo_0009 (*tnaA*). (A)** pTA1615 *ΔtnaA::hdrB+* deletion construct **(B)** Candidate colonies were screened with a radiolabelled *Age*I digested fragment from pTA875. Colonies that did not hybridise were *ΔtnaA*.

In order that *hdrB* selection can be utilised in episomal expression plasmids, H2167 and H2169 containing *ΔtnaA::hdrB+* were transformed with pTA1508 (*ΔtnaA*) to remove the *hdrB+* marker from the *ΔtnaA* locus, Figure 6.4A. Pop-outs were patched in duplicate on media containing +/- thy to select for colonies lacking the *hdrB+* marker. No *hdrB-* strains were found (data not shown), indicating that attempts to remove the *hdrB+* marker this way were unsuccessful. The *ΔtnaA* partial deletion construct, pTA1730, was constructed by the removal of a 588 bp *Eco*NI-*Eco*NI fragment from *tnaA* and transformed into H2169 to generate the strain H2682 (*tnaAΔEcoNI*, *ΔhdrB)*, Figure 6.4B. The *tnaAΔEcoNI* partial deletion was confirmed by colony hybridisation, Figure 6.4C, Southern blot and by the inability ability of colonies to grow on media lacking Thy.

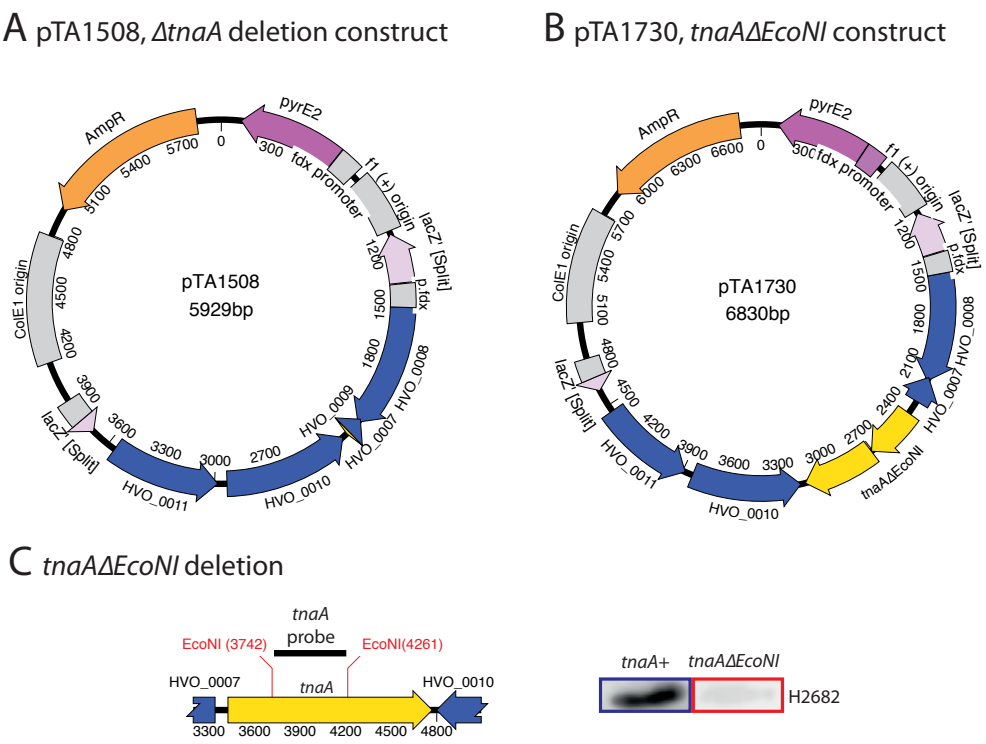


Figure 6.4: ***ΔtnaA* and *tnaAΔEcoNI* deletion constructs. (A)** *ΔtnaA* deletion construct pTA1508. **(B)** *tnaA* partial deletion construct containing *tnaAΔEco*NI fragment, pTA1730. **(C)** Candidate colonies were screened with a radiolabelled *Eco*NI digested fragment from pTA875. Colonies that did not hybridise were *tnaAΔ Eco*NI.

In the first instance, *tnaA* could only be deleted using a *hdrB*+ marker using the deletion construct pTA1615 (*ΔtnaA::hdrB*+) to generate the strains H2167 and H2169. The removal of the *hdrB*+ marker could only be achieved using the *tnaA* partial deletion construct pTA1730 (*tnaAΔEcoNI*) but not with the 'whole' *tnaA* deletion construct pTA1508 (*ΔtnaA*). These results suggest that a full *tnaA* deletion is not possible; perhaps a full deletion interrupts the read through of transcription of a potentially essential neighbouring gene to *tnaA*. Perhaps the integrity of the gene neighbourhood is preserved when *tnaA* is only partially deleted, or when the *hdrB*+ marker which is a similar size to *tnaA* is present. Alternatively, the difficulty in generating a full *tnaA* deletion could be due to unknown technical reasons.

Due to time constraints and the difficulties in generating a *tnaA* deletion in the H1424 and H1895 backgrounds, the resulting strain H2682 (*tnaAΔEcoNI*,

ΔpilB3C3) was not tested for improved growth upon the addition of tryptophan and improvements in protein overexpression by tryptophan induction.

To determine if the partial deletion of *tnaA* in H2682 (*tnaAΔEcoNI*, *ΔpilB3C3*) results in improved growth by blocking indole-induced quiescence, growth assays would have been performed. Growth assays in liquid culture would have been performed by measuring the A_{600} of H2682 (*tnaAΔEcoNI*, *ΔpilB3C3*) along side H1895 (*ΔpilB3C3)* using an Epoch2 Microplate Spectrophotometer (BioTek) in the presence and absence of tryptophan. A range of 2 – 5 mM would be used to mimic the concentrations of tryptophan used in protein overexpression protocols. This assay would determine if strains deleted for *tnaA* were able to grow better than (or as well as) strains not deleted for *tnaA*. The efficiency of protein overexpression by tryptophan induction would be assayed by transformation with a plasmid containing a tagged protein under the control of a tryptophan inducible promoter. Levels of protein expression would have been measured by comparing amounts of protein purified from H2682 (*tnaAΔEcoNI*, *ΔpilB3C3)* to its parent strain H1895 (*ΔpilB3C3)*. Alternatively, the level of protein expression could be measured by inserting a reporter gene such as *bgaH* could be placed under the control of the tryptophan inducible promoter.

6.3.2 Development of episomal overexpression plasmid constructs

The episomal overexpression constructs pTA963 and pTA1228 currently used in *H. volcanii* allows for the expression and purification of N-terminally His_6-tagged recombinant proteins. However, *H. volcanii* contains many histidine-rich proteins, which cause contamination during metal affinity chromatography, Figure 6.2. Therefore, an alternative method of affinity-tag protein purification is required for use in *H. volcanii*. An important point of consideration is that the affinity tag of choice must be compatible with high salt conditions since proteins in *H. volcanii* are adapted to function in approximately 2 M salt. Examples of commonly used alternatives to His_6-tags are shown in Table 6.1.

Table 6.1: **Commonly used tags for protein purification.**

Tag	Size (aa)	Matrix
Maltose binding protein (MBP)	396	Amylose
Glutathione *S*- transferase (GST)	218	GSH-sepharose
Calmodulin-binding peptide (CBP)	28	Calmodulin affinity
Chitin-binding domain (CBD)	51	Chitin
StrepII	8 (WSHPQFEK)	Strep-Tactin sepharose
FLAG	8 (DYKDDDDK)	Anti-FLAG M2 mAb agarose
Heavy chain of protein C (HPC)	12	Anti-Protein C mAb matrix
Covalent yet dissociable (CYD)	5	InaD
S	15	S-protein of RNase A
HA	9	Anti-HA epitope mAb
c-Myc	11	Anti-Myc epitope mAb

Adapted from (Kimple et al 2013, Zhao et al 2013)

Large protein tags such as MBP and GST are likely to mis-fold in high salt conditions and therefore not be functional in the purification of halophilic proteins. The large size of these tags may also hinder protein function. Similarly, antibody based purification systems such as FLAG, HPC, HA and c-Myc may not be suitable as they are expensive and likely to be unstable in high salt conditions. Purification protocols using CBP and S require protein binding in low salt and protein elution in high salt. Since the protein will need to be maintained in high salt concentrations to ensure correct folding, these tags are not suitable for halophilic proteins (Kimple et al 2013). A possible candidate for purification of halophilic proteins is the CBD tag, which is stable up to 1 M NaCl, however this tag is bulky and could hinder protein function (Terpe 2003). StrepII tags are stable up to 5 M NaCl and Humbard and colleagues have successfully used StrepII tags to purify proteins expressed natively in *H. volcanii* (Humbard et al 2009). For these reasons, the StrepII tag was the tag of choice for the development of improved generic overexpression plasmids for use in *H. volcanii* within this study.

Plasmid Constructs

Plasmid constructs pTA1392 and pTA1403 were developed to allow for conditional overexpression of *H. volcanii* proteins with the option of a His_6-tag and/or a StrepII tag.

pTA1392 allows for in-frame insertion of a gene of interest with the option to have an N-terminal *his_6-tag* and/or a C-terminal *strepII-tag*, this is achieved by integrating a $(CAC)_6$ tract and a WSHPQFEK motif downstream of p.*tnaA*, Figure 6.5. A number of sites are available for inserting genes of interest in-frame with the *his_6-tag* and *strepII-tag*. To generate a gene without a N-terminal *his_6-tag*, the start codon of the gene of interest is replaced with an *Nde*I site and ligated with the *Nde*I end of the plasmid. For the inclusion of an N-terminal *his_6-tag* genes are inserted at *Pci*I (AC**ATG**T) or at *Nsp*I (RC**ATG**Y, where R = A or G and Y = C or T), both contain internal in-frame ATG start codons (bold). Additional enzymes that produce compatible ends are *Nco*I (CC**ATG**G), *Bsp*HI (TC**ATG**A) and *Sph*I (GC**ATG**C). The restriction site at the 5' end of the gene of interest is selected based on the first residue in the gene after the ATG start site. Therefore, the 4th residue of the gene should correspond to the 6th residue of the restriction site: *Pci*I is used when second codon in the coding sequence starts with a T, *Nco*I when it begins with a G, *Bsp*HI when it begins with an A and *SphI* when it begins with a C.

To generate a gene without a C-terminal *strepII-tag,* an *Eco*RI *or Bam*HI site is incorporated after the stop codon and ligated, and for the incorporation of a C-terminal *strepII-tag* the stop codon of the gene is replaced with an *Nhe*I or a compatible site *(Avr*II*, Spe*I *or Xba*I*)* and ligated with the *Nhe*I end of the plasmid.

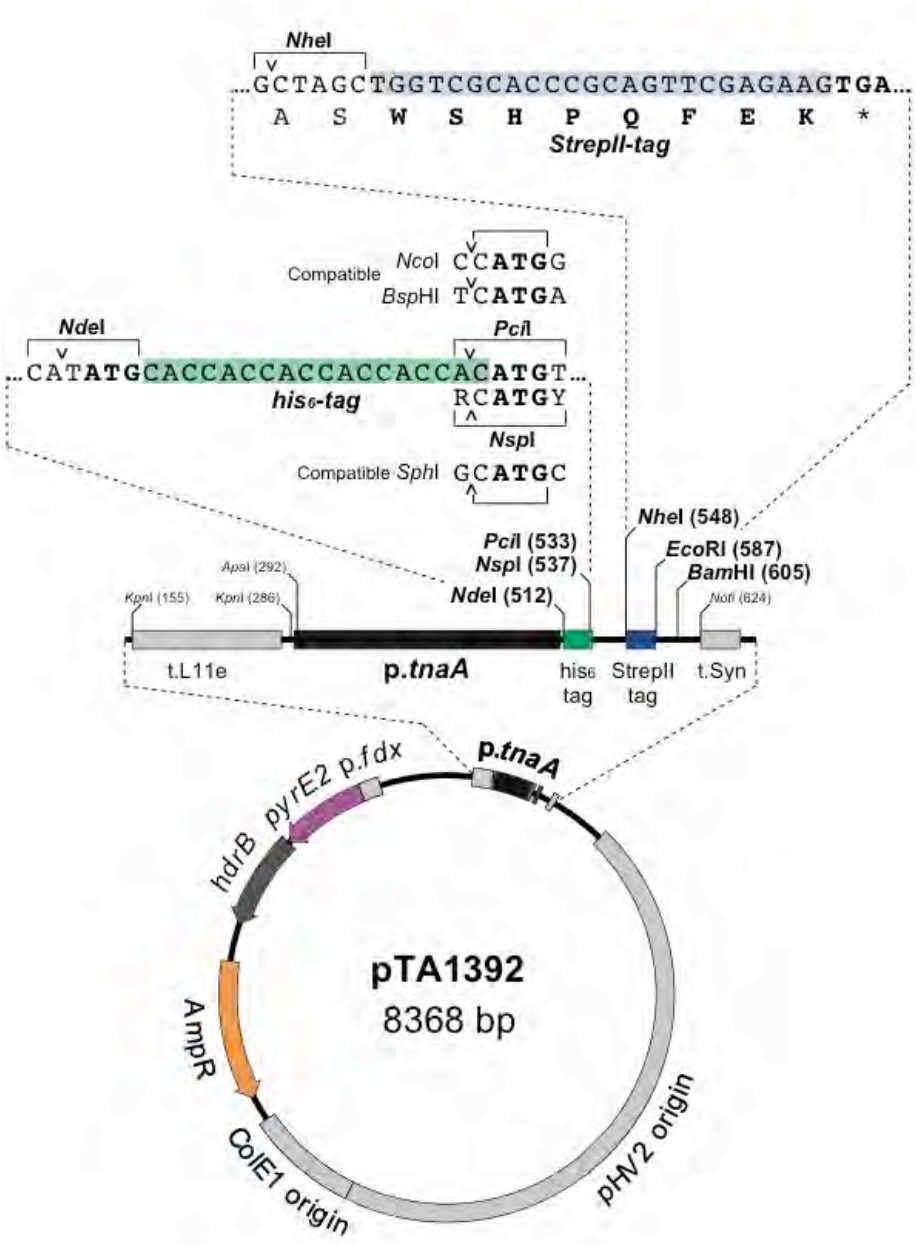


Figure 6.5: **pTA1392.** A plasmid for the conditional overexpression of N-terminally *his*$_6$ and/or C-terminally *strepII* tagged proteins. The plasmid contains a tryptophan inducible promoter (*p.tnaA*), *pyrE2* and *hdrB* markers for selection in *H. volcanii*. Restriction sites for inserting gene of interest are shown. Abbreviations: *ampR* (ampicillin resistance gene, *E. coli*), *colE1 ori* (*E. coli* origin of replication), *hdrB* (thymidine biosynthesis, *H. volcanii*), *his*$_6$*-tag* (hexahistidine tag), *pHV2* (*H. volcanii* origin of replication), *p.tnaA* (tryptophan inducible promoter), *p.fdx* (ferredoxin promoter), *pyrE2* (uracil biosynthesis, *H. volcanii*), *StrepII tag* (streptavidinII tag) *t.L11e* (terminator), *t.Syn* (terminator). In the restriction sites, Y = C or T and R = A or G.

pTA1403 allows for in-frame insertion of a gene of interest with the option to have an N-terminal *strepII-tag* and/or a C-terminal *his*$_6$*-tag*, this is achieved by

integrating a WSHPQFEK motif and a $(CAC)_6$ tract downstream of p.*tnaA*, Figure 6.6.

To generate a construct without a N-terminal *strepII-tag* the start codon is replaced with an *Nde*I site and ligated with the *Nde*I end of the plasmid. For the incorporation of an N-terminal *strepII-tag* the start codon is replaced with an *Nco*I, *Bsp*HI, *Pci*I or an *Sph*I site and ligated with the *Pci*I or *Nsp*I end of the plasmid. The restriction site is selected based on the first residue in the gene of interest after the ATG start site as described for inclusion of an N-terminal his_6*-tag* gene in pTA1392.

To generate a gene without a C-terminal his_6*-tag*, an *Eco*RI *or Bam*HI site is incorporated after stop codon and ligated, and for the inclusion of a C-terminal his_6*-tag* the stop codon of the gene is replaced with an *Nhe*I or compatible site *(Avr*II, *Spe*I *or Xba*I*)* and ligated with the *Nhe*I end of the plasmid.

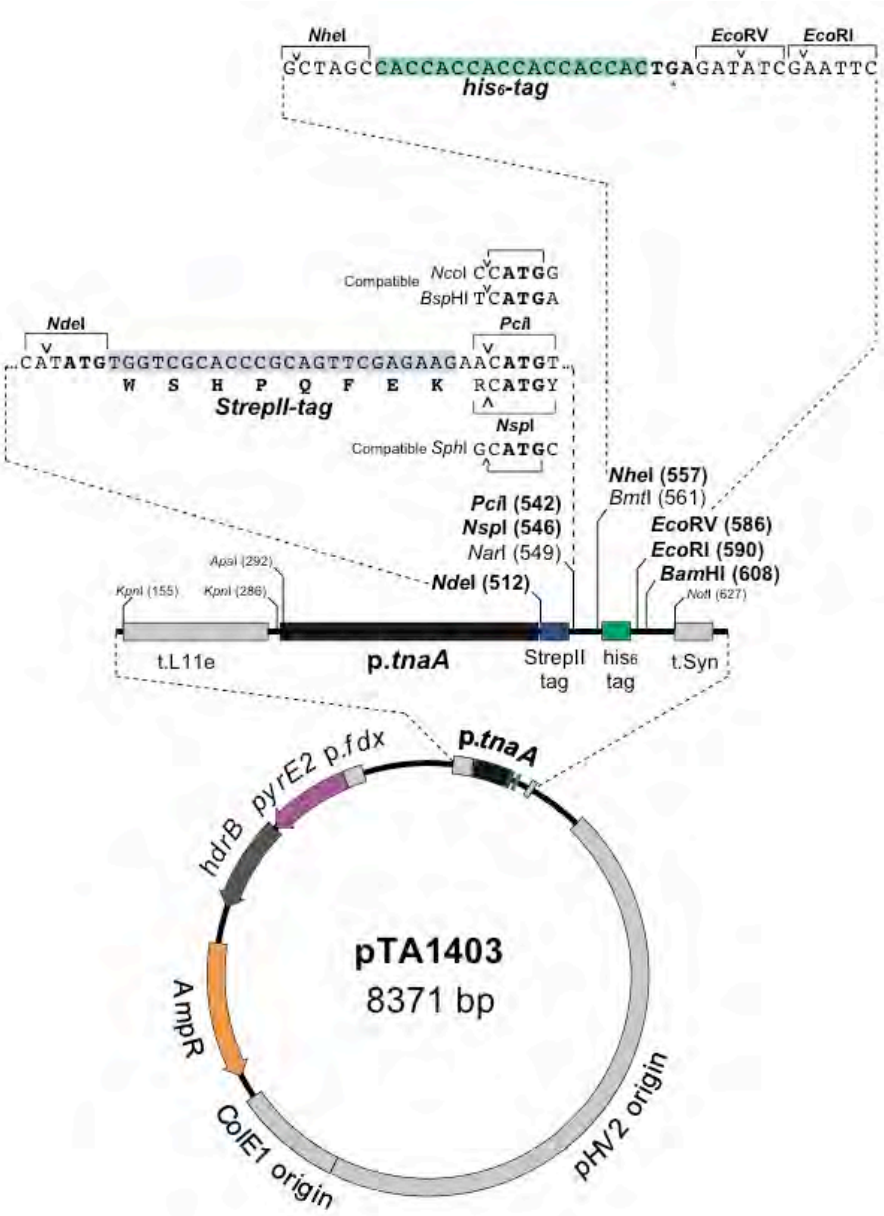


Figure 6.6: **pTA1403**. A plasmid for the conditional overexpression of N-terminally *strepII* and/or C-terminally his_6 tagged proteins. The plasmid contains a tryptophan inducible promoter (*p.tnaA*), *pyrE2* and *hdrB* markers for selection in *H. volcanii*. Restriction sites for inserting gene of interest are shown. Abbreviations: *ampR* (ampicillin resistance gene, *E. coli*), *colE1 ori* (*E. coli* origin of replication), *drB* (thymidine biosynthesis, *H. volcanii*), *his6-tag* (hexahistidine tag), *pHV2* (*H. volcanii* origin of replication), *p.tnaA* (tryptophan inducible promoter), *p.fdx* (ferredoxin promoter), *pyrE2* (uracil biosynthesis, *H. volcanii*), *StrepII tag* (streptavidinII tag) *t.L11e* (terminator), *t.Syn* (terminator). In the restriction sites, Y = C or T and R = A or G.

Induction of gene expression

The tightly controlled tryptophan inducible promoter p.*tnaA* is used for the induction of gene expression in pTA1392 and pTA1403 (Large et al 2007).

Tryptophan is an energetically costly amino acid for a cell to produce, requiring the equivalent of 75 ATPs for synthesis, therefore genes involved in tryptophan biosynthesis and degradation are tightly regulated. *p.tnaA* is the promoter for a tryptophanase gene, which encodes a protein that is involved in the degradation of excess tryptophan. p.*tnaA,* has a basal activity that is effectively zero, and is rapidly and strongly induced in the presence of ≥ 1 mM tryptophan which gives up to 100-fold induction (Large et al 2007).

Termination of gene expression

The transcriptional terminators *t.L11e* and *t.Syn* are present outside of the *p.tnaA::his$_6$-tag/strepII-tag* and *p.tnaA::strepII-tag/his$_6$-tag* region in pTA1392 and pTA1403 respectively. These terminators insulate the gene of interest from read-though transcription by cryptic promoters present elsewhere on the plasmid. *t.L11e* is an L11e rRNA terminator from *H. volcanii* and *t.Syn* is a synthetic terminator (comprising a T tract flanked by G/C-rich sequences) that function in *H. volcanii* (Allers et al 2010, Shimmin & Dennis 1996).

Replication of plasmid in H. volcanii

pTA1392 and pTA1403 contain the origin of replication *ori-pHV2,* which allows for replication in *H. volcanii*. *ori-pHV2* originated from pHV2, a plasmid that was cured from the laboratory strain of *H. volcanii* (Wendoloski et al 2001). Plasmids containing *ori-pHV2* are present at approximately six copies per genome (Charlebois et al 1987).

Selection of plasmid in H. volcanii

In order to select for pTA1392 or pTA1403, *pyrE2*+ and *hdrB*+ markers are present on the plasmids. These plasmids are transformed into *H. volcanii* strains that are Δ*pyrE2* (uracil) and Δ*hdrB* (thymidine) (Allers et al 2004, Bitan-Banin et al 2003).

Replication of plasmid in E. coli

Due to speed of growth and higher copy number, plasmids are often generated in *E. coli* rather than *H. volcanii*. For replication of the plasmids in *E. coli* to occur, the *E. coli* origin of replication, *ColE1 ori*, is present.

Selection of plasmid in E. coli

The plasmids contain the ampicillin resistance marker, *ampR*, which allows for selection of the plasmid when transformed into *E. coli* and plated on media containing ampicillin.

6.3.3 Protein overexpression and purification

The N-terminal his_6-tag and C-terminal StrepII-tag (pTA1392) and N-terminal StrepII-tag and C-terminal His_6-tag (pTA1403) constructs were tested for their suitability for improved protein overexpression and purification in *H. volcanii,* using Hel308 as a candidate protein.

Induction of expression

Protein induction is under the control of the tryptophan inducible promoter p.*tnaA*. Concentrations of >1 mM tryptophan affect the growth of *H. volcanii* therefore induction was delayed until 2 hours before harvesting (Allers et al 2010). To induce protein expression a final concentration of 3 mM tryptophan was added to the culture and incubated for 1 hour then the tryptophan was increased to 4.5 mM final concentration and incubated for a further hour.

Protein purification

Metal Affinity Chromatography

Proteins containing a His_6-tag were purified from cell lysate using metal affinity chromatography. Due to their ease of use in small-scale purification, gravity columns packed with charged Sepharose beads were used. In this study

Ni^{2+} charged beads were used, since they result in a higher protein yield than Co^{2+} charged beads in conditions of >1 M NaCl (Allers et al 2010). The tagged protein was eluted using 100-500 mM imidazole. If the protein was to be subsequently purified using a Strep-Tactin column, the sample was diafiltrated using appropriate MWCO Vivaspin 20 ultrafiltration spin columns to remove residual imidazole and to concentrate the protein sample.

Strep-Tactin Chromatography

Proteins containing a StrepII-tag were purified either from the cell lysate or from elutions following metal affinity chromatography, using gravity columns packed with *Strep*-Tactin Sepharose. Protein was eluted using 5 mM D-desthiobiotin. Following purification, protein samples were precipitated using trichloroacetic acid (TCA) to aid visualisation of protein bands on SDS-PAGE gels.

Purification of N-terminally His_6-tagged and C-terminally StrepII-tagged Hel308

In order to purify N-terminally His_6-tagged and C-terminally StrepII-tagged Hel308, H1424 was transformed with pTA1419 (*p.tnaA:: his_6tag-hel308+strepIItag pyrE2+ hdrB+*) to generate the strain H1737. For details of plasmid construction, see Chapter 3: *Plasmid and Strain Construction*. His_6-Hel308-StrepII was expressed and purified by metal affinity chromatography and Strep-Tactin chromatography, Figure 6.7A, or by Strep-Tactin chromatography alone Figure 6.7B. Details of these methods are described previously.

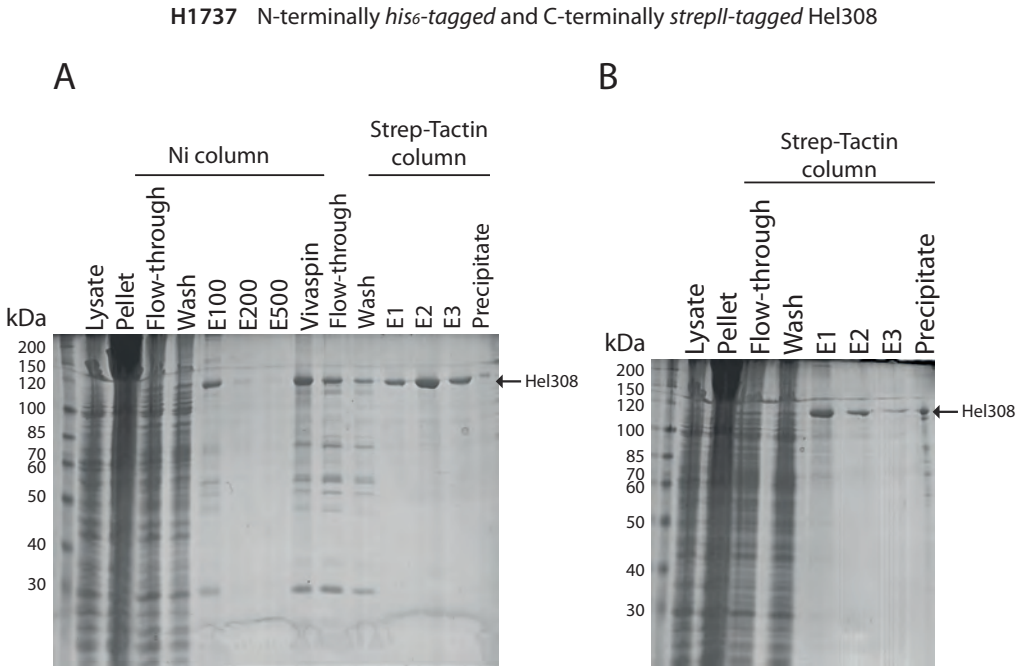


Figure 6.7: **Purification of N-terminally his_6-tagged and C-terminally strepII-tagged Hel308. (A)** His_6-Hel308-strepII was expressed and purified from H1737 by metal affinity chromatography using a step gradient of elution buffer containing imidazole (E refers to the imidazole concentration in mM), following with Strep-Tactin chromatography using a 3 step elution with 0.4 ml, 0.7 ml and 0.4 ml of elution buffer containing 5 mM D-desthiobiotin (E refers to elution number). **(B)** His_6-Hel308-strepII was expressed and purified from H1737 by Strep-Tactin chromatography alone using a 3 step elution with 0.4 ml, 0.7 ml and 0.4 ml of elution buffer containing 5 mM D-desthiobiotin (E refers to elution number) Hel308 indicated, confirmed by mass spectrometry, Susan Liddell, University of Nottingham, data not shown.

As expected after purification of His_6-Hel308-StrepII with only Ni^{2+} metal affinity chromatography, the contaminating naturally histidine-rich proteins are still present in the elutions E100-E500 (Figure 6.7A, all bands not indicated as Hel308). After further purification with Strep-Tactin chromatography, a large number of these contaminating bands have been removed (Figure 6.7A, E1-E3). After purification of His_6-Hel308-StrepII with Strep-Tactin chromatography alone, some contaminating proteins are seen in the elutions E1-E3 (Figure 6.7B, all bands not indicated as Hel308). This is likely due to overloading of the Strep-Tactin sepharose beads by the complete cell lysate. Furthermore, the overall amount of Hel308 recovered in E1-E3 appears to be less than that recovered in E1-E2 in Figure 6.7A, this is probably due to overloading of the Strep-Tactin sepharose beads. Therefore, purification of His_6-Hel308-StrepII is most successful by Ni^{2+} metal affinity chromatography

followed by Strep-Tactin chromatography, rather than by Strep-Tactin chromatography alone.

Purification of N-terminally StrepII-tagged and C-terminally His_6-tagged Hel308

In order to purify N-terminally StrepII-tagged and C-terminally His_6-tagged Hel308, H1424 was transformed with pTA1425 (*p.tnaA::strepII tag-hel308+ his_6-tag pyrE2+ hdrB+*) to generate the strain H1743. For details of plasmid construction, see Chapter 3: *Plasmid and Strain Construction*. StrepII-Hel308-His_6 was expressed and purified by metal affinity chromatography and Strep-Tactin chromatography, Figure 6.8A, or by Strep-Tactin chromatography alone Figure 6.8B. Details of these methods are described previously.

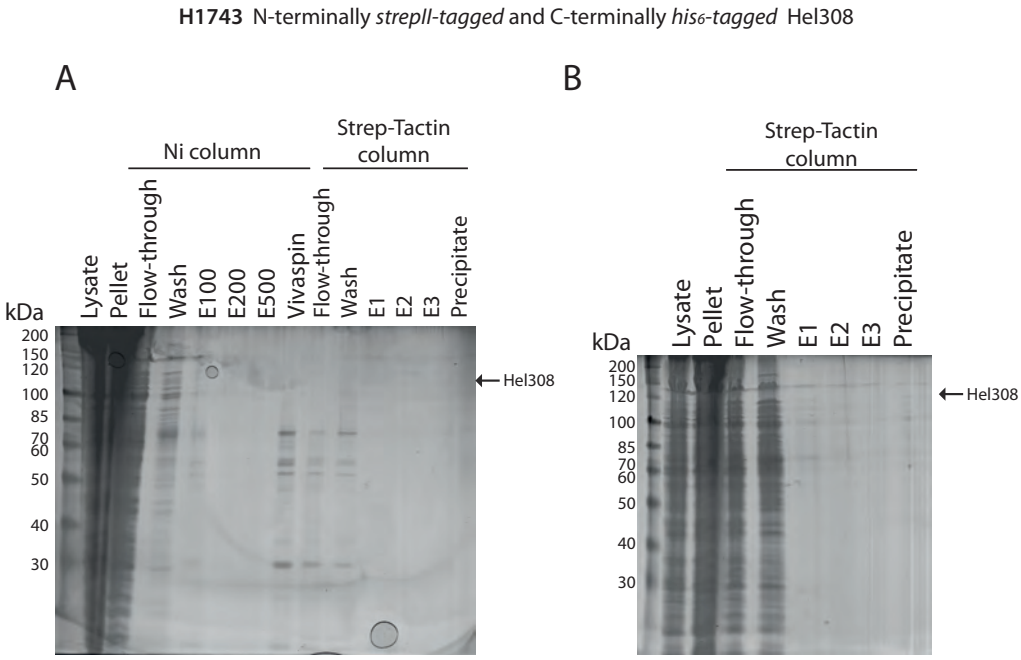


Figure 6.8: **Purification of N-terminally strepII-tagged and C-terminally his_6-tagged Hel308. (A)** StrepII-Hel308-his_6 was expressed and purified from H1743 by metal affinity chromatography using a step gradient of elution buffer containing imidazole (E refers to the imidazole concentration in mM), following with Strep-Tactin chromatography using a 3 step elution with 0.4 ml, 0.7 ml and 0.4 ml of elution buffer containing 5 mM D-desthiobiotin (E refers to elution number). **(B)** StrepII-Hel308-his_6 was expressed and purified from H1743 by Strep-Tactin chromatography alone using a 3 step elution with 0.4 ml, 0.7 ml and 0.4 ml of elution buffer containing 5 mM D-desthiobiotin (E refers to elution number). Hel308 indicated, confirmed by mass spectrometry, Susan Liddell, University of Nottingham, data not shown.

As expected, after purification of StrepII-Hel308-His_6 with only Ni^{2+} metal affinity chromatography, the contaminating naturally his rich proteins are

present in the elutions E100-E500 (Figure 6.8A, all bands that are not indicated as Hel308). However, the recovery of Hel308 is very poor, even in the Vivaspin lane where the elutions from Ni^{2+} metal affinity chromatography have been concentrated. Again, the contamination is removed via application of the elutions to a Strep-Tactin column but Hel308 is still not present. After purification of StrepII-Hel308-His_6 with Strep-Tactin chromatography only, contaminating bands can be seen in E1-E3 (Figure 6.8B, all bands that are not indicated as Hel308). However, the recovery of Hel308 was very poor.

From Figures 6.7 and 6.8 it is evident that the N-terminal His_6-tag and C-terminal StrepII-tag orientation is better for purifying Hel308 than the N-terminal StrepII-tag and C-terminal His_6-tag orientation. The N-terminal His_6-tag and C-terminal StrepII-tag orientation was also seen to be better for purifying RadA (data not shown). Other members of the group have also found this orientation to be preferable for the purification of proteins such as RadB, Mre11-Rad50 and BktAB (beta-ketothiolase).

Purification of His_6-Hel308-StrepII was successful using Ni^{2+} metal affinity chromatography followed by Strep-Tactin chromatography. However, as many of the co-eluting proteins have been removed in this process, the identification of protein:protein interactions may be limited. The use of a zero-length protein cross-linking agent such as 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide (EDC) and *N*-hydroxysuccinimide (NHS) could be advantageous in detecting protein:protein interactions in this system (Grabarek & Gergely 1990).

6.3.4 Development of Chromosomally Tagged Expression Strains

At overexpressed levels, Hel308 is toxic to *H. volcanii* (Thorsten Allers, personal communication). Therefore, the identification of protein:protein interaction partners of overexpressed Hel308 may not give a true representation of protein interactions within the cell. Furthermore, with excess Hel308 within the cell the fraction of Hel308 that will interact with other proteins will be small. For this reason, a strain where the genomic copy of

hel308 is replaced with an N-terminally *his*$_6$-tagged and C-terminally *strepII*-tagged version of *hel308* was used to study protein:protein interaction partners of Hel308. The chromosomally *his*$_6$*-hel308-strepII* is under control of the natural promoter of *hel308* resulting in native levels of expression.

The chromosomal N-terminally *his*$_6$-tagged and C-terminally *strepII*-tagged *hel308* strain was generated in the H2047 background:

H2047 (*ΔpyrE2, Nph-pitA, Δmrr, cdc48d-ct, ΔtrpA)* is derived from H1424 via H1611 for improved protein expression and purification. To generate H1611, *trpA* was deleted from H1424 to allow for the selection of integrated *trpA*+ marked gene replacement or deletion constructs. H1611 was then transformed with a linear fragment containing the *hdrB* gene to make the resulting strain H2407 *hdrB*+. This was undertaken because *hdrB* selection is not required, and to ensure that strains do not encounter problems with thymidine starvation.

hel308::trpA+ was deleted from H2047 using pTA1277 (Δ*hel308::trpA*+ deletion construct) generating the strain H2131. An N-terminally *his6-tagged* and C-terminally *strepII-tagged hel308* was introduced into H2131 using the gene replacement construct pTA1662 to generate the strain H2418. Both strains were confirmed by colony hybridisation and Southern blot, for full details of plasmid and strain construction see Chapter 3: *Plasmid and Strain Construction*.

6.3.5 *In vivo* Protein:Protein Interactions

The study of protein:protein interactions is a central tool in the understanding of protein function and cellular processes.

MBP-tagged Hel308 from *M. thermautotrophicus* imobilised on amylose beads was shown to interact *in vitro* with His$_6$-tagged purified replication protein A (RPA) after incubation overnight at room temperature (Woodman et al 2011). RPA is the eukaryotic and archaeal single stranded binding (SSB) protein that binds unwound single stranded DNA to protect it from degradation and secondary structure formation.

In the hyperthermophilic archaeon *Sulfolobus tokodaii*, gel filtration, affinity pulldown, and yeast two-hybrid analyses revealed that the Hel308 homologue Hjm physically interacts with Hjc *in vitro* (Li et al 2008). Hjc is a Holliday junction resolvase that binds specifically to Holliday junctions and cleaves two opposing strands symmetrically to generate two recombinant duplexes to restart stalled replication forks (Lestini et al 2010).

In *in vivo* immunoprecipitation analyses were performed and the proliferating cell nuclear antigen (PCNA) was found to co-precipitate with Hel308 homologue Hjm through an interaction in the C-terminal domain of Hjm. PCNA is the DNA sliding clamp which anchors DNA polymerase to the template DNA to prevent dissociation. Furthermore, PCNA was shown to stimulate the helicase activity of Hjm at fork structured DNA (Fujikane et al 2006).

The mammalian Hel308 homologue, HelQ was shown to associate with the RAD51 paralogues RAD51B/C/D and XRCC2, and with the DNA damage-responsive kinase ATR *in vivo* (Takata et al 2013).

In order to gain a further insight into the role of Hel308, it was interest to identify other interacting proteins in *H. volcanii*. This was achieved by studying *in vivo* protein:protein interactions, with Hel308 expressed at a native level.

Proof of Principle - *Mre11* and *Rad50*

In order to test whether the method would be successful, a control experiment was carried out using Mre11 and Rad50. The *mre11* and *rad50* genes are found on an operon and their resulting proteins are known to interact. Mre11 and Rad50 are involved in the first step of homologous recombination in eukaryotes and archaea. Mre11 and Rad50 form a heterodimer that binds to and processes double strand breaks (D'Amours & Jackson 2002). Genetic analysis has indicated that this is also true in *H. volcanii* (Delmas et al 2009). Therefore, Mre11-Rad50 is an ideal control to use for the study of *in vivo* protein:protein interactions.

The strain H2239 was constructed in the H2047 background where the chromosomal *mre11* was N-terminally *his$_6$-tagged* and the chromosomal *rad50* was C-terminally *strepII tagged* under the control of their native promoter (Wickham-Smith 2015). The result of the co-purification assay is shown in Figure 6.9A and B. Bands were identified by mass spectrometry.

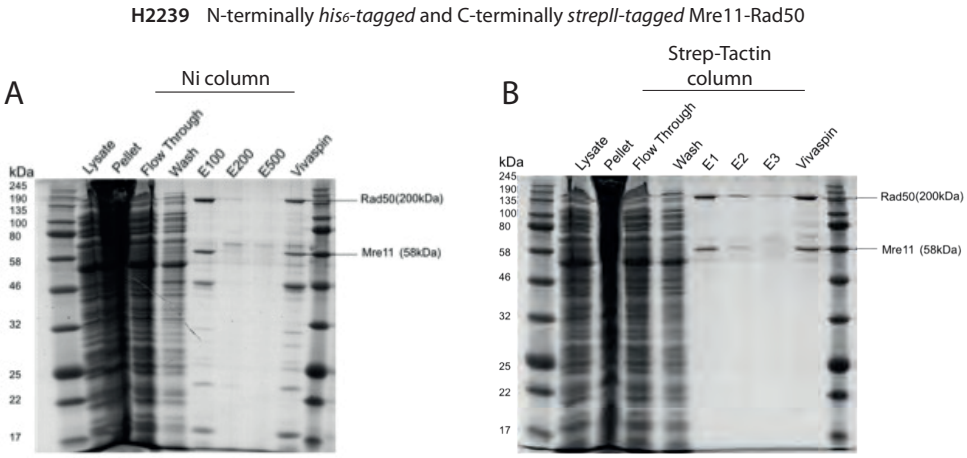


Figure 6.9: **Protein:protein interaction of N-terminally his$_6$-tagged and C-terminally strepII-tagged Mre11-Rad50. (A)** His$_6$-Mre11-Rad50-StrepII was expressed and purified from H2239 by metal affinity chromatography using a step gradient of elution buffer containing imidazole (E refers to the imidazole concentration in mM), Vivaspin sample obtained by pooling of elution samples. **(B)** His$_6$-Mre11-Rad50-StrepII was expressed and purified from H2239 by by Strep-Tactin chromatography alone using a 3 step elution with 0.4 ml, 0.7 ml and 0.4 ml of elution buffer containing 5 mM D-desthiobiotin (E refers to elution number). Vivaspin sample obtained by pooling of elution samples. Bands were identified by mass spectrometry (performed by Susan Liddell, University of Nottingham) Figure adapted from (Wickham-Smith 2015).

Rad50 was seen to co-purify with N-terminally His$_6$-tagged Mre11 after metal affinity chromatography, Figure 6.9A and Mre11 was seen to co-purify with C-terminally StrepII tagged Rad50 after Strep-Tactin chromatography. This confirms that this method can be used successfully to detect interactions between proteins known to form stable complexes.

Screening for Interaction Partners of Hel308

The following *in vivo* protein:protein interaction assay was preformed by Rebecca Lever (Doctoral Training Program rotation student) under my supervision.

In order to screen for interacting partners of Hel308 at native levels of expression, Hel308 was purified from the strain H2418 containing a chromosomal N-terminally *his6-tagged* and C-terminally *strepII-tagged hel308*. Because *Δhel308* has a marked survival deficiency after treatment with MMC, Hel308 is thought to act in the repair of DNA crosslinks (Adelman et al 2013, Takata et al 2013). The H2418 strain was incubated with 0.5 µg/ml of the DNA crosslinking agent MMC for 1 hour prior to protein purification, and only Strep-Tactin chromatography was performed. The result of this preliminary co-purification assay is shown below Figure 6.10. All bands were identified by mass spectrometry, data shown in Table 6.2,

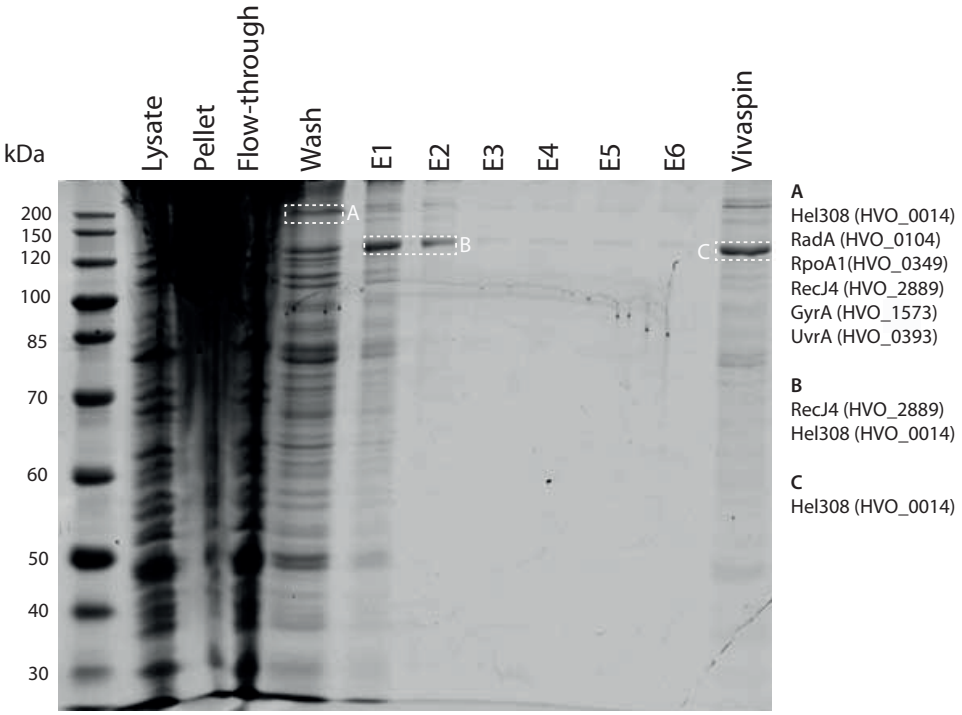


Figure 6.9: **Protein:protein interaction of N-terminally His_6-tagged and C-terminally StrepII-tagged Hel308.** His_6-Hel308-StrepII was expressed and purified from H2418 by Strep-Tactin chromatography alone using a 6 step elution with 0.4 ml, 0.7 ml and 4x 0.4 ml of elution buffer containing 5 mM D-desthiobiotin (E refers to elution number). Vivaspin sample obtained by pooling of elution samples. The DNA and RNA metabolism proteins Hel308 (HVO_0014), RadA (HVO_0104), RpoA1 (HVO_0349), GyrA (HVO_1573), RecJ4 (HVO_2889)and UvrA (HVO_0393) were identified in the Wash lane. Hel308 (HVO_0014) and RecJ4 (HVO_2889) was identified in Elutions 1 and 2. Bands were identified by mass spectrometry (performed by Susan Liddell, University of Nottingham).

Table 6.2: **Proteins identified by mass spectrometry**

Band A

Protein accession	Protein name	HVO_ number	Predicted MW	MASCOT score	# of peptides	Peptide sequences
D4GZY6	Tef1a1	0359	45731	1580	39	12
D4GTM3	CitB1	1955	70362	262	9	8
D4GT33	NrdJ	2452	224305	204	9	7
Q9HHA2	Cct3	0778	55208	167	13	6
Q48328	**RadA**	**0104**	**38251**	**162**	**4**	**3**
D4GXM0	**RecJ4**	**2889**	**79112**	**156**	**4**	**4**
D4GZX6	**RpoA1**	**0349**	**108846**	**155**	**6**	**6**
D4GZY3	Tef2	0356	80391	106	10	9
D4GXF1	PorA	1305	68427	106	2	1
D4GZ02	**GyrA**	**1573**	**94795**	**94**	**3**	**2**
D4GP52	CobN	B0050	141168	92	2	2
D4GYD4	GnaD	1488	45765	77	2	2
D4GYK9	**Hel308a**	**0014**	**90300**	**56**	**2**	**2**
D4GXE9	PorB	1304	34475	53	1	1
D4GR47	HyuA2	A0379	72923	53	1	1
D4H019	**UvrA**	**0393**	**108611**	**51**	**2**	**2**
P25062	Csg	2072	85138	40	1	1
Q48332	AtpA	0316	64474	33	3	1
D4GV93	Muc19	2160	232675	30	3	1
D4GZF1	AlaS1	0206	102401	28	1	1
D4GUC3	MetS	0809	81886	27	1	1

Band B

Protein accession	Protein name	HVO_ number	Predicted MW	MASCOT score	# of peptides	Peptide sequences
D4GYK9	**Hel308a**	**0014**	**90300**	**3495**	**125**	**30**
D4GZY3	Tef2	0356	80391	110	5	4
D4GXM0	**RecJ4**	**2889**	**79112**	**44**	**2**	**2**
D4GQB8	UspA27	A0086	14050	31	3	1
D4GZ93	UreA	0149	14136	29	10	1
D4GP28	2-keto-3-deoxyxylonate dehydratase	B0027	31842	28	9	1

Band C

Protein accession	Protein name	HVO_ number	Predicted MW	MASCOT score	# of peptides	Peptide sequences
D4GYK9	**Hel308a**	**0014**	**90300**	**5790**	**193**	**35**
D4GZY3	Tef2	0356	80391	300	8	5
D4GP28	2-keto-3-deoxyxylonate dehydratase	B0027	31842	35	19	1
D4GS50	Bacterio-opsin activator-like protein	0513	24295	26	12	1
D4GZ93	UreA	0149	14136	25	11	1

Protein Accession, from the UniProt database, e.g., D4GYK9; predicted MW, predicted molecular weight (Da) of the protein sequence identified by MASCOT; MASCOT score, MASCOT score associated with protein identification. Ions score is -10*Log(P), where P is the probability that the observed match is a random event. Individual ions scores > 26 indicate identity or extensive homology ($p<0.05$); number of peptides, no. of peptides associated with protein identification by MASCOT;. Peptide sequences, the number of distinct peptide sequences associated with the protein identified by MASCOT. DNA and RNA processing enzymes are highlighted in bold; proteins with a MASCOT score below 26 are shown in grey.

In the 'Wash' lane the DNA and RNA metabolism proteins RadA (HVO_0104), RpoA1 (HVO_0349), GyrA (HVO_1573) and UvrA (HVO_0393) were identified. After loading of cell lysate onto the gravity column packed with Strep-Tactin Sepharose beads, the column was washed with 5 x 0.5 ml Buffer B (20 mM HEPES pH 7.5, 2 M NaCl, 1 mM PMSF). The proteins found in the 'Wash' are included in the analysis because the Strep-Tactin Sepharose beads used were old and likely had reduced affinity for the StrepII-tagged protein. Repetition of this preliminary assay with fresh reagents is urgently needed to conclusively confirm if these protein interactions are indeed true. In elutions 1 and 2 RecJ4 (HVO_2889) was identified. All potential protein:protein interactions found are described below.

RadA (HVO_0104)

RadA is the archaeal recombinase that catalyses strand exchange during homologous recombination. With the aid of RadB, RadA forms a nucleoprotein filament on ssDNA that can then bind to dsDNA molecules and search for a region of homology (Wardell 2013). Once homology is found, RadA catalyses strand invasion and D-loop formation (Kil et al 2000).

RpoA1 (HVO_0349)

RpoA1 is a DNA-directed RNA polymerase subunit A'. In archaea, RpoA is the largest subunit of the RNA polymerase is divided into two polypeptides, A′ and A″ subunits, which are encoded by separate genes in an operon (Jun et al 2011, Langer et al 1995).

GyrA (HVO_1573)

DNA gyrase is the topoisomerase II found primarily in bacteria and some archaea. It consists of two polypeptide subunits, GyrA and GyrB, which form a heterotetramer. The C-terminal domain of GyrA is thought to bind DNA and help mediate a positive superhelical wrap about the protein (Corbett et al 2004).

UvrA (HVO_0393)

In bacteria, the proteins UvrA, UvrB, UvrC and UvrD are responsible for nucleotide excision repair (NER). The UvrA2B heterotrimer detects DNA damage as distortions of the DNA helix. Interaction of UvrA2B with the lesion causes local unwinding (Zou & Van Houten 1999). Homologs of the bacterial UvrABCD system are found in *H. volcanii* and are thought to act in NER (Crowley et al 2006, Lestini et al 2010).

RecJ4 (HVO_2889)

H. volcanii contains four RecJ proteins: RecJ1-RecJ4. RecJ proteins are 5'→3' single-stranded DNA exonucleases that play roles in homologous recombination, mismatch repair and base excision repair. RecJ, in combination with the DNA helicase RecQ, produces ssDNA tails that are required to initiate recombination from a double-strand break (Han et al 2006). It is also proposed that RecJ is the bacterial and archaeal homologue of the eukaryotic CDC45 protein, which forms part of the CMG (CDC45, MCM, GINS) complex (Makarova et al 2012).

The validity of these interactions will be discussed in further detail in Section 6.4: Discussion.

6.4 Discussion

Development of Methods

Methods developed and employed in this chapter have been used successfully to purify proteins expressed natively in *H. volcanii*, and to investigate *in vivo* protein:protein interactions.

Prior to this study, recombinant proteins expressed in *H. volcanii* were His_6-tagged and purified by metal affinity chromatography. However, many naturally occurring histidine-rich proteins from *H. volcanii* co-purify with the tagged protein of interest and hinder protein:protein interaction studies. The utilisation of a StrepII-tag in combination with a His_6-tag has solved this problem. The optimal method to purify StrepII-tagged and His_6-tagged Hel308 was via Ni^{2+} metal affinity chromatography followed by Strep-Tactin purification. For Hel308 and other proteins mentioned in this chapter, the optimal orientation for tags are an N-terminal His_6-tag and C-terminal StrepII-tag. The use of N-terminal StrepII-tags and C-terminal His_6-tags recovered little to no protein during purification.

The use of StrepII tags for protein purification in *H. volcanii* could be developed further in the future. If a protein requires tagging at the N-terminus only, a tandem His_6 and StrepII could be utilised. Currently under development by Alexandra Schindl and Thorsten Allers are plasmid constructs containing the following N-terminal tags MHHHHHHHGTSG**WSHPQFEK**GGSG**WSHPQFEK**GGDM (His_6 tag shown by underlining, StrepII tags shown in bold, linkers shown in uppercase only). Two tandem StrepII tags are included in this array to improve the affinity and stringency of protein purification.

The *tnaA* tryptophanase gene was deleted to reduce the degradation of added tryptophan during tryptophan-induced gene expression from the *tnaA* promoter, and thereby increase the efficiency of recombinant protein production. The development of improved expression strains by the deletion of

tnaA was successful, however due to time constraints these strains could not be tested for their efficacy in protein purification protocols. Before use, these strains should be tested by comparison with strains not deleted for *tnaA*, to measure protein expression at various levels of tryptophan induction (1 mM to 4.5 mM). Alternatively, the level of induction could be measured by placing a reporter gene such as *bgaH* under control of the tryptophan inducible promoter.

Interaction partners of Hel308

A preliminary protein:protein interaction assay identified five DNA/RNA metabolism proteins as possible interaction partners of Hel308: RadA, RpoA1, GyrA, UvrA and RecJ4. However, repetition of this assay is urgently required to validate these findings. A large amount of flow over can be seen in the lysate, pellet, flow-through and wash lanes in the protein pull down gel, (Figure 6.9). Many bands in the Elution 1 lane are of comparable molecular weights to ones seen in the Wash lane, implying that either the stringency of the Strep-Tactin beads is poor or that flow over has occurred between these lanes also. Therefore, the likelihood that the proteins identified in bands A, B and C are true interaction partners of Hel308 is poor. Additionally, most proteins identified in this assay appear to be running at a higher molecular weight in figure 6.9. For example the molecular weight of RadA is ~ 37 kDa however the band in which radA was identified from was ~ 150-200 kDa, indicating that protein identification in this assay may not be reliable. Furthermore, RadA, RpoA1, GyrA, UvrA and RecJ4 were only identified by a few (2 to 6) peptide sequences, Table 6.2. A high number of peptides and high peptide coverage is desirable when confirming protein;protein interaction partners using mass spectrometry.

If these interaction partners were confirmed to be true in the future, this would have interesting implications for Hel308.

For example, an interaction with RecJ could indicate that Hel308 (a RecQ family helicase) could play a similar role to the bacterial RecQ helicase, which acts in concert with RecJ 5'-3' exonuclease to generate ssDNA intermediates during DNA repair (Morimatsu & Kowalczykowski 2014). In homologous

recombination, the ssDNA tails formed by DNA end resection (by RecQ and RecJ) are coated with the RecA-family recombinases to generate a nucleoprotein filament. Since Hel308 is proposed to have a role in the regulation of homologous recombination (possibly through the interaction with the recombinase mediator RadB, see Chapter 4; Genetic analysis of *hel308*) the interaction of Hel308 and RecJ places Hel308 in the correct time and place for this regulation of homologous recombination to occur. More interestingly, RecJ is proposed to be the archaeal counterpart to the eukaryotic CDC45 protein within the DNA replication initiation CMG complex (Makarova et al 2012). Further evidence to support this proposal is that in *H. volcanii*, RecJ4 was seen to co-purify with the replication initiation protein Orc1 (Darya Ausiannikava, University of Nottingham, personal communication). The interaction of Hel308 and RecJ4 places Hel308 at the site of DNA replication.

Furthermore, an interaction with RpoA1 and UvrA could mean that Hel308 could act during transcription-coupled repair (TCR). TCR is widely described in bacteria and eukaryotes but less so in archaea (Deaconescu et al 2006, Eisen & Hanawalt 1999, Svejstrup 2002). *H. volcanii* contains a TCR mechanism that is not dependent on UvrA (Stantial et al 2016). During TCR, the RNA polymerase arrests upon encountering a DNA lesion, which is then repaired via nucleotide excision repair (NER) using the UvrABC complex (Hanawalt & Spivak 2008). Since Hel308 is proposed to act as a regulator of recombination, Hel308 could act at the site of arrested RNA polymerases during TCR (or with UvrA at DNA lesions outside TCR), to promote DNA repair via NER rather than by homologous recombination.

6.5 Future Perspectives

What is the role of Hel308?

The preliminary protein:protein interaction assays employed in this study indicate the potential to uncover possible protein interaction partners of Hel308 in the future.

To confirm the possible protein interactions in this study, the protein:protein interaction assays require repetition with fresh Strep-Tactin Sepharose alongside an empty vector control. Only then can more reliable determinations of the role of Hel308 be made.

The use of zero length cross-linking agents such as EDC and NHS may also aid in the discovery of transient interaction partners of Hel308 (Grabarek & Gergely 1990). Protein:protein interaction assays in this study were performed under DNA damaging conditions (MMC) and it would be of interest to compare these to results obtained under native conditions.

Once confirmed by protein:protein interaction assays, it would be of interest to delete the genes encoding these interacting partners on combination with *hel308*, and then determine growth, DNA damage and recombination phenotypes as described in Chapter 4: Genetic Characterisation of Hel308. This analysis would aid in understanding the role of Hel308 in DNA repair.

6.6 Conclusion

Strains and plasmids have been developed to allow for the overexpression and purification of halophilic proteins in *H. volcanii*. These methods have also been adapted for *in vivo* protein:protein interaction studies.

Chapter 7: Phylogenetic and Genetic Analysis of *hel308b*

7.1 Background

In *H. volcanii* a second *hel308* gene is present in addition to the canonical *hel308*, which is described in detail in Chapters 4-6. The second Hel308 helicase is termed Hel308b, it is 639 amino acids in length (Hel308 is 827 amino acids in length) and the *hel308b* gene (HVO_0971) is found at bp 880821-882740 in the *H. volcanii* genome. *hel308b* is a previously unstudied gene.

Duplicates of genes are found in abundance across all three domains of life (Zhang 2003). Gene duplication can result from unequal crossing-over, retroposition or chromosome/genome duplication. Unequal crossing-over tends to generate tandem gene duplications i.e. duplicated genes that are linked on the chromosome and depending on the position of the crossing-over the duplicated region can contain part of a gene, an entire gene, or several genes. In retroposition, mRNA is retrotranscribed to cDNA and then inserted into the genome, this process results in a loss of introns and regulatory sequences and a presence of polyA tracts and flanking short direct repeats (Brosius 2003). Due to the lack of regulatory elements, genes duplicated in this way tend to become pseudogenes. A pseudogene is a DNA sequence derived from a functional gene but has been rendered non-functional by accumulation of mutations (Weiner et al 1986). The duplicated gene generated by retroposition is usually found to be unlinked from the original gene as gene insertion happens at random. Chromosomal or genome duplication can when daughter chromosomes separate incorrectly following DNA replication.

Many duplicated genes are lost from the genome, however some duplicated genes do become fixed. Since duplication generates functional redundancy, eventually mutations will accumulate in one of the gene copies creating a pseudogene, which will either, be deleted from the genome or become diverged

from the original gene. If the presence of the duplicated gene is beneficial to the cell then the gene is fixed. In this scenario the two paralogues will have similar sequences and functions, this is termed concerted evolution (Elder & Turner 1995). If the presence of extra gene product from the duplication event is not advantageous then the two genes with identical functions are unlikely to be maintained within the genome (Nowak et al 1997). Here the duplicates will be stably maintained if they each adopt differing functions, for example each daughter gene could adopt part of the function of the original gene, this is called sub-functionalisation (Hughes 1994). Another outcome of gene duplication is that the generation of a gene with a novel function, this is called neo-functionalisation (Rastogi & Liberles 2005).

It is likely that *hel308b* was generated from a gene duplication event from the parental (canonical) Hel308 within *H. volcanii*. Gene duplication in the history of Hel308 has already been documented, mammalian *polQ* encodes a DNA helicase and a DNA polymerase which show similarity to *hel308* and *polN* respectively, it is thought that the similarity of *polQ* and *hel308* arose from a ancient duplication event (Marini et al 2003).

7.2 Aims

Very little is currently known about the cellular role of Hel308b in *H. volcanii*, to elucidate the function of Hel308b the following steps were undertaken:

- Analyse the protein sequence of Hel308b to identify protein structure and conserved domains, which may predict the function of the protein.
- Analyse the phylogenetic lineage of *hel308b* and its homologs. Examine which species it is found in. Also analyse its gene neighbourhood location in order to identify any conserved genes located in the vicinity of *hel308b*.
- Analyse the effect of *hel308b* deletion on cell growth to determine if it is required during normal cellular growth. Examine the effect of *hel308b* deletion on DNA content and cell size, to examine if it has an effect on

DNA replication.

- Since *hel308* has characteristic DNA damage phenotypes and recombination defects, the role of *hel308b* in DNA repair using DNA-damage assays and recombination assays was studied. Analyse the expression pattern of *hel308b* to determine if it is up-regulated in response to DNA-damaging agents.
- Carry out an analysis of *hel308b* deletion in combination with *hel308* and analyse the strains for growth rate and DNA-damage sensitivity.

7.3 Results

7.3.1 Phylogenetic analysis of Hel308b

Domain analysis of Hel308b

The domain structure of Hel308b was analysed by protein sequence alignment with Hel308 from *H. volcanii* and other halophilic archaea. Alignment of the Hel308b and Hel308 was carried out in MacVector using ClustalW (Gonnet; penalty for open gap = 10; extend gap = 0.2), Figure 7.1.

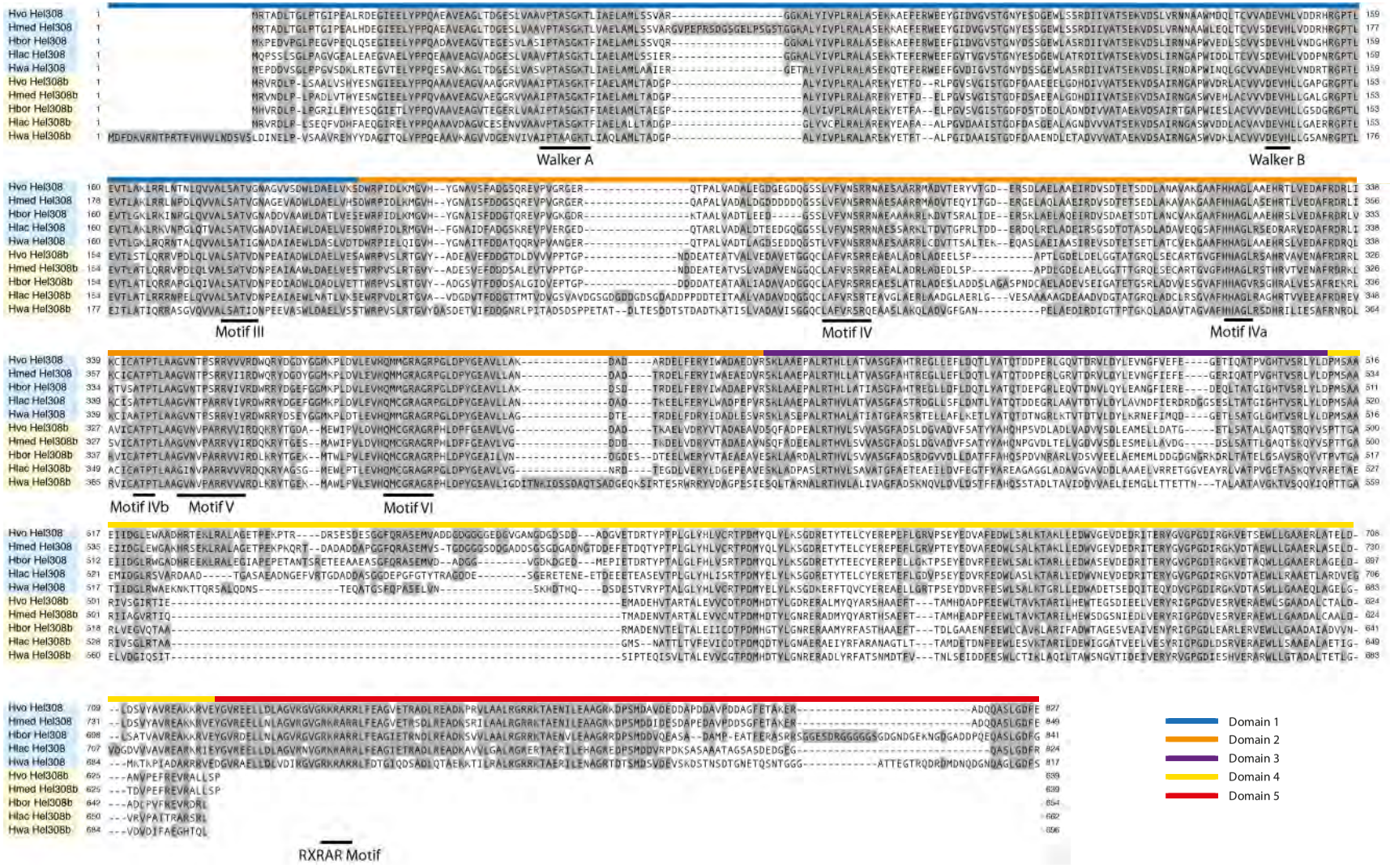


Figure 7.1: **Hel308b protein sequence alignment**. Multiple sequence alignment of Hel308b and Hel308 from *Haloferax volcanii*, *Haloferax mediterranei*, *Halorubrum lacusprofundi*, *Haloquadratum walsybi* and *Halogeometricum borinquense*. Hel308b shown in yellow and Hel308 shown in blue. Hel308 motifs and domains are indicated. Alignment carried out in MacVector using ClustalW (Gonnet; penalty for open gap = 10; extend gap = 0.2)

Hel308b shows an overall high level of sequence similarity to Hel308b from other halophilic archaeal species as well as to Hel308.

Domain 1

Hel308b contains the conserved Walker A and B ATPase motifs, GXXXXGKT/S where X can be any residue and DEAD/DEVH respectively. Walker A and B motifs have been discussed in further detail in Chapter 5: *Genetic Analysis of hel308 Point Mutants*, Section 5.3.1: *K53R Walker A and D145N Walker B mutations*. The conservation of these motifs suggests that Hel308b has the ability to bind and hydrolyse ATP.

Domain 2

Hel308b also contains the domain 2 conserved Motifs IV, IVa and IVb, as previously discussed in Chapter 5: *Genetic Analysis of hel308 Point Mutants*, Section 5.3.2: *Domain 2 mutations F316A, H317G and E330G,* these motifs make numerous contacts with the the unwound 3' DNA tail as it passages through the central pore of the helicase. It is proposed that the position of the ratchet helix is modulated by an ATP dependent movement in domain. The helicase motifs V and VI in domain 2 are also conserved in Hel308b, upon binding of the γ-phosphate of ATP a conserved motif VI arginine is thought to drive conformational closure of domains 1 and 2 (Buttner et al 2007). The conservation of these domain 2 motifs suggests that Hel308b could be a fully functioning helicase.

Domain 3

Domain 3 shows a high level of conservation between all Hel308b and Hel308 sequences. This domain houses a non-canonical winged helix (WH) fold, comprising four α- and two parallel β- strands. WH domains are known binders of nucleic acids, the WH fold in Hel308 may interact with an extended lagging strand of branched or forked DNA, WH domains have also been implicated in protein-protein interactions (Deng et al 2007, Woodman & Bolt 2011). Due to high conservation, the WH fold in Hel308b may act in a similar way to that of Hel308.

Domain 4

A large stretch of poorly conserved residues is found in domain 4 between all Hel308b and Hel308 sequences. Domain 4 been predicted to function as a helicase 'ratchet', which positions nucleic acids for translocation, this is linked to ATPase functions of the RecA folds in domains 1 and 2 (Woodman & Bolt 2011).

Domain 5

The most striking difference between Hel308b and Hel308 from this alignment is that Hel308b lacks the C-terminal domain 5, Figure 7.2. Domain 5 is an auto-inhibitory domain that couples ATP hydrolysis to DNA unwinding, after emerging from the central pore of Hel308 the 3' tail of unwound DNA binds a helix-link-helix (HLH) structure that contains a RXRAR motif in domain 5 via the phosphate-sugar backbone. The arginine residues in the RXRAR motif are cruicial to coordinating DNA binding to the rate of ATP hydrolysis and helicase unwinding (Richards et al 2008, Woodman et al 2007). The lack of domain 5 and the RXRAR motif suggests that Hel308b could have defects in DNA binding and therefore modulating ATP hydrolysis and helicase unwinding. Further details of the function of domain 5 can be found in Chapter Chapter 5: *Genetic Analysis of hel308 Point Mutants*, Section 5.3.4: *Domain 5 mutation R743A*.

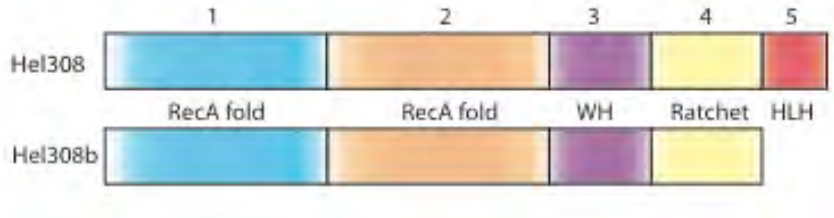


Figure 7.2: **Domain structure of Hel308 and Hel308b.** Hel308b has a highly similar domain structure to Hel308, however Hel308b lacks the 'auto-inhibitory' domain 5. Numbers indicate domains.

In order to further analyse Hel308b, Phyre2 was used to predict 3D structure (Kelley et al 2015). Phyre2 uses PSI-BLAST to analyse sequences, this alignment is then compared to alignments generated from database of known 3D structures. This technique is able to predict 3D structure based on remote

homology, Figure 7.3. The overall structure of Hel308b shows conservation to domains 1-4 of Hel308. Consistent with the multiple sequence alignment of Hel308b and Hel308, Figure 7.1 a degree of structural disorder can be seen in domain 4 of Hel308b when compared to Hel308.

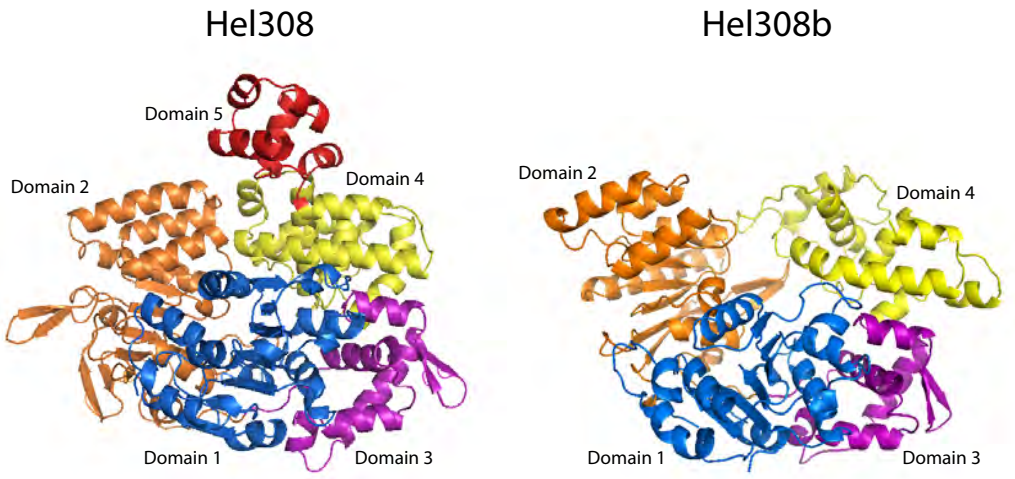


Figure 7.3: **Predicted structure of Hel308 and Hel308b from *H. volcanii*.** Structures predicted using Phyre2 (Kelley et al 2015). The overall arrangement of domains 1-4 in Hel308b is similar to Hel308.

In order to determine if the lack of domain 5 in Hel308b is genuine or due to a mis-annotation in the genome of *H. volcanii*, the genomic sequence following the stop codon of Hel308b was analysed Figure 7.4.

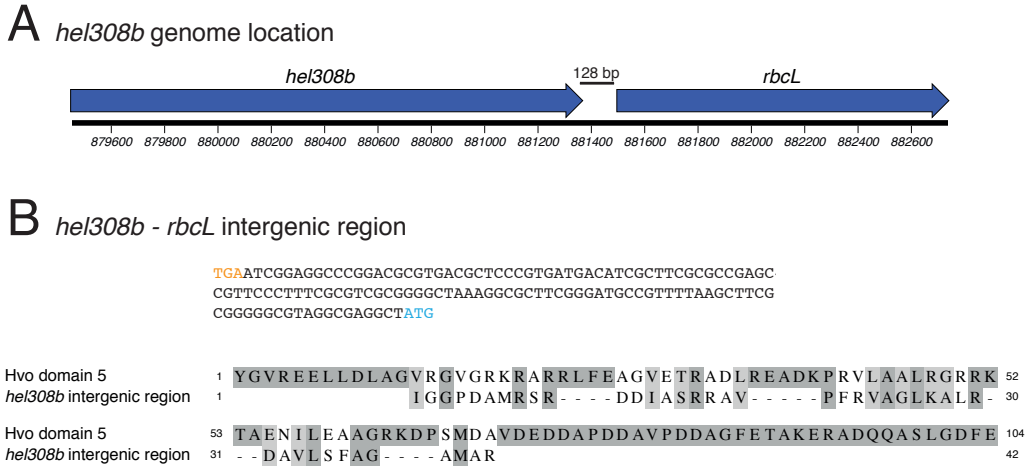


Figure 7.4: ***hel308b* genome location.** (A) Genome location of *hel308b*, showing downstream gene *rbcL* and 128 bp intergenic region. (B) *hel308b- rbcL* intergenic region, Stop codon of *hel308b* (TGA) shown in orange and start codon of *rbcL* (ATG) shown in blue. Intergenic region was translated *in silico* using the ExPASy Translate tool (Gasteiger et al 2003) and aligned to domain 5 of *H. volcanii*.

On the genome of *H. volcanii* a 128 bp intergenic region can be seen after the stop codon of *hel308b* and before the start codon of *rbcL* (ribulose bisphosphate carboxylase). Domain 5 of *H. volcanii* is 104 amino acids in length, which equates to around 312 nt of gene sequence. The nucleotide sequence of the intergenic region was taken and translated *in silico* using the ExPASy Translate tool (Gasteiger et al 2003) and aligned to Hel308 domain 5 of *H. volcanii*, carried out in MacVector using ClustalW (Gonnet; penalty for open gap = 10; extend gap = 0.2), Figure 7.4B. Very limited alignment is seen between the translated Hel308b – RbcL intergenic region and domain 5 of Hel308 and the intergenic region is too short to contain domain 5, indicating that the genome annotation is correct and Hel308b does not have a domain 5.

Distribution of *hel308b*

In order to determine the distribution of Hel308b, a BLAST search was carried out using the protein sequence from *H. volcanii* (Hel308b UniProt accession number: D4GV78). Hel308b is only found in the archaea and more specifically only in a small group of five closely related haloarchaeal species: *Haloferax volcanii, Haloferax mediterranei, Halorubrum lacusprofundi*, *Haloquadratum walsbyi* and *Halogeometricum borinquense*, Figure 7.5.

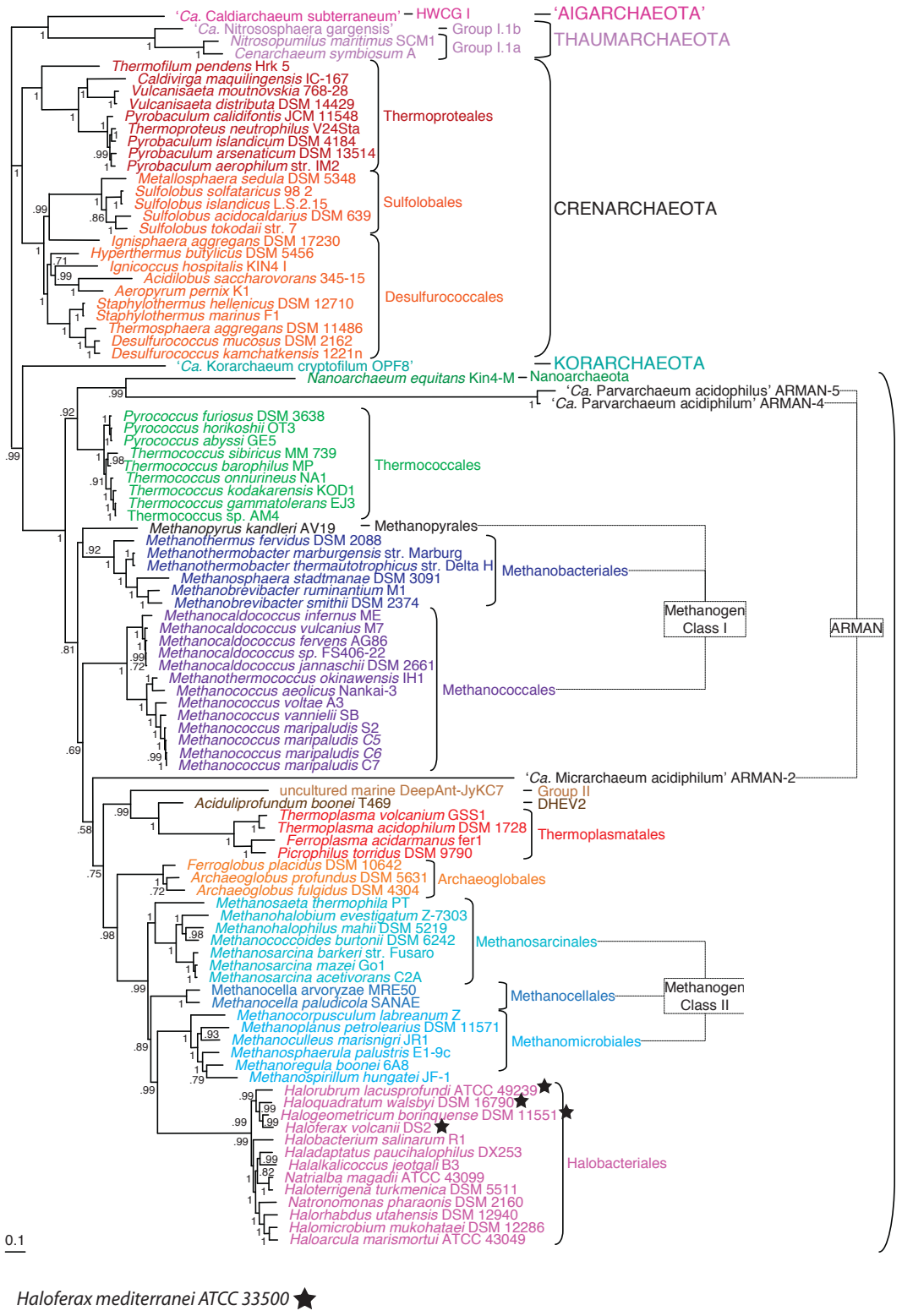


Figure 7.5: **Distribution of Hel308b amongst archaea.** Unrooted tree of the archaeal domain based on a concatenation of 57 ribosomal proteins. Hel308b is only found in five closely related halophiles, indicated by black stars. Hel308b is also found in *Haloferax mediterranei ATCC 33500* (a close relative of *H. volcanii)* however this strain was not included in this tree. Phylogenetic tree adapted from (Brochier-Armanet et al 2011). The scale bar indicates the average number of substitutions per site. Numbers at branches represent posterior probabilities.

A phylogenetic tree of Hel308 and Hel308b shows that Hel308b most likely arose out of an ancestral gene duplication event in the Halobacteriales (Figure 7.6).

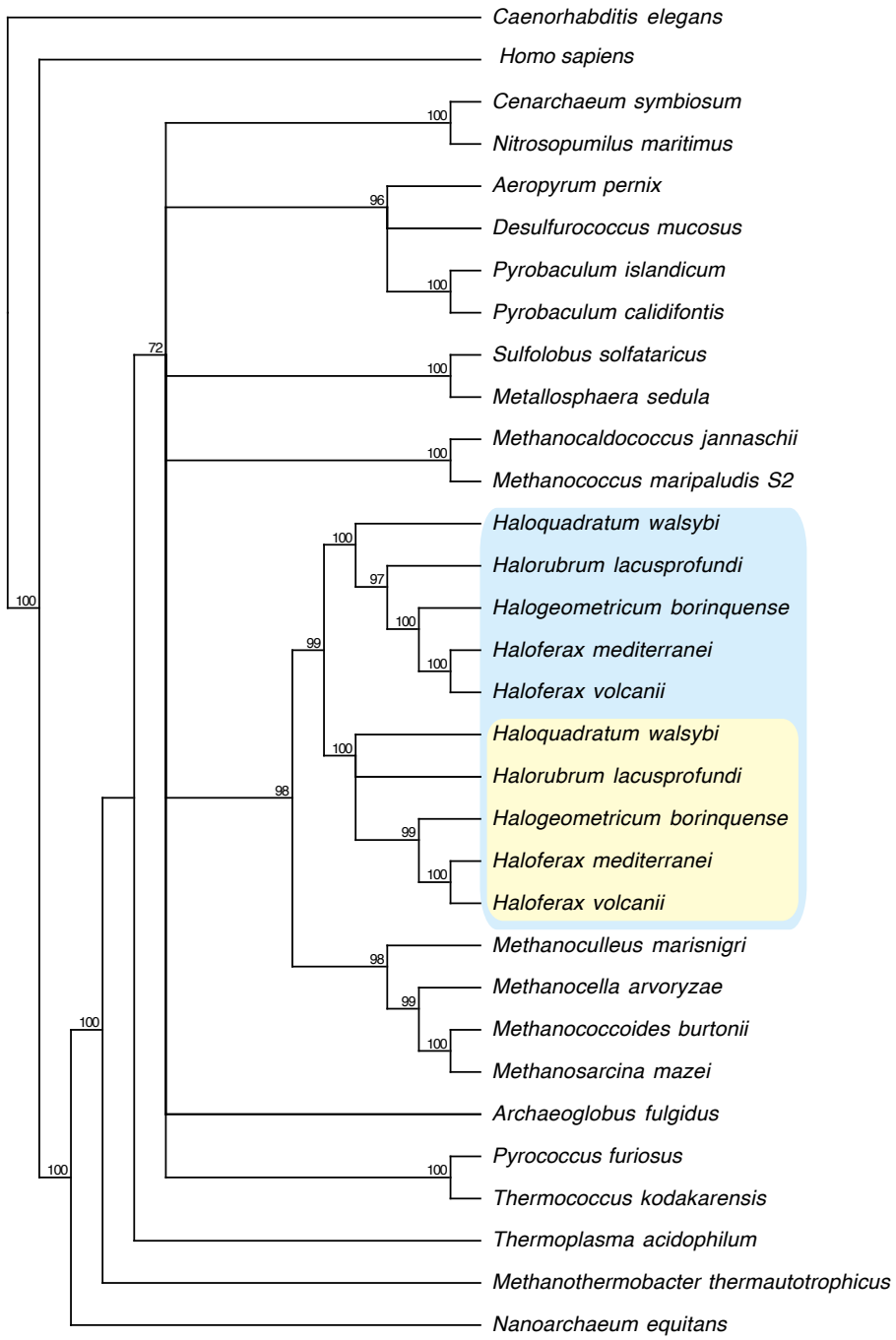


Figure 7.6: **Phylogenetic tree of Hel308 and Hel308b.** Rooted phylogenetic tree of Hel308 from Archaea, *Homo sapiens* and *Caenorhabditis elegans*. Tree rooted to *Caenorhabditis elegans*. Halophiles shown in blue and Hel308b is highlighted in yellow within the halophiles. Calculated using Neighbour Joining method (Bootstrap 1000 reps). Numbers above branches indicate the likelihood supporting the nodes (in percent). Generated using Macvector 12.0

Neighbourhood analysis of hel308b

hel308b (Hvo_0971) is found at position 880821-882740 on the genome of *Haloferax volcanii*. *hel308b* is not found in the vicinity of *hel308* (Hvo_0014) which has a location of 12388-14871 on the genome.

A gene neighbourhood analysis was carried out in order to predict the potential gene function of *hel038b*. Tight gene linkages often indicate that proteins function in the same pathway (Gabaldon & Huynen 2004, Korbel et al 2004, Wolf et al 2001).

Genes located 5 upstream and 5 downstream of *hel308b* were identified for the five halophilic species in which *hel308b* is present; *Haloferax volcanii, Haloferax mediterranei Halorubrum lacusprofundi*, *Haloquadratum walsbyi* and *Halogeometricum borinquense,* Figure 7.7. Halophiles are known to undergo a large amount of lateral gene transfer between species and between non-halophiles. Therefore, tightly associated gene linkage is more likely to be functionally relevant in halophiles compared to other species (Rhodes et al 2011).

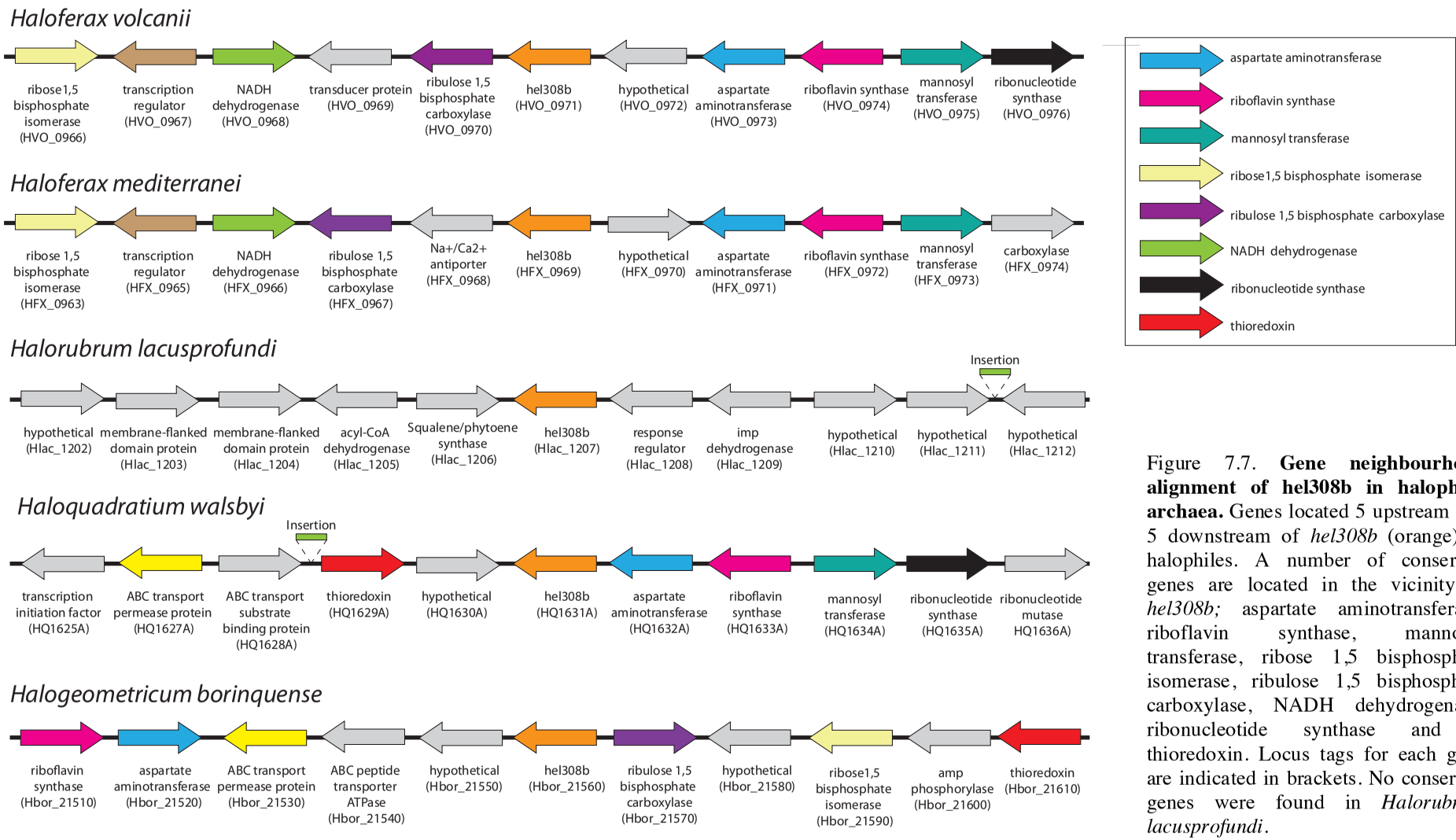


Figure 7.7. **Gene neighbourhood alignment of hel308b in halophilic archaea.** Genes located 5 upstream and 5 downstream of *hel308b* (orange) in halophiles. A number of conserved genes are located in the vicinity of *hel308b;* aspartate aminotransferase, riboflavin synthase, mannosyl transferase, ribose 1,5 bisphosphate isomerase, ribulose 1,5 bisphosphate carboxylase, NADH dehydrogenase, ribonucleotide synthase and a thioredoxin. Locus tags for each gene are indicated in brackets. No conserved genes were found in *Halorubrum lacusprofundi*.

A number of conserved genes were found to be in the vicinity of *hel308b* in the species *Haloferax volcanii, Haloferax mediterranei*, *Haloquadratum walsbyi* and *Halogeometricum borinquense*. These are aspartate aminotransferase (*aspC3*), riboflavin synthase (*ribH*), mannosyltransferase, ribose 1,5 bisphosphate isomerase, ribulose 1,5 bisphosphate carboxylase (*rbcL)*, NADH dehydrogenase (*nuoD*), ribonucleotide synthase (*purK)* and thioredoxin (*trxA3*). However, no conserved genes are found around *hel308b* in *Halorubrum lacusprofundi*.

aspC3-aspartate aminotransferase

Aspartate aminotransferase (AspC) is a multifunctional enzyme that catalyzes the synthesis of aspartate and the aromatic amino acids phenylalanine and tyrosine. AspC is catalytically active as a dimer (Fotheringham et al 1986, Gelfand & Steinberg 1977).

ribH-riboflavin synthase

Riboflavin synthase (ribH) also known as lumazine synthase in *Bacillus subtilis* is involved catalysing the conversion of GTP and ribulose-5-phosphate to riboflavin (Mack et al 1998, Ritsert et al 1995). Riboflavin synthase from *Schizosaccharomyces pombe* monomeric in the crystal structure (Gerhardt et al 2002). Riboflavin is part of the vitamin B group. It is the central component of the coenzyme flavin adenine dinucleotide (FAD) and flavin mononucleotide (FMN). FAD and FMN are electron carriers in a many redox reactions in various metabolic pathways (Oprian & Coon 1982)

Mannosyltransferase

Mannosyltransferases are involved in mannan biosynthesis, mannans are polysaccharides consisting of mannose units and are a key component in the yeast cell wall (Hall et al 2013, Hawkins 1973)

ribose 1,5 bisphosphate isomerase

In the archaea, ribulose-1,5-bisphosphate isomerase is involved in providing a

substrate for ribulose-1,5-bisphosphate carboxylase-oxygenase (RuBisCO) through an isomerization reaction for adenosine 5′-monophosphate (AMP) metabolism (Nakamura et al 2012, Sato et al 2007).

rbcL-ribulose 1,5 bisphosphate carboxylase

In archaea, *rbcL* is the ribulose-1,5-bisphosphate carboxylase-oxygenase (RuBisCO) which acts in the final step of AMP metabolism. In this step RuBisCO catalyzes the conversion of ribulose 1,5-bisphosphate CO_2, and H_2O to two molecules of 3-phosphoglycerate (3-PGA), which is an intermediate of central sugar metabolism. Genome sequences indicate that this pathway is distributed broadly among the Archaea, including all members of the *Thermococcales, Archaeoglobales, Methanomicrobiales, Methanosarcinales Halobacteriales, Methanococcales, Desulfurococcales, and Thermoproteales*. At the present, this pathway seems to be confined to the Archaea, since a complete set of genes cannot be found in any of the genomes of the Bacteria and Eucarya (Aono et al 2012).

nuoD-NADH dehydrogenase

In *Escherichia coli*, a 14 subunit complex (Complex 1) containing *nuoD* is responsible for coupling the oxidation of NADH to the generation of a proton motive force in the first enzyme complex of the respiratory chain (Falk-Krzesinski & Wolfe 1998). NuoD is involved quinone binding in complex I (Sinha et al 2015)

purK-ribonucleotide synthase

In prokaryotes, *purK* is one of the 12 enzymes that are involved in the biosynthesis of purine (Li et al 1999). *purK* is a member of the ATP-grasp enzymes that contain an atypical ATP-binding site (Fawaz et al 2011). *purK* catalyses the ATP dependent conversion of 5-aminoimidazole ribonucleotide (AIR) and HCO_3 to N5-carboxyaminoimidazole ribonucleotide (N5-CAIR) (Thoden et al 1999).

trxA3-thioredoxin

Thioredoxin is a disulfide oxido-reductase that catalyses a wide spectrum of redox reactions in the cell and is involved in the oxidative stress response in *Lactobacillus casei*. It is thought that *trxA3* is partly involved in maintaining the balance of cellular thiol and disulfide levels (Serata et al 2012).

Many of the conserved genes located in the neighbourhood *hel308b* are involved in biosynthesis of small molecules such as aspartate, aromatic amino acids, riboflavin, mannans, AMP and purine. Other conserved genes found in this neighborhood are involved in redox reactions as part of multisubunit enzymatic complexes such as *nuoD* and *trxA3*. This analysis suggests that *hel308b* could play a role in the regulation of small molecule metabolism or redox reactions within the cell. Alternatively (more likely), the presence of *hel308b* in this genome neighbourhood could be an accident of evolutionary history and without functional significance.

7.3.2 Expression of *hel308b*

Analysis of hel308b transcript levels

Expression of genes involved in DNA repair can be constitutive, or up-regulated in response to DNA-damaging agents. Expression of *hel308b* has not previously been studied, expression was measured following treatment with UV and mitomycin C (MMC), and these agents induce ss/dsDNA breaks and inter-strand crosslinking respectively.

Strains were grown to mid-exponential phase and either UV irradiated at 20 J/m^2 or incubated for 1 hour with 2 μg/ml MMC. Samples with no UV irradiation or MMC treatment were used as a control. This level of UV irradiation and MMC exposure results in a small amount of cell death. Following treatment, cells were resuspended in fresh growth media and allowed to recover. RNA was extracted and DNase treated to remove any contaminating DNA. End-point RT-PCR (reverse transcriptase PCR) was used to measure transcript levels of *hel308b* using the primers Hel308b and

Hel308bRTR which gives a product size of 297 bp. As a control, expression of *rpoA* (RNA polymerase subunit A) was analysed, as transcript levels of this gene do not change depending on the growth phase of the cell (Tom Batstone, University of Nottingham, personal communication). *rpoA* expression levels were analysed using the primers rpoARTR and rpoARTF (product size 328 bp). Results are shown in Figure 7.8.

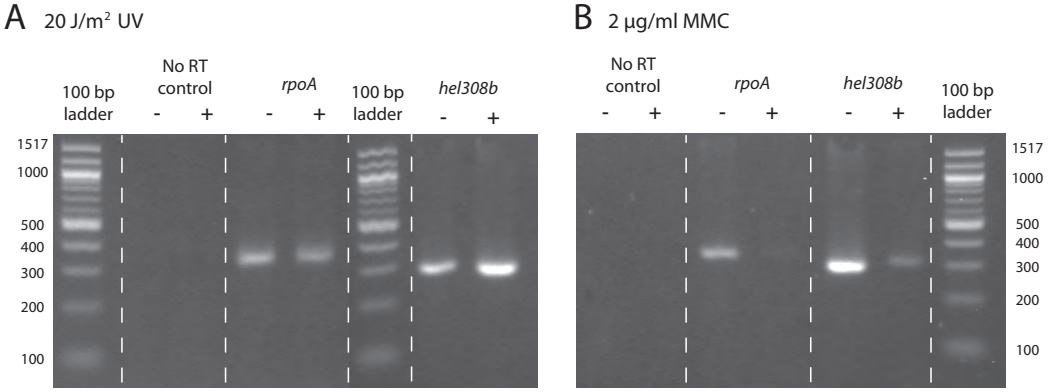


Figure 7.8: **RT-PCRs showing expression levels of *hel308b* and *rpoA* (control). (A)** Expression following treatment with 20 J/M2 UV-irradiation. *hel308b* expression does not change. **(B)** Expression following treatment with 2 µg/ml MMC. *hel308b* expression does not change. In all cases, '-' indicates no UV or MMC treatment, and '+' indicates treatment with UV or MMC.

This preliminary data shows that *hel308b* is indeed expressed. This demonstrates that although the existence of *hel308b* is probably due to a historical gene duplication event of *hel308* in halophilic archaea, *hel308b* has not lost its gene expression and therefore is likely not to be a pseudogene. No difference in the expression of *hel308b* was seen following treatment with UV-irradiation. A small reduction in expression can be seen following treatment with MMC, Indicating that *hel308b* could have some function in the repair of DNA crosslinks.

rpoA expression was used as a control for RT-PCRs as transcript levels of this gene do not change depending on the growth phase of the cell (Tom Batstone, University of Nottingham, personal communication). Results show that this control is not suitable for treatment with MMC, as the expression level of *rpoA* is drastically reduced following treatment.

7.3.3 Genetic Analysis of *Δhel308b*

A genetic analysis of *hel308b* was carried out. In order to determine whether *hel308b* was essential for cell viability, it was deleted from the genome of H26 to generate the strain H1843 (See Chapter 3: *Plasmid and Strain Construction* Section: *3.2.2.2*) Strains deleted for *hel308b* are viable; the deletion was confirmed by colony hybridisation, Southern blot and PCR.

Growth Rate

The growth rate of H1843 (*Δhel308b*) was qualitatively compared to that of H26 (*hel308b+*) by streaking onto complete media, Figure 7.9.

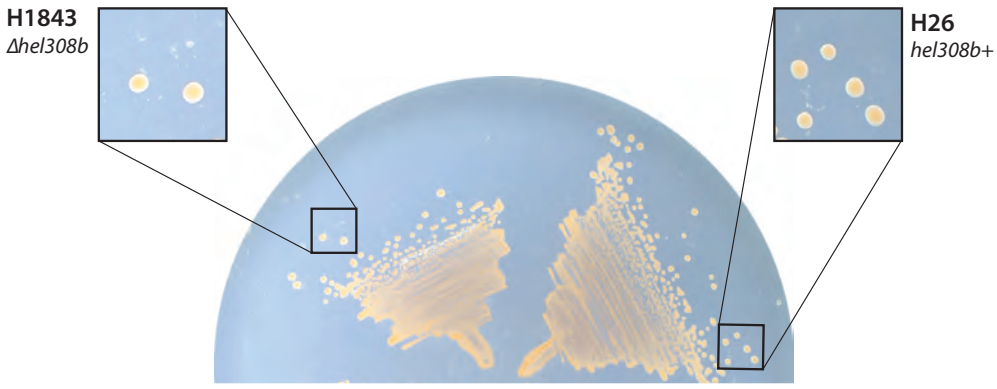


Figure 7.9: **Growth of a strain deleted for *hel308b*.** H1843 (*Δhel308b*) has no observable growth defect compared to wild type H26 (*hel308b+*).

No difference in growth can be seen between H1843 (*Δhel308b)* and wild type H26 (*hel308b+*). Strains with similar growth rates cannot be distinguished by this qualitative method, and for this reason growth assays in liquid culture were performed, Figure 7.10. The method for this assay has been described previously, Chapter 4: *Genetic Analysis of hel308*, Section: 4.3.2.2: *Deletion in combination with radB*.

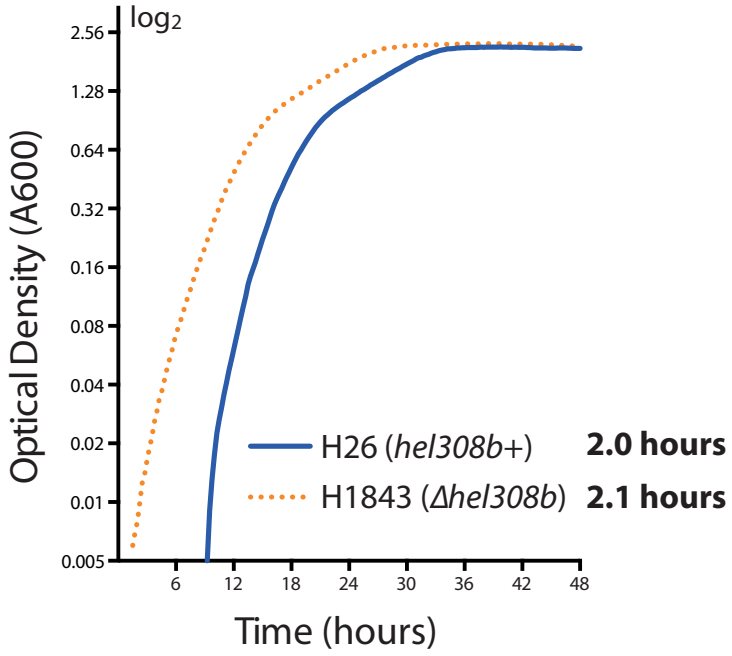


Figure 7.10: **Exponential growth rate of a strain deleted for *hel308b*.** Growth was measured by A_{600}. Generation time is indicated at the side of the strain name. H1843 (Δ*hel308b*) has a generation time similar to wild type H26 (*hel308b*+). Graph plotted on a $\log_2$ scale. 3 repeats were carried out for each strain. Generation time calculated from linear growth phase. All strains and repeats were incubated on the same 48 well plate and measured simultaneously using an Epoch 2 Microplate Spectrophotometer (BioTek).

H1843 (Δ*hel308b*) has a growth similar to wild type H26 (*hel308b*+) with generation times of 2.0 hours and 2.1 hours respectively. This suggests that *hel308b* is not essential for cell viability.

Survival Following Treatment with DNA-damaging Agents

Since *hel308* deleted strains are known to have a DNA damage phenotype after treatment with the DNA damaging agent MMC and the expression of *hel308b* was seen to decrease slightly after treatment with MMC, it was of interest to determine the survival of H1843 (Δ*hel308b)* after treatment with UV irradiation and MMC, Figure 7.11. These agents induce ss/dsDNA breaks and inter-strand crosslinking respectively, the method for these assays have been described previously Chapter 4: *Genetic Analysis of hel308*, Section: 4.3.2.2: *Deletion in combination with radB*.

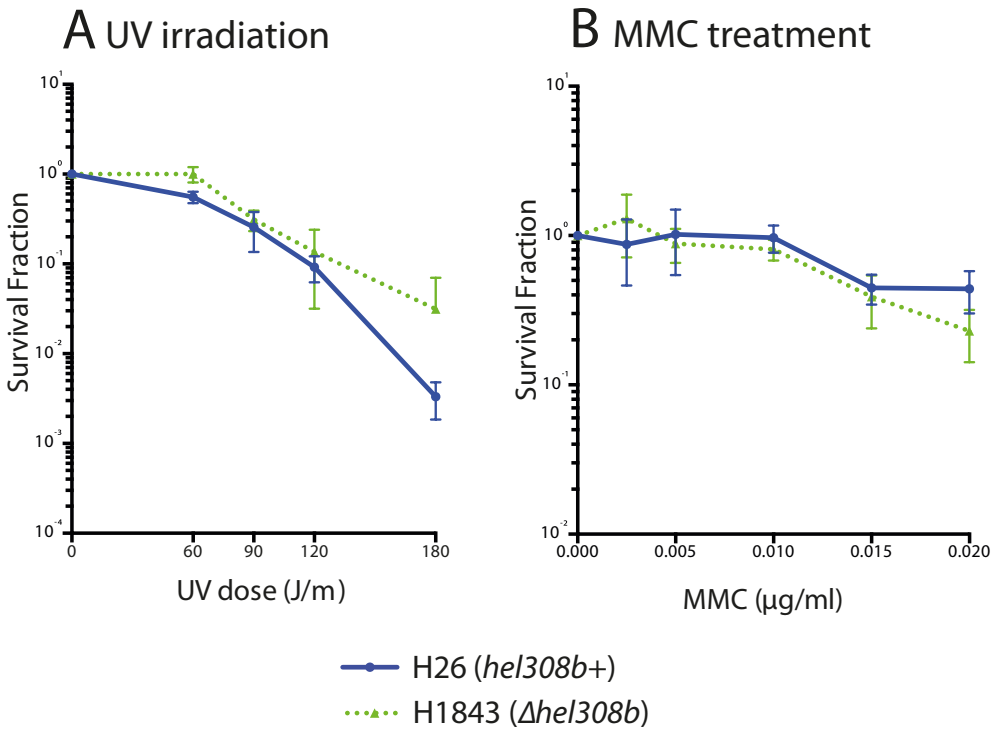


Figure 7.11: **Survival frequency of a strain deleted *hel308b* following treatment with DNA-damaging agents. (A)** Survival following treatment with UV irradiation. No diference in survival can be seen between H1843 (Δ*hel308b*) and wild type H26 (*hel308b*+). **(B)** Survival following treatment with MMC. No diference in survival can be seen between H1843 (Δ*hel308b*) and wild type H26 (*hel308b*+). Survival fraction is calculated relative to un-treated control. Each data point is generated as an average of at least 3 independent trials. Standard error is shown.

Following treatment with UV irradiation or MMC H1843 (Δ*hel308b)* has no significant difference in survival compared to wild type H26 (*hel308b*+), (P-values > 0.05, calculated from a two-tailed *t*-test). Indicating that *hel308b* has no involvement in the repair of DNA breaks or DNA crosslinks.

DNA Content and Cell Size

In order to determine the effect of a *hel308b* deletion on DNA content and cell size, flow cytometry was used, Figure 7.12.

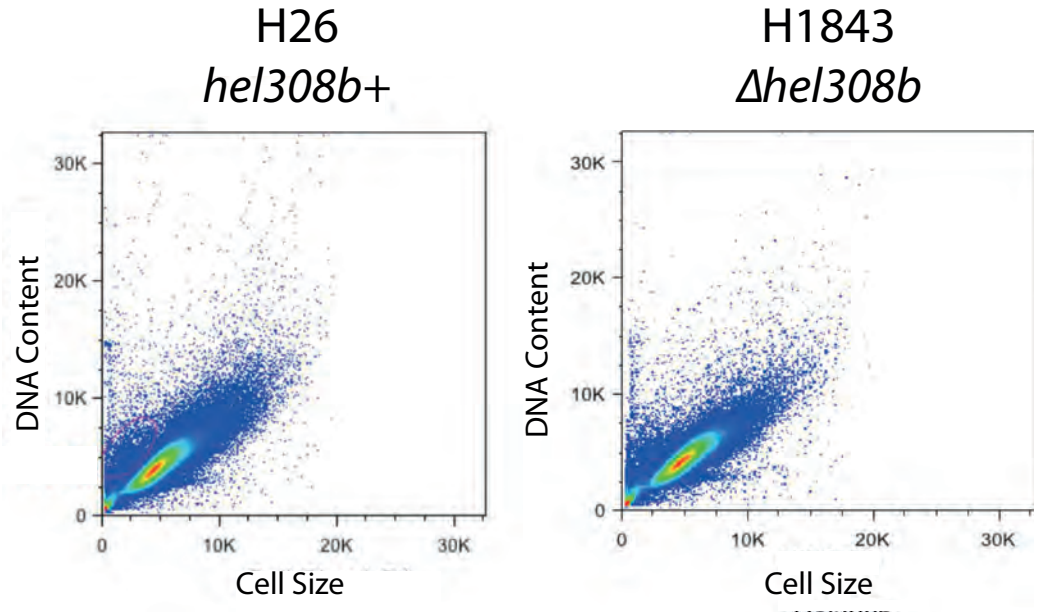


Figure 7.12: **Flow cytometry analysis of a strain deleted for *hel308b*.** Profiles show cell size vs DNA content. The profile for H1843 (*Δhel308b*) is identicle to wild type H26 (*hel308b+*).

No difference in the cell size vs DNA content profile for H1843 (*Δhel308b*) can be seen compared to wild type H26 (*hel308b+*). This indicates that the presence or absence of *hel308b* has no effect on DNA replication or cellular segregation.

Recombination Frequency

Since Hel308 has been suggested to play a role in the regulation of homologous recombination (Chapter 4: *Genetic Analysis of hel308*, Section: 4.3.2.2: *Deletion in combination with radB*) it was of interest to analyse the recombination frequency of *hel308b*. The method for this assay has been described previously; *Δhel308* has been included for comparason, Table 7.1.

Table 7.1: **Recombination frequency of a *Δhel308b* strain.**

Strain	**H164**	**H2117**	**H2007**
	hel308b+	*Δhel308*	*Δhel308b*
Recombination Frequency (RF)	4.94×10^{-5} (+/- 3.01×10^{-5})	3.23×10^{-5} (+/- 1.17×10^{-5})	7.52×10^{-6} (+/- 6.2352×10^{-7})
Transformation Efficiency (TE)	1.07×10^{-5} (+/- 3.25×10^{-6})	3.00×10^{-5} (+/- 0.00)	1.25×10^{-5} (+/- 5.26×10^{-6})
Relative recombination frequency (normalised by TE)	4.62×10^{0}	1.08×10^{0}	6.02×10^{-1}
	1×	**0.23×**	**0.13×**
Crossover fraction	13.49% (126)	8.75% (120)	6.67% (120)
Non-crossover fraction	86.51% (126)	91.25% (120)	93.33% (120)

Values in bold indicate the amount of recombination, crossover or non-crossover events compared to wild-type H164 (*hel308b+*). Values are generated as an average of at least 3 independent trials, +/- standard error is shown in brackets. Cells are shaded blue to indicate recombination defect and red to indicate hyper-recombination. Fraction of crossover and non-crossover events represented as a percentage, cells are shaded where values differ significantly from the wild type (P =0.05), blue indicates a decrease, red indicates an increase. Number of colonies assayed for crossover and non-crossover is indicated in brackets underneeth the percentages.

H2007 (*Δhel308b*) shows a low recombination frequency of 0.13× that of wild type H164 (*hel308b+*), this is lower than that of H2117 (*Δhel308*), which has a recombination frequency of 0.23×. Crossover and non-crossover events were observed to be significantly different to wild type (with two degrees of freedom with a chi-squared test) at 6.67% and 93.33% that of wild type respectively. These results suggest that *hel308b* may play a role in the regulation of homologous recombination.

7.3.4 Genetic Interactions of Hel308b

Since *Δhel308b* like *Δhel308* shows a recombination defect, it was of interest to investigate if *hel308b* and *hel308* acted in related pathways for DNA repair and regulation of homologous recombination. In order to study this, the strain H2488 (*Δhel308b Δhel308*) was generated. Strains deleted for *hel308b* and *hel308* are viable; the deletion was confirmed by colony hybridisation, Southern blot and PCR. For strain construction see Chapter 3: *Plasmid and*

Strain Construction, Section 3.2.2.2: *Generation of deletion strains for Chapter 7*.

Growth Rate

In order to compare the growth rate of H2488 (*Δhel308b Δhel308*) with H1843 (*Δhel308b hel308+*) and H1391 (*Δhel308*) strains were streaked onto solid media along side wild type H26 (*hel308b+ hel308+*), Figure 7.13.

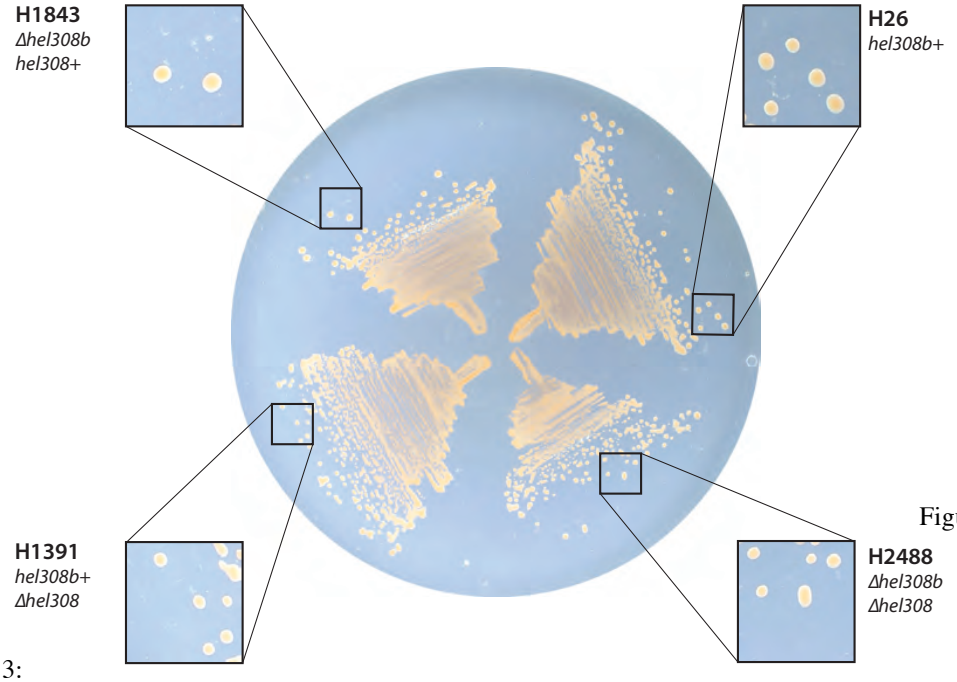


Figure 7.13: **Growth of a strain deleted for *hel308b* and *hel308*.** H2488 (*Δhel308b Δhel308*) has no observable growth defect compared to H1843 (*Δhel308b*), H1391 (*Δhel308*) and wild type H26 (*hel308b+*).

No difference in growth was observed between H2488 (*Δhel308b Δhel308*) has no observable growth defect compared to H1843 (*Δhel308b*), H1391 (*Δhel308*) and wild type H26 (*hel308b+ hel308+*). Indicating that *Δhel308b* and *Δhel308* are not synthetically lethal.

Strains with similar growth rates cannot be distinguished by this qualitative method, and for this reason growth assays in liquid culture were performed, Figure 7.14. The method for this assay has been described previously.

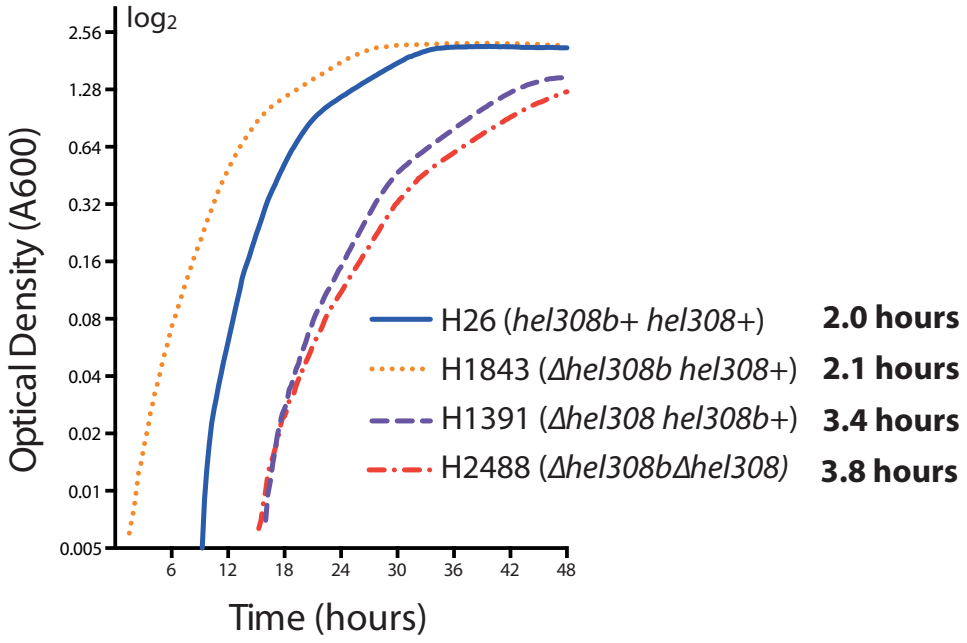


Figure 7.14: **Exponential growth rate of strains deleted for *hel308b* and/or *hel308*.** Growth was measured by A_{600}. Generation time is indicated at the side of the strain name. H1843 (*Δhel308b hel308+*) has a generation time similar to wild type H26 (*hel308b hel308+*). Whereas H2488 (*Δhel308b Δhel308*) has a generation time longer than H1391 (*Δhel308 hel308b+*). Graph plotted on a $\log_2$ scale. 3 repeats were carried out for each strain. Generation time calculated from linear growth phase. All strains and repeats were incubated on the same 48 well plate and measured simultaneously using an Epoch 2 Microplate Spectrophotometer (BioTek).

H1843 (*Δhel308b hel308+*) has a growth similar to wild type H26 (*hel308b+ hel308+*) with generation times of 2.0 hours and 2.1 hours respectively. However, in contrast to growth rates seen on solid media H2488 (*Δhel308b Δhel308*) has a generation time slower than that of H1391 (*Δhel308 hel308b+*) suggesting that *hel308b* and *hel308* could have a synthetic lethality.

Survival Following Treatment with DNA-damaging Agents

Strains deleted for *hel308b* do not have a sensitivity to treatment with MMC, however they have a slight increase in survival rate after treatment with high doses of UV. By constrast, *hel308* characteristically shows a sensitivity to MMC treatment but no sensitivity after UV irradiation. Given the differences

in survival, it would be of interest to determine the survival of H2488 (*Δhel308b Δhel308*) after treatment with DNA damaging agents, Figure 7.15.

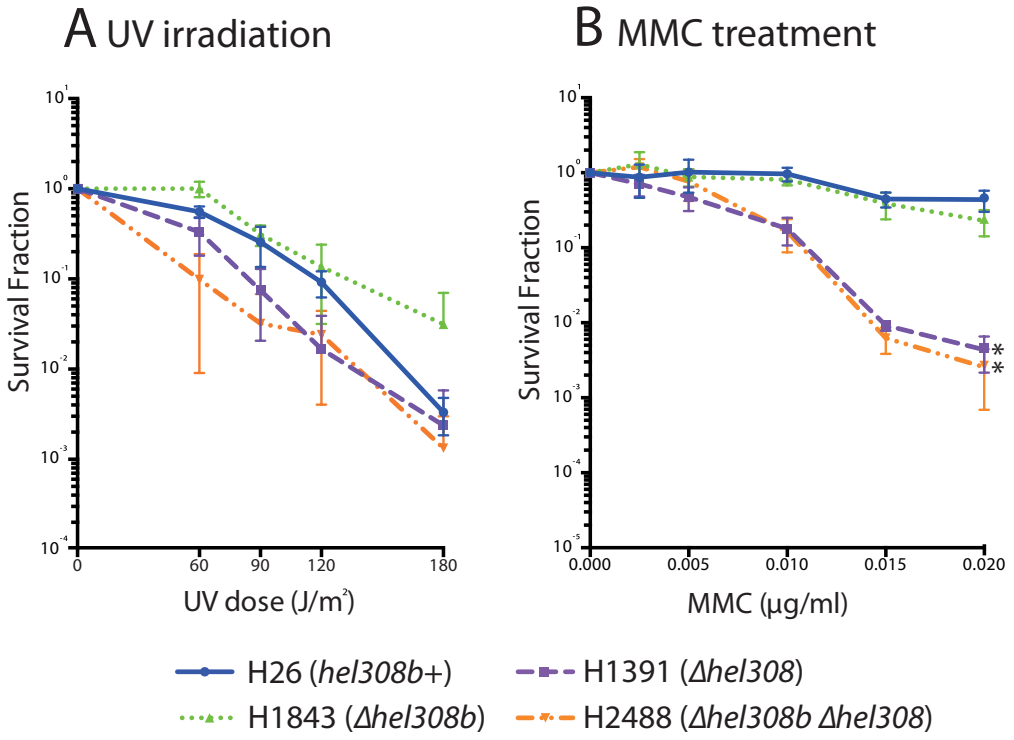


Figure 7.15: **Survival frequency of a strains deleted *hel308b* and *hel308* following treatment with DNA-damaging agents. (A)** Survival following treatment with UV irradiation. H2488 (*Δhel308b Δhel308*) has growth comparable to H1391 (*Δhel308*) and wild type H26 (*hel308b+ hel308+*). **(B)** Survival following treatment with MMC. H2488 (*Δhel308b Δhel308*) has growth comparable to H1391 (*Δhel308*). Survival fraction is calculated relative to un-treated control. Each data point is generated as an average of at least 3 independent trials. Standard error is shown. Asterisk (*) indicates that the highest dose of UV (180 J/m^2) or MMC (0.02 μg/ml) is significantly different to H26 (wild type) with $P < 0.05$. P-value calculated from two-tailed *t*-test in mutated strains compared to H26 (wild-type).

H2488 (*Δhel308b Δhel308*) shows a level of survival the same as H1391 (*Δhel308*), H1843 (*Δhel308b*) and wild type H26 (*hel308b+ hel308+*) following treatment with UV-irradiation. No significant difference is observed between these strains (P-values > 0.05, calculated from a two-tailed *t*-tests). After MMC treatment H2488 (*Δhel308b Δhel308*) shows a significant reduction in survival compared to wild type (P-value < 0.05, calculated from a two-tailed *t*-test). However, the survival shows no significant difference compared to H1391 (*Δhel308*) (P-values > 0.05, calculated from a two-tailed *t*-

test) suggesting that *hel308b* does not play a role in the repair of MMC induced DNA crosslinks in the presence or absence of *hel308*.

DNA Content and Cell Size

The cell size vs DNA content profile for H1843 (*Δhel308b*) is no different to wild type, however the profile for H1391 (*Δhel308 hel308b+*) shows an increased cell size and DNA content (Chapter 4: *Genetic Analysis of hel308*, Section: 4.3.2.2: *Deletion in combination with radB*). Due to these different profiles it is of interest to analyse the profile of H2488 (*Δhel308b Δhel308*), Figure 7.16.

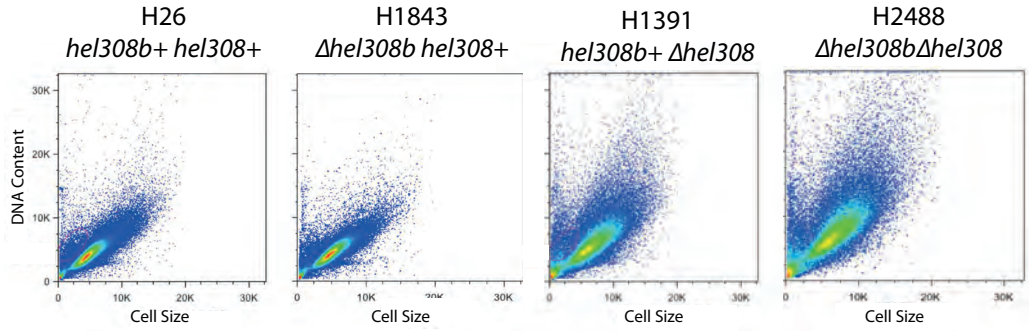


Figure 7.16: **Flow cytometry analysis of strains deleted for *hel308b* and *hel308*.** Profiles show cell size vs DNA content. The profile for H1843 (*Δhel308b*) is identicle to wild type H26 (*hel308b+*), H1391 (*Δhel308 hel308b+*) has higher cell sizes and DNA content compared to wild type and H2488 (*Δhel308b Δhel308*) has higher DNA content still.

The cell size vs DNA content profile of H2488 (*Δhel308b Δhel308*) is more scattered than that of H1391 (*Δhel308 hel308b+*) with a higher proportion of larger cell sizes and DNA content. This suggests that a *hel308b hel308* double deletion causes DNA replication and cell segregation defects that are worse than those seen in a *Δhel308* mutant.

Recombination Frequency

Since *Δhel308b* and *Δhel308* display severe recombination defects, it was of interest to analyse the recombination frequency of *hel308b* and it was of interest to analyse the recombination frequencey of *hel308* double deletion, Table 7.2.

Table 7.2: **Recombination frequency of *Δhel308b and Δhel308* strains.**

Strain	**H164**	**H2117**	**H2007**	**H2643***
	hel308b+ hel308+	*Δhel308*	*Δhel308b*	*Δhel308b Δhel308*
Recombination Frequency (RF)	4.94×10^{-5} ($+/- 3.01\times10^{-5}$)	3.23×10^{-5} ($+/- 1.17\times10^{-5}$)	7.52×10^{-6} ($+/- 6.2352\times10^{-7}$)	2.37×10^{-6} ($+/- 3.17\times10^{-6}$)
Transformation Efficiency (TE)	1.07×10^{-5} ($+/- 3.25\times10^{-6}$)	3.00×10^{-5} ($+/- 0.00$)	1.25×10^{-5} ($+/- 5.26\times10^{-6}$)	7.74×10^{-5} ($+/- 2.97\times10^{-5}$)
Normalised Recombination Frequency (RF/TE)	4.62×10^{0}	1.08×10^{0}	6.02×10^{-1}	1.87×10^{-2}
	1×	**0.23×**	**0.13×**	**0.04×**
Crossover fraction	13.49% (126)	8.75% (120)	6.67% (120)	8.75% (80)
Non-crossover fraction	86.51% (126)	91.25% (120)	93.33% (120)	91.25% (80)

*Recombination assay for this strain performed by Charlie Wickham-Smith (Doctoral Training Program rotation student) under my supervision.

Values in bold indicate the amount of recombination, crossover or non-crossover events compared to wild-type H164 (*hel308b+*). Values are generated as an average of at least 3 independent trials, +/- standard error is shown in brackets. Cells are shaded blue to indicate recombination defect and red to indicate hyper-recombination. Fraction of crossover and non-crossover events represented as a percentage, cells are shaded where values differ significantly from the wild type (P =0.05), blue indicates a decrease, red indicates an increase. Number of colonies assayed for crossover and non-crossover is indicated in brackets underneeth the percentages.

H2643 (*Δhel308b Δhel308*) has a severe recombination defect with a recombination frequency of 0.04 times that of wild type H164 (*hel308b+ hel308+*). This defect is worse than those seen with individual *hel308b* or *hel308* deletions, which have recombination frequencies of 0.13× and 0.23× that of wild type respectively. This result suggests that *hel308b* and *hel308* have a synthetic defect with regards to the regulation of homologous recombination. However, the crossover and non-crossover fractions of the *Δhel308b Δhel308* strain is not significantly different to wild type (with two degrees of freedom with a chi-squared test).

7.4 Discussion

Hel308b is not essential in *H. volcanii*; strains deleted for *hel308b* are viable and have growth phenotypes comparable to wild type strains. Double deletions of *hel308b and hel308* are also possible, and exhibit growth phenotypes no worse than strains deleted for *hel308* only.

By RT-PCR analysis it was confirmed that *hel308b* is constitutively expressed under both native and DNA damaging conditions, suggesting that the gene product of *hel308b* has a potential role in the cell and therefore is not a pseudogene.

Structure of Hel308b

Hel308b consists of four domains (domains 1-4) that show high sequence and structural homology to the canonical Hel308. However, the most notable difference between these two helicases is the absence of domain 5 in Hel308b, which is present in Hel308.

Domain 5 is proposed to be an auto-inhibitory domain in Hel308 that couples the hydrolysis of ATP to the unwinding of DNA using a conserved RXRAR motif. Mutations in Hel308 from *Methanothermobacter thermautotrophicus* have shown that the RXRAR motif (where X can be any residue) to be critical in the regulation of Hel308 helicase activity. Mutations in the first arginine result in increased ATPase activity, suggesting that the normal role of this amino acid is to restrain ATP hydrolysis until DNA is correctly engaged. Mutation of the central arginine leads to DNA binding defects and reduced DNA-stimulated ATP hydrolysis. By contrast, mutations in the final arginine completely abolish helicase actvity of Hel308 *in vitro* (Woodman & Bolt 2011, Woodman et al 2007). However, Hel308b lacks domain 5 entirely and therefore does not contain the RXRAR motif. This strongly suggests that Hel308b is unable to couple the hydrolysis of ATP to DNA binding and helicase-mediated unwinding. This could either mean that Hel308b is unable to function as a helicase, or that the helicase action of Hel308b is not regulated by an autoinhibitory mechanism. Since Hel308b lacks domain 5 and mutations

in the RXRAR motif of Hel308 can lead to the total loss of helicase activity, this suggests that Hel308b may not function as a helicase at all.

A crystal structure of the canonical Hel308 from *A. fulgidus* in complex with unwound DNA shows that the 3' tail of unwound DNA emerges from the central pore of the helicase and binds a helix-l-helix (HLH) structure in domain 5 via the phosphate-sugar backbone (Buttner et al 2007). The preferred substrate for Hel308 is a branched DNA structure and it is thought that this interaction is crucial for Hel308 to recognise replication fork structures (Fujikane et al 2005, Guy & Bolt 2005). The lack of domain 5 and therefore the HLH structure from Hel308b suggests that Hel308b is unable to recognise specific DNA substrates, in particular branched and replication fork-like structures. This potential lack of replication fork-like substrate -specificty suggests that Hel308b may play a different role in the cell to that of Hel308.

Distribution of *hel308b*

The *hel308b* gene is only found in five closely-related halophilic euryarchaeal species: *Haloferax volcanii, Haloferax mediterranei Halorubrum lacusprofundi*, *Haloquadratum walsbyi* and *Halogeometricum borinquense*. The presence of Hel308b in only closely-related species suggests that a gene duplication event of *hel308* is likely to have occurred in a common ancestor to these species rather than *hel308b* arising through lateral gene transfer. If *hel308b* arose through lateral gene transfer, then it would be present in a wider range of species.

hel308b is Found in a Semi-conserved Gene Cluster

Genes that are found in conserved neighbourhoods tend to have related or co-operative functions within the cell (Rhodes et al 2011). The relationship of genes found within gene clusters has previously been documented within *H. volcanii*. A phylogenetic study in *H. volcanii* and other archaea identified that the recombinase mediator *radB* is often found within a gene neighbourhood alongside the genes *rcrA* (Hvo_2384), *cdc48* (Hvo_2380) and genes coding for two hypothetical proteins, *ndnR* (Hvo_2382) and *hrp* (Hvo_2381) (Wardell 2013). All the identified genes in this neighborhood code for proteins involved

in DNA metabolism or protein turnover. This study also determined that RadB and RcrA physically interact. Furthermore, the putative resolvase NdnR is essential in a *radB*+ background, suggesting the interrelatedness of the two genes within this cluster.

In all species in which *hel308b* is present (apart from *Halorubrum lacusprofundi*), it is found in a loosely-conserved neighbourhood containing genes involved in the biosynthesis of small molecules such as aspartate, aromatic amino acids, riboflavin, mannans, AMP and purine. Other conserved genes found in this neighborhood are involved in redox reactions as part of multi-subunit enzymatic complexes such as *nuoD* and *trxA3*. Since the gene cluster shown in Figure 7.7 is not particularly conserved across all 5 haloarchaeal species and the conserved genes within this cluster do not share related functions, it is unlikely that the gene neighbourhood of *hel308b* can give insight into its function. It is more likely that the presence of *hel308b* within this gene neighbourhood is coincidental. For example, after a gene duplication event in the ancestor of the five haloarchaea, the *hel308b* gene might have been stable in this position and therefore persisted.

Hel308b is not Involved in Interstrand Crosslink Repair

Unlike *Δhel308* mutants, strains deleted for *hel308b* are not sensitive to DNA interstrand crosslinking agents such as MMC. This indicates that Hel308b does not play a role in the repair of DNA crosslinks and therefore suggests that the role of Hel308b within the cell is different to that of Hel308.

Hel308b as a Putative Regulator of Homologous Recombination?

A striking result of this study is the recombination defect observed upon the deletion of *hel308b* from *H. volcanii*. A *hel308b* deletion strain exhibits a recombination frequency of 0.13 times that of wild type. For reference, *hel308* deletion results in a recombination frequency of 0.23 times that of wild type. This significant reduction in recombination frequency suggests a role for Hel308b in the regulation and perhaps the promotion of homologous recombination in the cell.

Since Hel308b lacks domain 5 (which links ATP hydrolysis to helicase function), this observation suggests that the regulatory role of Hel308b and therefore Hel308 within homologous recombination could be independent of helicase function or unwinding. Furthermore, since a *hel308b* deleted strain is able to elicit a recombination defect, it suggests that the region of Hel308 that is essential for recombination regulation resides in domains 1-4.

Lastly, a double deletion of *hel308b* and *hel308* results in a recombination defect that is more severe than either of the singly-deleted strains, with a recombination frequency of 0.04 times that of wild type. The crossover fraction of the *hel308b* deleted strain is significantly decreased and the non-crossover fraction is significantly increased compared to wild type. However a *Δhel308bΔhel308* strain has crossover and non-crossover fractions comparable to wild type. The synthetic defect and differences in crossover and non-crossover fractions suggests that Hel308b and Hel308 act in separate pathways or at different points to regulate homologous recombination.

7.5 Future perspectives

Due to time constraints, a full genetic and biochemical characterisation of Hel308b was not possible. Many questions about the function and role of Hel308b within the cell remain, and future study should aid in the deeper understanding of Hel308b.

How does Hel308b function?

Since Hel308b lacks domain 5 and is therefore missing essential regulatory motifs that are key to helicase function, an important question to ask is whether Hel308b has helicase activity?

Biochemical analysis will be able to answer this question. Hel308b is only found in halophiles and halophilic proteins are unstable in low salt conditions, therefore routine helicase assays will need to be adapted to high salt conditions. Protocols for the expression and purification of halophilic proteins in *H. volcanii* is described in detail in Chapter 6: *in vitro* Analysis of Hel308. Helicase assays such as those described for Hel308 from

Methanothermobacter thermautotrophicus (Guy & Bolt 2005, Woodman et al 2007) will not only determine whether Hel308b is a helicase but its polarity and any substrate preferences.

Since Hel308b lacks domain 5, which is a key domain in the regulation of ATP hydrolysis, it will be interesting to determine if there are differences in ATPase activity between Hel308 and Hel308b. Simple colorimetric ATPase assays using malachite green will be able to determine the ATP turnover of each helicase. Malachite green molybdate interacts with free orthophosphate liberated during ATP hydrolysis to result in a yellow-to-green colour change, which can be measured at 620-640 nm (Feng et al 2011).

In which pathway does Hel308b act?

Revealing the protein:protein interactions of Hel308b will give an insight into which pathway it acts. Protein:protein interactions as determined by protein pull-down assays and mass spectrometry are described in Chapter 6: *in vitro* Analysis of Hel308. Interaction partners revealed using pull-down assays would be candidates for deletion in combination with *hel308b*. Complementation of biochemical findings with genetic characterisation will improve an understanding of the role of Hel308b in *H. volcanii*.

Because a striking recombination defect was seen in strains deleted for *hel308b,* it would be interesting to determine which growth, DNA damage repair and recombination defects are exhibited by strains deleted for *hel308b* and/or *hel308* in combination with the recombinase and recombination mediator *radA* and *radB* respectively. The phenotypes that are observed in these strains will give a better understanding of the roles that Hel308b and Hel308 play in the regulation of homologous recombination in *H. volcanii*.

7.6 Conclusion

Hel308b is the previously unstudied second Hel308 helicase found in *H. volcanii*. Structural analysis has determined that Hel308b lacks the domain 5 found in canonical Hel308 helicases, which is critical for coupling ATP hydrolysis to helicase function. *hel308b* is not essential and double *hel308b*

and *hel308* deletions are viable and do not exhibit growth defects. However, *hel308b* and double *hel308b* and *hel308* deletions show striking recombination defects, suggesting a role for Hel308b in the regulation of homologous recombination.

Chapter 8: Novel Haloviral DNA Processing Enzymes for the use in Nanopore DNA Sequencing Technologies

As part of a BBSRC-CASE studentship, the work carried out in this chapter was completed in collaboration with the industrial partner Oxford Nanopore Technologies.

8.1 Background

8.1.1 Haloviruses

In this study the term "halovirus'' will be used to describe viruses found in hypersaline systems, including bacterial, archaeal, and eukaryotic viruses.

Viruses are the most abundant reservoirs of nucleic acid-encoded information in the biosphere and outnumber cells 10 to 100 fold. This is also true for haloviruses, where counts in hypersaline waters are reported to be at least 10^7 viruses per ml and up to 2 x 10^9 per ml in crystallizer ponds (Aalto et al 2012, Dyall-Smith et al 2003, Guixa-Boixereu et al 1999). Although organisms from all domains of life are found in hypersaline environments, the majority of haloviruses to date infect only archaea. Just 9 bacterial and 5 eukaryotic haloviruses have been found to date (Atanasova et al 2015b). However, compared to bacteriophages and eukaryotic viruses, very little is known about archaeal haloviruses, this partially due to difficulties in cultivating haloarchaea that can acts as hosts.

To date only around 90 viruses have been described for halophilic archaea and of those approximately 50 have fully sequenced genomes (Atanasova et al 2015b, Luk et al 2014). At the start of this study in 2012, only 17 archaeal haloviral genomes were published and no bacterial or eukaryotic haloviral

genomes were known. All archaeal haloviral genomes are dsDNA, the majority are linear and genome sizes range from 7 kb to 143.9 kb (Pietila et al 2009, Sencilo et al 2013). No RNA haloviruses infecting archaea have yet been isolated (Pietila et al 2014). Most open reading frames (ORFs) of archaeal haloviruses lack any significant matches to sequences in databases; for example, the halovirus HF1 has 117 ORFs with 102 being unique (Tang et al 2004). Genes that are annotated have shown to have homology with organisms from all three domains of life, providing an explanation for the majority of archaeal haloviruses having highly mosaic genomes (Sencilo et al 2013, Tang et al 2002). Archaeal haloviruses have highly diverse genomes, studies have determined that the mutation frequency is 7.65×10^{-3} substitutions per nucleotide with a 24% higher mutation frequency in coding regions than in non-coding regions (Santos et al 2011, Santos et al 2010). Archaeal haloviruses have abnormally high mutation rates compared to DNA and RNA bacteriophages which have mutation rates of 10^{-8} to 10^{-6} and 10^{-6} to 10^{-4} substitutions per nucleotide respectively (Sanjuan et al 2010).

8.1.2 Halophilic proteins

There are two alternative strategies employed by halophilic organisms for maintaining an osmotic balance between intracellular and extracellular salt concentrations. The first is a 'salt out' strategy used predominantly by halophilic bacteria and eukaryotes, where salts are actively pumped out from the cell and the cytoplasm is packed with organic solutes such as glycerol or glycine betaine to maintain the osmotic balance (Christian & Waltho 1962, Oren 1999, Oren 2008). Halophilic archaea and a few bacteria maintain osmotic balance by accumulating high levels of salt in the cytoplasm this is termed a 'salt-in' approach (Oren et al 2002). Proteins in halophilic archaea have adapted to function in high salt and low water conditions by several different strategies (Mevarech et al 2000). Halophilic proteins tend to contain a high amount of acidic residues on their surface such as aspartic acid and glutamic acid, but contain a small amount of nonpolar residues. This generates an overall low isoelectric point (pI) and high density of negative charges on the

surface of the protein that will co-ordinate a network of hydrated cations, allowing the protein to stay soluble in solution (Lanyi 1974). Halophilic proteins have a reduced surface hydrophobicity by replacing large hydrophobic side groups with small hydrophilic ones. Some halophilic proteins have extra domains or peptide insertions that are extremely rich in acidic residues, which are essential for correct protein folding (Graziano & Merlino 2014). Halophilic proteins show high levels of activity in organic media and solvents, making them ideal candidates for use as biocatalysts and in other biotechnological applications (DasSarma et al 2010, Lanyi 1974). An extensive list of halophilic proteins and enzymes used in biotechnology can be found in (Wackett 2012).

8.1.3 Haloviral proteins

Due to limitations in genome size, viruses tend to have simpler proteins with fewer cofactors than their cellular counterparts. For example, polymerases are usually active as single subunits in viruses, but are composed of multisubunit complexes in their prokaryotic and eukaryotic hosts (Choi 2012).

Since viruses have to survive inside and outside of the protective host environment, their proteins are unusually robust. Archaeal and bacterial haloviruses have been shown to withstand being subjected to a wider range of salinities than their cellular hosts (Luk et al 2014). Haloviruses can endure rapid changes of considerable magnitudes of salinity that cellular halophiles cannot match. This is thought to be due to haloviral proteins having an extremely low pI combined with covalent cross linking that promotes protein stability not seen in cellular counterparts (Kukkaro & Bamford 2009, Pauling 1982).

The ‘stripped-down’ nature and hardiness of haloviral proteins could be exploited for use in biotechnology.

8.1.4 Nanopore sequencing

The first biological nanopore to be used as a biosensor to sequence nucleic acids was alpha hemolysin (α-HL) (Braha et al 1997). α-HL is a toxic protein secreted by bacteria in monomeric form that spontaneously inserts into lipid bilayers and assembles to form heptameric transmembrane channels (Gouaux et al 1994). The transmembrane stem section of the pore is a 5 nm β-barrel that has a ring of 14 alternating lysine and glutamic acids that form a 1.5 nm "limiting aperture" between the vestibule and the stem. This limiting aperture is large enough to accommodate a single molecule of single stranded DNA or RNA. When a nanopore is inserted into a synthetic membrane and a potential is then applied across the membrane, the current runs only through the aperture of the nanopore (Braha et al 1997). Single molecules, such as proteins, DNA or RNA that enter the nanopore cause characteristic disruptions in the current. The DNA or RNA strands to be sequenced are mixed with a processive 'motor' enzyme such as a helicase, nuclease or a polymerase. As the DNA:enzyme complex approaches the nanopore, the ssDNA is pulled though the aperture of the nanopore and the enzyme ratchets the DNA through the nanopore one base at a time. This ratcheting slows the translocation rate of the ssDNA through the nanopore down to the millisecond time scale allowing for better resolution of the signal produced (Deamer 2010). As the ssDNA passes through the pore, each base perturbs the electrical current in a characteristic way enabling the order of the bases on the DNA strand to be determined.

MinION nanopore sequencing technology developed by Oxford Nanopore is continually improving, however this technology currently has an error rate in base calling of around 12% to 40% (Goodwin et al 2015, Ip et al 2015, Laver et al 2015). This is speculated to be due to challenges in the signal processing of the ionic current measurements. The pores used in this technology are more than a single base in height so that the ionic signal measurements are not of individual nucleotides but of approximately 5 nucleotides at a time (Goodwin et al 2015). If any diffusion of the ssDNA within the pore occurs, the bases will be miscalled. To generate accurate base calling, diffusion of the ssDNA has to be kept to a minimum, this could be achieved by fine-tuning the way that the

ssDNA travels through the nanopore. A method to do this could be through the development of new 'motor' enzymes that ratchet the ssDNA through the pore. A candidate protein for this role would have to be a robust DNA processing enzyme that can withstand the salinity of the bathing solution within the nanopore flow cell.

8.2 Aims

The aim of this study was to discover novel DNA processing enzymes such as helicases and polymerases from novel haloviruses. These enzymes would hopefully be candidates for improved 'motor' proteins to ratchet ssDNA through the nanopore in Oxford Nanopore's DNA sequencing technology.
The following steps were carried out to achieve this aim:

- Sample for novel haloviruses from hypersaline waters and crystallizer ponds in Israel and Alicante, Spain.
- Isolate the viral fraction from the samples through host-dependent enrichment assays and by host-independent precipitation methods.
- Extract viral genomic DNA and RNA and sequence using Illumina MiSeq.
- Assemble sequencing reads and mine novel haloviral genomes for novel DNA processing enzymes such as helicases and polymerases.
- Express and purify helicases and polymerases in an *Haloferax volcanii* expression strain.
- Test suitability of the helicases and polymerases for use in Nanopore sequencing technologies.

8.3 Results

On the 13th and 14th October 2011, a total of 35 litres of water samples were taken from several sites at the Dead Sea, Israel and from evaporation pools at a salt works outside Eilat, Israel, Table 8.1 and Figure 8.1. Samples were taken at three geographically different locations along the west shore of the Dead Sea to capture haloviral diversity. Green algae was seen to be growing at the pool

at Mitspe Shalem suggesting that the water was brackish, this site was picked as mesohalic viruses may be present in this pool. Different hypersaline pools at the salt works in Eilat were chosen in order to collect samples with a range of high salt concentrations. Samples were collected in a plastic beaker and crudely filtered at the location through fabric to remove brine shrimp and debris. The samples were decanted into fresh jerry cans, sealed with parafilm and transported back to England via surface mail.

On the 29th January 2013 a 50 litre sample was collected from a hypersaline pond at the Bras del Port salterns, Alicante, Spain with help from Rodriguez-Valera laboratory at the University Miguel Hernandez, Alicante, Table 8.1 and Figure 8.1. The sample was collected in a plastic beaker and decanted into 2 x 25 litre jerry cans, sealed with parafilm and transported back to England by airmail.

Table 8.1: **Halovirus sampling locations.**

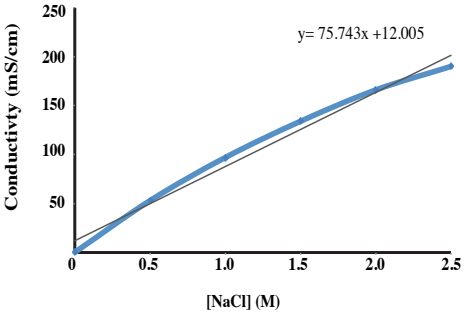
Sample ID	Country of Origin	Latitude	Longitude	Sample Size	Date Collected
Eilat 1	Israel	29º 33' 33" N	34º 57' 41" E	5 L	13th Oct 2011
Eilat 2	Israel	29º 33' 39" N	34º 57' 45" E	5 L	13th Oct 2011
Eilat 3	Israel	29º 33' 33" N	34º 57' 41" E	5 L	13th Oct 2011
Ein Gedi	Israel	31º 27' 35" N	35º 24' 00" E	5 L	14th Oct 2011
Kalya	Israel	31º 45' 41" N	35º 30' 13" E	5 L	14th Oct 2011
Mitspe Shalem	Israel	31º 34' 54" N	35º 24' 42" E	5 L	14th Oct 2011
Mitspe Shalem Pool	Israel	31º 34' 54" N	35º 24' 44" E	5 L	14th Oct 2011
Bras del Port Saltern ID - #30	Alicante, Spain	38º 11' 47.4"N	0º 35' 0.8" W	50 L	29th Jan 2013

Figure 8.1 **Halovirus sampling locations. (A)** Map of Israel with sampling sites in Eilat and the west shore of the Dead Sea highlighted by yellow boxes. **(B)** Mitspe Shalem Pool on the west shore of the Dead Sea. **(C)** Thorsten having a paddle and hunting halophiles in the Dead Sea. **(D)** Ein Gedi on the west shore of the Dead Sea. **(E)** An evaporation pool at the salt works, Eilat. **(F)** Aerial view of Bras del Port Salterns, Alicante, Spain. **(G-H)** Saltern #30 Bras del Port Salterns, Alicante, Spain.

8.3.1 Salinity of Sea Water Samples

As haloviruses from hypersaline environments were desired, the salinity of the water samples from each location was determined. A standard curve was made by measuring the conductivity of known concentrations of NaCl. The conductivity of samples was measured and NaCl concentration calculated using the equation of the line from the standard curve, (for ease of transportation some samples were pooled or split, this is indicated in the sample name), Figure 8.2. It was found that all water samples were hypersaline, ranging from 16.44 to 31.94 % NaCl.

The salinity of the water sample from the Bras del Port salterns, Alicante was 32%; this was measured using a hand refractometer.



Sample	mS/cm	[NaCl] M	% NaCl
Eilat 1 + 3	106	1.24	29.15
Eilat 2 # 1	87	0.99	23.26
Eilat 2 # 2	85	0.96	22.64
Ein Gedi + Kalya	115	1.36	31.94
Mitspe Shalem	110	1.29	30.39
Mitspe Shalem Pool	65	0.70	16.44

Figure 8.2: **Salinity of Sea Water Samples.** The conductivity (measured in mS/cm) of known concentrations of NaCl solutions were measured using an AKTA Prime Plus (Amersham Biosciences) to generate a standard curve. From this the salinity of the sea water samples were calculated.

8.3.2 Viral Enrichment and Viral Plaque Assays

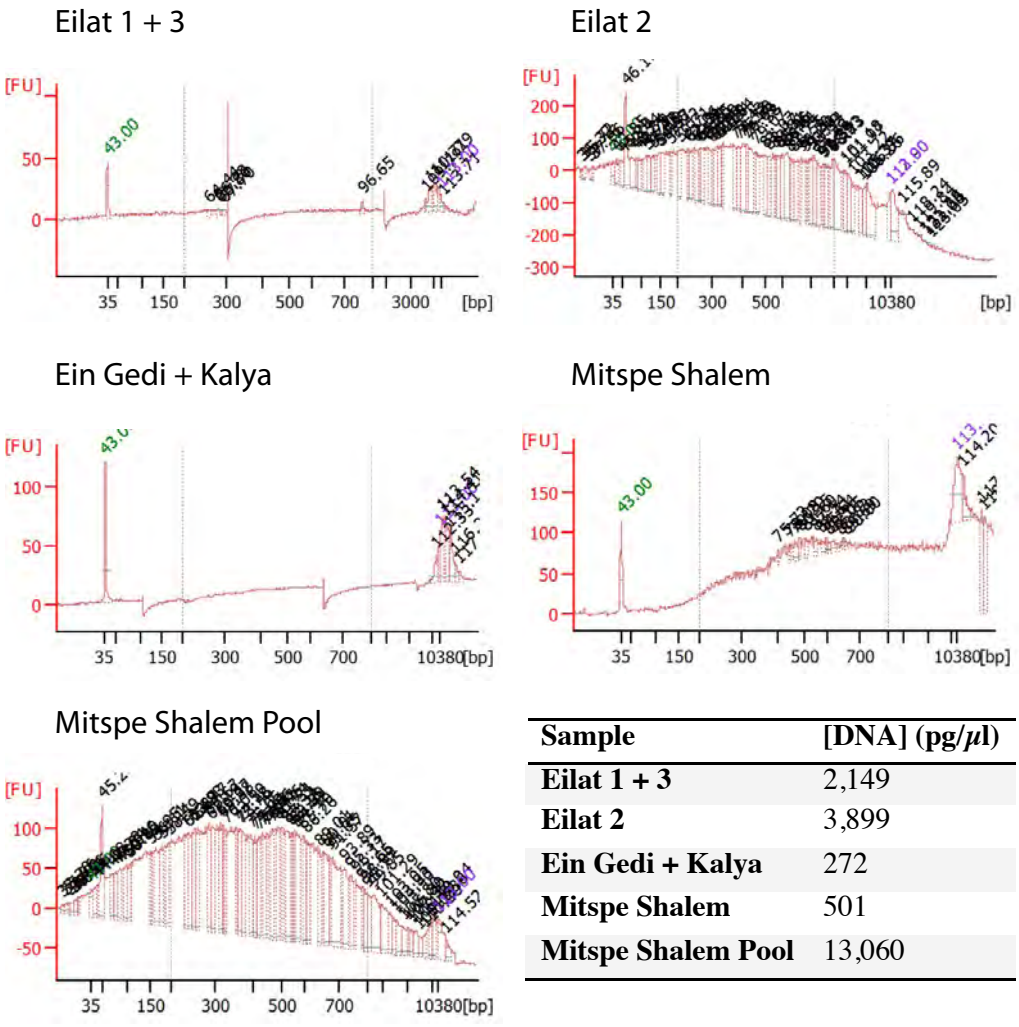
Viral enrichment was carried out on the Israeli samples to increase the concentration of virus particles and therefore genomic content for sequencing. Sea water was filtered to remove cellular matter (but to retain virus particles) and incubated with a culture of *Haloferax volcanii* to enrich virus numbers. The enriched viral fraction was spotted onto a lawn of *Haloferax volcanii*. After 48 hours no viral plaques were visible on the *Haloferax volcanii* lawn, indicating that the viruses contained in the sea water were unable to infect *Haloferax volcanii*. As enhancing viral numbers via enrichment was unsuccessful, the viral fraction was isolated through a series of filtration and precipitation techniques.

8.3.3 Analysis of DNA/RNA extracted from haloviruses

The sea water samples collected in Israel were filtered to remove cellular matter and concentrated using tangential flow filtration. Unlike normal flow filtration, in tangential flow filtration (TFF) the sea water is flowed in at a tangent to the membrane, this means that the virus particles do not build up on the membrane surface but are swept along and therefore can be collected in the

retentate (Grzenia et al 2008). The virus particles were precipitated using PEG8000 (Colombet et al 2007). DNA and RNA were isolated from haloviruses using TRIzol. The quality and amount of viral DNA extracted from these samples was measured using the Agilent High Sensitivity DNA Assay, performed by Deep Seq, University of Nottingham, Figure 8.3.

It was found that the DNA in the Eilat 2 and Mitspe Shalem Pool samples is highly degraded with 63 and 67 DNA fragment peaks between 35 bp and 10,380 bp respectively. Although the DNA in the Eilat 2 is highly degraded the DNA concentration is high, at 3,899 pg/μl because of this, this sample was taken forward for sequencing. The Mitspe Shalem Pool sample has a higher DNA concentration (13,060 pg/μl) than Eilat 2, but this sample was not taken forward for sequencing. This is because the sample was highly viscous due to PEG8000 persisting in the sample from the precipitation step. Excess PEG contamination could effect downstream enzymatic reactions during library preparation and sequencing. Unsuccessful attempts were made to remove the PEG8000 from the sample. Due to its high concentration of 2,419 pg/μl and lack of degradation the sample Eilat 1 + 3 was also taken forward for sequencing. Although the DNA from the Ein Gedi + Kalya and Mitspe Shalem samples had high integrity, the samples were viscous due to PEG8000 contamination and the DNA concentrations are low (272 pg/μl and 501 pg/μl respectively). Therefore these samples were not taken forward for sequencing.

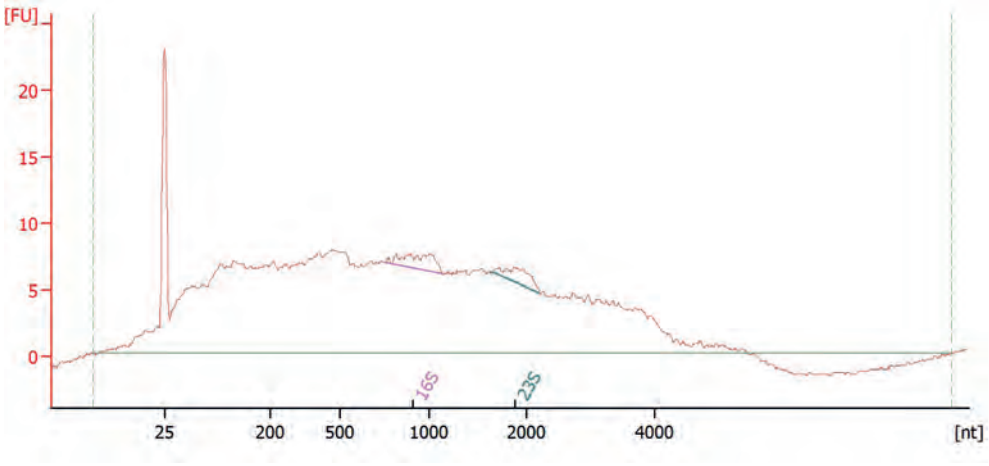


Sample	[DNA] (pg/μl)
Eilat 1 + 3	2,149
Eilat 2	3,899
Ein Gedi + Kalya	272
Mitspe Shalem	501
Mitspe Shalem Pool	13,060

Figure 8.3: **Analysis of DNA extracted from haloviruses using Agilent High Sensitivity DNA Assay.** The traces show length of DNA fragments in base pairs (bp) vs. Fluorescence Units (FU) for all DNA samples. The DNA concentration of each sample is also shown.

Due to volume sizes during extraction, RNA was only purified from Mitspe Shalem Pool and Eilat 1 + 3. The quality and amount of viral RNA extracted from these samples was measured using an Agilent RNA 6000 Pico Kit, performed by Deep Seq, University of Nottingham, Figure 8.4. Both RNA traces show a wide range of nucleotide lengths are present in the samples, indicating a degree of RNA degradation has occurred. For Mitspe Shalem Pool and Eilat 1 + 3 the RNA concentrations were 20 ng/μl and 50 ng/μl respectively. The Eilat 1 + 3 RNA sample shows a higher level of degradation and was highly viscous due to PEG8000 contamination, therefore this sample was not taken forward for sequencing.

Mitspe Shalem Pool



Eilat 1 + 3

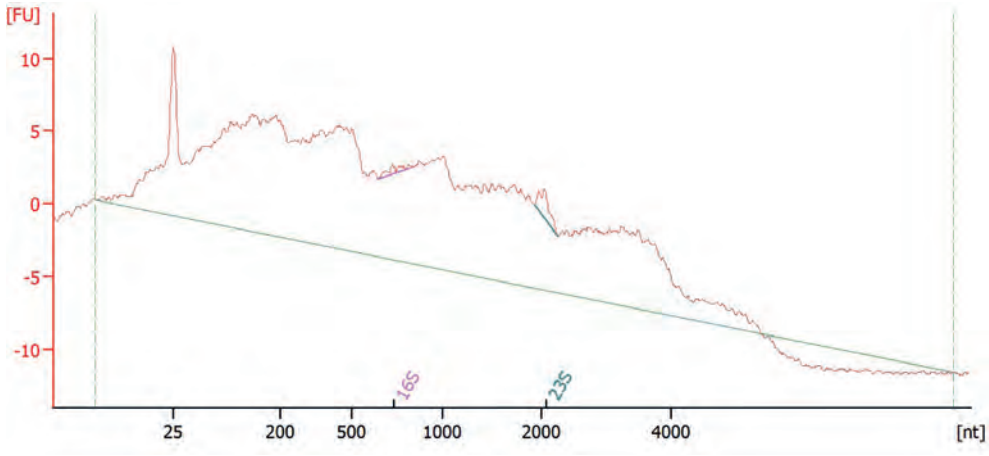


Figure 8.4: **Analysis of RNA extracted from viruses using an Agilent RNA 6000 Pico Kit.** The traces show length of RNA fragments in nucleotides (nt) vs. Fluorescence Units (FU) for all RNA samples.

The sea water samples collected at the Bras del Port salterns, Alicante, Spain were filtered to remove cellular matter and concentrated using tangential flow filtration and followed with viral precipitation using iron chloride flocculation (John et al 2011). DNA and RNA were isolated from the haloviruses using TRIzol. The viral DNA and RNA extracted was analysed using the Agilent High Sensitivity DNA Assay and an Agilent RNA 6000 Pico Kit respectively, performed by Deep Seq, University of Nottingham. The total amount of DNA and RNA in the samples was 14 ng and 4 μg respectively; both samples were highly degraded (data not shown) and so library construction for sequencing could not be carried out.

8.3.4 Sequencing and Bioinformatic Analysis

In summary, the following samples were put forward for sequencing by Illimina MiSeq (2 x 250 bp paired-end): Eilat 1 + 3 DNA, Eilat 2 DNA and Mitspe Shalem Pool RNA.

After sequencing with MiSeq, the Eilat 1 + 3 and Eilat 2 DNA reads were highly under-represented and so these samples were not taken forward for further bioinformatic analysis. The Mitspe Shalem Pool RNA sample had 93% of reads passing filter. The overlapping paired end reads were merged and error corrected using FLASH and Musket respectively, details of the reads from this sample are shown in Table 8.2.

Table 8.2: **MiSeq reads for the Mitspe Shalem Pool RNA sample.**

Sequencing ID	Number of reads	Minimum length (bp)	Maximum length (bp)	Average length (bp)
Mitspe Shalem Pool RNA	2,632,721	35	251	246

The 2,632,721 reads were assembled into contigs using the CLC Assembly Cell *de novo* assembler, assembly metrics are shown in Table 8.3.

Table 8.3: **Assembly metrics for the first contig assembly.**

Sequencing ID	Number of contigs	Average length (bp)	N50*
Mitspe Shalem Pool RNA	161,999	358.39	355

* The N50: the sum of sequences of this length or longer is at least 50% of the total length of all sequences.

The contigs were screened against 50 known haloarchaeal and halobacterial genomes, to remove any contaminating 'host' sequences. 833 contigs had total matches to known genomes and 2,793 contigs were partial matches to host genomes. The 833 total matched contigs were removed leaving 161,166 contigs remaining. The reads were then aligned to the *de novo* contigs using Bowtie 2, and assembled again using the CLC Assembly Cell *de novo* assembler; assembly metrics for the second assembly are shown in Table 8.4.

Table 8.4: **Assembly metrics for second contig assembly**

Sequencing ID	Number of contigs	Average length (bp)	N50*
Mitspe Shalem Pool RNA	143,674	365.64	365

* The N50: the sum of sequences of this length or longer is at least 50% of the total length of all sequences.

The contigs could not be assembled into long lengths; the average contig length recovered was 365.64 bp, which is not significantly greater than the average read length of 251 bp. Whole genes could not be found within the assembled contigs.

The contigs were analysed by BLASTX, a protein database searched using a translated nucleotide query. The contig sequences generated 25,324 proteins hits with E values ranging from 6.00E-176 to 1.30E-007. However, as contig length was low, only partial matches to proteins were found. Furthermore, this analysis would not be able to identify novel haloviral sequences.

To check if the sample contained any haloviral sequences at all, the contigs were aligned to 17 known haloviral genomes. Sequences were found to match to the following 6 genomes: Archaeal BJ1 virus, Halovirus HVTV-1, Halovirus HSTV-2, Halovirus HF1, Halovirus HF2 and the His1 virus.

8.4 Discussion

Isolation of haloviruses in this study was partially successful, but due to difficulties in contig assembly, novel haloviruses could not be identified. However, some contigs did show matches to published haloviral genomes, indicating that haloviruses were indeed present in the sample. The least successful part of this study was the assembly of contigs into genomes or even into lengths sizable enough to contain whole genes. After the second assembly, the average contig size was 365.64 bp, which is not significantly more than the average read length of 251 bp.

A significant reason for this outcome could be due to the low quality and quantity of DNA and RNA extracted from the viral isolate. The degradation of the viral isolate and therefore DNA and RNA could have occurred at several points through this study. Firstly, samples were collected and then shipped back to the UK for processing via surface or airmail (Israel and Alicante samples respectively). In both cases the samples would not have been in a controlled environment, and could have been subjected to a range of temperatures and possibly X-rays. The surface mail samples took several weeks to arrive, therefore allowing for increased degradation of the sample. If this were to be attempted again, processing of the samples close to the site of collection would be advantageous. Secondly, degradation could have occurred during processing of the water samples. The filtering apparatus was cumbersome and so it was not practical to process the samples next to a sterile flame or within a lamina flow hood. Efforts were made to keep the filtering apparatus clean (by flushing with NaOH) however the apparatus was intricate so it was possible that contaminants such as nucleases could have been present in the system. Loss of sample could have occurred during 0.22 μm filtration and some viruses could have been removed. However all published haloviruses have capsids smaller than 0.1 μm so this is unlikely (Dyall-Smith et al 2003). Lastly, the DNA and RNA samples were highly viscous due to PEG8000 persisting in the sample from the viral precipitation step. In high salt conditions the number of adsorbed cations on the side chains of PEG is increased which leads to higher viscosity of the solution (Brunchi & Ghimici 2013, Gonzalez-

Tello et al 1994) therefore using PEG as a precipitant in a hypersaline system is not ideal. Excess PEG contamination may have effected downstream enzymatic reactions during library preparation and sequencing. Furthermore, different sizes of DNA can be isolated by using different concentrations of PEG in solution. Higher concentrations of PEG results in the isolation of smaller sized DNA fragments (Lis & Schleif 1975). Therefore the excess PEG contamination could have increased the presence of highly degraded DNA in downstream processing as seen in the Agilent High Sensitivity DNA Assay (Figure 8.2). Due to the problems with PEG, the sea water samples from Alicante, Spain were precipitated using iron chloride (John et al 2011). 1 ml of a 10 g/litre Fe stock solution was added to each 10 litres of sea water and incubated. The Fe treated sea water was filtered through a 1.0 μm polycarbonate membrane, where the $FeCl_3$ precipitated viruses were captured on the surface of the membrane. Unfortunately, this method of viral concentration resulted in very little DNA or RNA being extracted. This indicates that the iron chloride failed to precipitate any haloviruses at all. An alternative way to isolate and concentrate the viral fraction would be via ultracentrifugation for several hours at 80,000 – 288,000 x g and purification using a sucrose or CsCl gradient (Emerson et al 2012, Garcia-Heredia et al 2012, Pietila et al 2013). However, ultracentrifugation was not available during the time of this study and furthermore would be cumbersome given the volumes involved.

Attempts to isolate haloviruses through culture-dependent methods were made, however, viral enrichment and plaque assays were unsuccessful in this study. The *Haloferax volcanii* DS70 strain was used as the 'host' in this protocol. One haloarchaeal virus (HF1) is known to infect the strain *Haloferax volcanii* WFD11 (Nuttall & Dyallsmith 1993, Tang et al 2004). Both strains are derivatives of the wild type isolate DS2. WFD11 was generated by loss of the pHV2 plasmid by ethidium bromide treatment; the large plasmid (pHV3) in this strain is unstable and so became slow growing and is prone to lysis upon transformation. Due to these difficulties, the DS70 strain (cured of pHV2 by non mutagenic methods) was generated; this strain is indistinguishable in

growth and transformation characteristics from the parent DS2 (Charlebois et al 1987, Wendoloski et al 2001). The enrichment protocol may have been more successful using different halophilic archaeal or bacterial species as the 'host', either using species that are amenable to being infected by haloviruses such as *Halorubrum* spp. and *Haloarcula* spp (Atanasova et al 2015a, Luk et al 2014) or use cellular isolates derived from the same sea water as the viral fraction. Unfortunately, no other haloarchaeal or bacterial species were available at the time and optimising growth conditions for new and unknown sea water derived species would be time consuming, during which the sea water sample would undergo further degradation. Enriching for viruses is selective, only viruses that are able to infect the 'hosts' are propagated, resulting in a loss of viral diversity. However, if only one or a few viruses are able to infect the host, the resultant concentration of those virus genomes will be high in the final sample, leading to a higher read depth during sequencing. This would result in better assembly of contigs compared to a metaviriome approach, however only one virus genome would be present.

The metaviriome sequencing approach was taken because viral enrichment was not successful. In this approach viral diversity is retained, this is an advantage as genetic information is not lost but a disadvantage as high viral diversity within a sample leads to a lower read depth and difficulties in contig assembly. These problems are compounded by the fact that only around 50 haloviral genomes are published (Luk et al 2014) and only 17 were published at the time of this study (2012). These genomes are highly diverse, meaning finding a suitable sequence to map reads against is difficult. A method to increase the read depth with metaviriome sequencing would be to construct a fosmid metagenomic library as described by Garcia-Heredia et al 2012. Fosmids are vectors based on the bacterial F-plasmid which can be used to house around 40 kb of DNA (the average size of a haloviral genome) and taken up by *E.coli*. Fosmids can be induced to have a copy number of 50 per cell immediately before DNA purification. This method woud increase the concentration of each genetic element whilst retaining viral diversity.

In this study the assembled contigs were screened against known haloarchaeal and halobacterial genomes in order to remove contaminating 'host' sequences. However, as viruses can share genetic elements with hosts and vice versa, it is likely that this screen could have removed genuine viral sequences by mistaking it for 'host'. A way to resolve this would be to check for contamination prior to sequencing by amplifying the 16S rRNA genes with universal primers for *Bacteria* (27F and 1492R) and *Archaea* (522F and 1354R) (Emerson et al 2012) to check for cellular DNA. Contamination of cellular matter could have also been checked by microscopy prior to DNA and RNA extraction.

The only sample from this study of sufficient quality to analyse bioinformatically was the Mitspe Shalem Pool RNA sample. No haloviruses containing RNA genomes have been reported to date (Pietila et al 2014), the RNA in this sample is most likely to be viral transcriptome rather than true RNA genome. The evidence for this is that matches from the sample were found to known viruses with DNA genomes. If the sample was predominantly transcriptome this could account for some of the short nucleotide lengths seen in the Agilent RNA analysis.

Another point to consider is that the Mitspe Shalem Pool sample was of the lowest salinity of all the samples at 16.44% NaCl. Although as previously mentioned that halovairuses have a wide tolerance to ranges in salinity (Kukkaro & Bamford 2009), this might not have been the most suitable sample to capture the most hyper halophilic viruses.

8.5 Future perspectives

Nanopore sequencing by Oxford Nanopore has a high error rate in base calling but improved versions of this technology are continually being released (Laver et al 2015). Therefore, Nanopore technology could be employed to sequence the metaviriome of halophilic environments in the future. This data could be mined for novel halophilic DNA processing enzymes to fulfil the objective of

this study or alternatively to advance the field of haloviruses in general. At the time of this study in 2012, Nanopore sequencing by Oxford Nanopore was not publically available.

Nanopore sequencing could circumvent some of the genome assembly problems encountered in this study. Firstly, the current MinION flow cell contains 512 channels; each channel is connected to 4 wells which may each contain a nanopore. Each channel provides data from one of the four wells at a time, allowing up to 512 independent DNA molecules to be sequenced simultaneously (Bayley 2015, Ip et al 2015). The data from each pore is also captured separately from that of data from other pores. Therefore, for samples containing mixed DNA, for example a metaviriome, each genome will be sequenced individually. This avoids the bioinformatic issue of separating different genome sequences from each other downstream of sequencing, a task that would have to be performed with conventional sequencing methods. Secondly, unlike conventional sequencing methods, nucleic acids do not have to be fragmented to a few hundred bp during library preparation (Bradley et al 2015). Nanopore sequencing can sequence an intact strand of DNA that can be up to hundreds of kb in length. The data output is a continuous sequence the length of the DNA provided, therefore the task of assembling short reads is avoided, which is a significant problem in metaviriome sequencing. Thirdly, as genomes are sequenced whole and not in short fragments, it would be easier to identify and discard contaminating cellular sequence data from a metaviriome sample than with conventional sequencing methods. Fourthly, during library preparation a hairpin adaptor is ligated to the dsDNA about to be sequenced, during sequencing the leading template strand passes through the nanopore followed by the hairpin adapter and then the complement strand. A consensus sequence from both strands of DNA can be made therefore improving the accuracy and read depth of sequence produced (Goodwin et al 2015). This proof reading technique is advantageous as metaviriomes are diverse and concentrations of individual virus genomes could be low. Lastly, as contig assembly is not required, this method of sequencing is less reliant on having a

good reference genome to map data against, which for haloviruses is advantageous as reference genomes are not available.

Since the time of this study in 2012, 33 more haloviral genomes have been published (Luk et al 2014). If novel DNA processing enzymes are still required for the advancement of nanopore technology, these genomes could be mined for novel helicase and polymerase sequences.

8.6 Conclusion

The attempt to identify novel halophilic DNA processing enzymes through the sequencing of the metaviriome of haloviral communities was unsuccessful. This was in part due to the poor quality of DNA and RNA provided for sequencing and the numerous difficulties in assembling reads from metaviriomes using conventional sequencing technologies such as Illumina MiSeq.

Chapter 9: Conclusion and Future Perspectives

Hel308 is a 3′ to 5′ RecQ family DNA helicase that is conserved in metazoans and archaea but is absent from bacteria and fungi. Hel308 family helicases are implicated in DNA repair, homologous recombination and genome stability, but the exact cellular role of Hel308 is largely unknown. Strains deleted for *hel308* are sensitive to DNA interstrand crosslinks, which are potent blocks to DNA replication. In humans and archaea, Hel308 localises at damaged DNA replication forks. In this study, *Haloferax volcanii* was used as a model archaeon to study the cellular role of Hel308.

Hel308 is an Anti-recombinase

Previous studies have shown that Hel308 homologues can disassemble recombinase filaments from ssDNA and D-loop structures, suggesting that Hel308 acts as an anti-recombinase. This study provides further evidence to support this proposal and provides insight into the mode of action of Hel308 as an archaeal anti-recombinase.

The recombination mediator RadB assists the polymerisation of RadA onto ssDNA during homologous recombination (HR), and deletion of RadB results in slow growth and a marked reduction in HR to ~ 5% that of a wild type strain (Haldenby 2007). Deletion of *hel308* in combination with *radB* resulted in a reduction in the severity of growth, DNA damage and recombination phenotypes, compared to a *radB* strain. This finding suggests that Hel308 antagonises RadB by acting as an anti-recombinase; it also suggests that RadB could have a regulatory role during homologous recombination (in addition to its role as a mediator to assist RadA polymerisation), namely to promote the repair of DNA damage by the HR pathway.

In a preliminary assay, Hel308 potentially interacted with RadA *in vivo,* reinforcing the suggestion that Hel308 acts as an anti-recombinase. An interaction between Hel308 and RadB was not seen, most likely due to low

levels of RadB in the cell compared to RadA. It is of interest to confirm this interaction, in order to understand the role of Hel308 fully.

Point mutations in domains 1, 2, 3 and 5 of Hel308 perturbed the frequency of homologous recombination in the cell, confirming that Hel308 is critical in the regulation of recombination. Several point mutations (F316A, E330G, D420A, E422G and R743A) had little effect on growth or survival of cells after treatment with DNA damaging agents, however some mutants such as F316A and R743A resulted in profound changes in recombination frequency. These observations suggest that the helicase unwinding activity of Hel308 is separate to its function in the regulation of recombination, and that correct regulation of recombination relies heavily on the correct structural conformation of Hel308. These suggestions pose interesting questions about how these two roles are managed by Hel308 and why structure in particular is important to the its regulatory role in HR.

The point mutants F316A and R743A resulted in around a 100-fold and 60-fold increase in recombination frequency, respectively, and generated only non-crossover (NCO) products. The synthesis dependent strand annealing (SDSA) pathway of homologous recombination is capable of generating only NCO products (San Filippo et al 2008). Therefore, Hel308 may regulate not only the level of homologous recombination in the cell but also influence the pathway choice by which recombination intermediates are resolved. It would be of interest to determine how Hel308 is involved in this pathway choice.

Interactions of Hel308

Could Hel308 interact with RecJ4?

Preliminary assays indicated that Hel308 may interact with RecJ4 *in vivo*. RecJ is a 5'-3' exonuclease that in concert with the RecQ helicase generates ssDNA intermediates during mismatch repair, nucleotide excision repair (NER) and homologous recombination in bacteria (Morimatsu & Kowalczykowski 2014). If this interaction is confirmed to be true this could suggest that Hel308 (a

RecQ family helicase) may act in a range of DNA repair pathways in archaea, alongside RecJ.

More interestingly, RecJ has been proposed to be the archaeal counterpart to the eukaryotic CDC45 protein in the DNA replication helicase CDC45-MCM-GINS (CMG) complex (Makarova et al 2012). In *H. volcanii*, RecJ4 was seen to co-purify with the replication initiation protein Orc1 (Darya Ausiannikava, University of Nottingham, personal communication). If RecJ4 is a part of the archaeal CMG complex and if its interaction with Hel308 is verified then Hel308 may play a role in the restart of stalled DNA replication forks.

Could Hel308 be Involved in Transcription Coupled Repair?

Preliminary assays indicated that Hel308 may interact with the RNA polymerase subunit RpoA1 and the NER protein UvrA *in vivo*. If this interaction is found to be true this suggests a role for Hel308 in transcription-coupled repair (TCR). During TCR, RNA polymerase stalls at a DNA lesion, which is then repaired via NER involving the UvrABC complex. In *H. volcanii*, TCR is not dependent on UvrA (Stantial et al 2016), which raises the possibility that Hel308 could be acting as the coupling factor in TCR. Confirmation of this possible interaction and deletion of *hel308* in combination with TCR (and NER) proteins could give insight into role of Hel308 in this DNA repair pathway.

Hel308b is a Regulator of Recombination?

In this study, it was found that *H. volcanii* and four closely-related haloarchaeal species contain a second Hel308 helicase named Hel308b. Hel308b likely arose from an ancient gene duplication event, and notably Hel308b lacks the 'auto-inhibitory' domain 5 found in the canonical Hel308 helicases. Deletion of *hel308b* results in a recombination deficiency that is more severe than that seen in *Δhel308* mutants, suggesting that Hel308b is involved in the regulation of homologous recombination. However, unlike Hel308, Hel308b does not appear to be involved in ICL repair.

Many questions remain about the role of Hel308b in *H. volcanii*. Why has a second Hel308 helicase been retained? Does Hel308b have a distinct role in the cell and is this role different or redundant to Hel308? Does the lack of domain 5 confer any functional advantage to Hel308b? Protein:protein pull-down assays of Hel308b could highlight interaction partners of Hel308b and provide further insights into its role.

Summary

This study has highlighted that Hel308 is a multifaceted helicase that could play roles in different overlapping DNA repair pathways in *H. volcanii*. Hel308 has shown to be a regulator of homologous recombination and is implicated in interstrand crosslink repair. The information presented in this study provides an insight into the understanding of DNA repair and genome stability in archaea.

References

Aalto AP, Bitto D, Ravantti JJ, Bamford DH, Huiskonen JT, Oksanen HM. 2012. Snapshot of virus evolution in hypersaline environments from the characterization of a membrane-containing Salisaeta icosahedral phage 1. *Proc Natl Acad Sci U S A* 109: 7079-84

Adams MD, McVey M, Sekelsky JJ. 2003. Drosophila BLM in double-strand break repair by synthesis-dependent strand annealing. *Science* 299: 265-7

Adelman CA, Lolo RL, Birkbak NJ, Murina O, Matsuzaki K, et al. 2013. HELQ promotes RAD51 paralogue-dependent repair to avert germ cell loss and tumorigenesis. *Nature* 502: 381-4

Albers SV, Forterre P, Prangishvili D, Schleper C. 2013. The legacy of Carl Woese and Wolfram Zillig: from phylogeny to landmark discoveries. *Nat Rev Microbiol* 11: 713-9

Albers SV, Meyer BH. 2011. The archaeal cell envelope. *Nat Rev Microbiol* 9: 414-26

Allers T. 2010. Overexpression and purification of halophilic proteins in Haloferax volcanii. *Bioeng Bugs* 1: 288-90

Allers T, Barak S, Liddell S, Wardell K, Mevarech M. 2010. Improved strains and plasmid vectors for conditional overexpression of His-tagged proteins in Haloferax volcanii. *Appl Environ Microbiol* 76: 1759-69

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

Allers T, Ngo HP, Mevarech M, Lloyd RG. 2004. Development of additional selectable markers for the halophilic archaeon Haloferax volcanii based on the leuB and trpA genes. *Appl Environ Microbiol* 70: 943-53

Altman-Price N, Mevarech M. 2009. Genetic evidence for the importance of protein acetylation and protein deacetylation in the halophilic archaeon Haloferax volcanii. *J Bacteriol* 191: 1610-7

Ammar R, Torti D, Tsui K, Gebbia M, Durbic T, et al. 2012. Chromatin is an ancient innovation conserved between Archaea and Eukarya. *Elife* 1: e00078

Aono R, Sato T, Yano A, Yoshida S, Nishitani Y, et al. 2012. Enzymatic characterization of AMP phosphorylase and ribose-1,5-bisphosphate isomerase functioning in an archaeal AMP metabolic pathway. *J Bacteriol* 194: 6847-55

Aravind L, Koonin EV. 2001. Prokaryotic homologs of the eukaryotic DNA-end-binding protein Ku, novel domains in the Ku protein and prediction of a prokaryotic double-strand break repair system. *Genome Res* 11: 1365-74

Arezi B, Kuchta RD. 2000. Eukaryotic DNA primase. *Trends Biochem Sci* 25: 572-6

Arias-Palomo E, O'Shea VL, Hood IV, Berger JM. 2013. The bacterial DnaC helicase loader is a DnaB ring breaker. *Cell* 153: 438-48

Atanasova NS, Demina TA, Buivydas A, Bamford DH, Oksanen HM. 2015. Archaeal viruses multiply: temporal screening in a solar saltern. *Viruses* 7: 1902-26

Atanasova NS, Oksanen HM, Bamford DH. 2015. Haloviruses of archaea, bacteria, and eukaryotes. *Curr Opin Microbiol* 25: 40-8

Atkinson J, McGlynn P. 2009. Replication fork reversal and the maintenance of genome stability. *Nucleic Acids Res* 37: 3475-92

Auchtung TA, Takacs-Vesbach CD, Cavanaugh CM. 2006. 16S rRNA phylogenetic

investigation of the candidate division "Korarchaeota". *Appl Environ Microbiol* 72: 5077-82
Azeroglu B, Mawer JS, Cockram CA, White MA, Hasan AM, et al. 2016. RecG Directs DNA Synthesis during Double-Strand Break Repair. *PLoS Genet* 12: e1005799
Bab-Dinitz E, Shmuely H, Maupin-Furlow J, Eichler J, Shaanan B. 2006. Haloferax volcanii PitA: an example of functional interaction between the Pfam chlorite dismutase and antibiotic biosynthesis monooxygenase families? *Bioinformatics* 22: 671-5
Babron MC, Kazma R, Gaborieau V, McKay J, Brennan P, et al. 2014. Genetic variants in DNA repair pathways and risk of upper aerodigestive tract cancers: combined analysis of data from two genome-wide association studies in European populations. *Carcinogenesis* 35: 1523-7
Bachrati CZ, Hickson ID. 2009. Dissolution of double Holliday junctions by the concerted action of BLM and topoisomerase IIIalpha. *Methods Mol Biol* 582: 91-102
Bansal T, Englert D, Lee J, Hegde M, Wood TK, Jayaraman A. 2007. Differential effects of epinephrine, norepinephrine, and indole on Escherichia coli O157:H7 chemotaxis, colonization, and gene expression. *Infect Immun* 75: 4597-607
Barry ER, Bell SD. 2006. DNA replication in the archaea. *Microbiol Mol Biol Rev* 70: 876-87
Bayley H. 2006. Sequencing single molecules of DNA. *Curr Opin Chem Biol* 10: 628-37
Bayley H. 2015. Nanopore sequencing: from imagination to reality. *Clin Chem* 61: 25-31
Bell SD. 2011. DNA replication: archaeal oriGINS. *BMC Biol* 9: 36
Bell SD, Jackson SP. 2001. Mechanism and regulation of transcription in archaea. *Curr Opin Microbiol* 4: 208-13
Bell SP, Dutta A. 2002. DNA replication in eukaryotic cells. *Annu Rev Biochem* 71: 333-74
Benner S, Chen RJ, Wilson NA, Abu-Shumays R, Hurt N, et al. 2007. Sequence-specific detection of individual DNA polymerase complexes in real time using a nanopore. *Nat Nanotechnol* 2: 718-24
Benson FE, Baumann P, West SC. 1998. Synergistic actions of Rad51 and Rad52 in recombination and DNA repair. *Nature* 391: 401-4
Bernander R, Lundgren M, Ettema TJ. 2010. Comparative and functional analysis of the archaeal cell cycle. *Cell Cycle* 9: 794-806
Betts M, Russell R. 2003. *Amino Acid Properties and Consequences of Substitutions*. John Wiley & Sons, Ltd.
Bitan-Banin G, Ortenberg R, Mevarech M. 2003. Development of a gene knockout system for the halophilic archaeon Haloferax volcanii by use of the pyrE gene. *J Bacteriol* 185: 772-8
Bizard AH, Hickson ID. 2014. The dissolution of double Holliday junctions. *Cold Spring Harb Perspect Biol* 6: a016477
Bjornson KP, Wong I, Lohman TM. 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: 411-22
Blastyak A, Pinter L, Unk I, Prakash L, Prakash S, Haracska L. 2007. Yeast Rad5 protein required for postreplication repair has a DNA helicase activity specific for

replication fork regression. *Mol Cell* 28: 167-75
Bocquet N, Bizard AH, Abdulrahman W, Larsen NB, Faty M, et al. 2014. Structural and mechanistic insight into Holliday-junction dissolution by topoisomerase IIIalpha and RMI1. *Nat Struct Mol Biol* 21: 261-8
Bolt EL, Lloyd RG, Sharples GJ. 2001. Genetic analysis of an archaeal Holliday junction resolvase in Escherichia coli. *J Mol Biol* 310: 577-89
Bowater R, Doherty AJ. 2006. Making ends meet: repairing breaks in bacterial DNA by non-homologous end-joining. *PLoS Genet* 2: e8
Boyd JB, Golino MD, Shaw KE, Osgood CJ, Green MM. 1981. Third-chromosome mutagen-sensitive mutants of Drosophila melanogaster. *Genetics* 97: 607-23
Boyd JB, Sakaguchi K, Harris PV. 1990. Mus308 Mutants of Drosophila Exhibit Hypersensitivity to DNA Cross-Linking Agents and Are Defective in a Deoxyribonuclease. *Genetics* 125: 813-19
Bradley P, Gordon NC, Walker TM, Dunn L, Heys S, et al. 2015. Rapid antibiotic-resistance predictions from genome sequence data for Staphylococcus aureus and Mycobacterium tuberculosis. *Nat Commun* 6: 10063
Braha O, Walker B, Cheley S, Kasianowicz JJ, Song LZ, et al. 1997. Designed protein pores as components for biosensors. *Chem Biol* 4: 497-505
Brendel J, Stoll B, Lange SJ, Sharma K, Lenz C, et al. 2014. A complex of Cas proteins 5, 6, and 7 is required for the biogenesis and stability of clustered regularly interspaced short palindromic repeats (crispr)-derived rnas (crrnas) in Haloferax volcanii. *J Biol Chem* 289: 7164-77
Breuert S, Allers T, Spohn G, Soppa J. 2006. Regulated polyploidy in halophilic archaea. *PloS one* 1: e92
Brochier-Armanet C, Boussau B, Gribaldo S, Forterre P. 2008. Mesophilic Crenarchaeota: proposal for a third archaeal phylum, the Thaumarchaeota. *Nat Rev Microbiol* 6: 245-52
Brochier-Armanet C, Forterre P, Gribaldo S. 2011. Phylogeny and evolution of the Archaea: one hundred genomes later. *Curr Opin Microbiol* 14: 274-81
Brosh RM, Jr. 2013. DNA helicases involved in DNA repair and their roles in cancer. *Nat Rev Cancer* 13: 542-58
Brosius J. 2003. The contribution of RNAs and retroposition to evolutionary novelties. *Genetica* 118: 99-116
Brunchi C-E, Ghimici L. 2013. PEG in aqueous salt solutions. Viscosity and separation ability in a TiO2 suspension. *Academia Romana* 58: 183-88
Bugreev DV, Yu X, Egelman EH, Mazin AV. 2007. Novel pro- and anti-recombination activities of the Bloom's syndrome helicase. *Genes Dev* 21: 3085-94
Bult CJ, White O, Olsen GJ, Zhou L, Fleischmann RD, et al. 1996. Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii. *Science* 273: 1058-73
Buttner K, Nehring S, Hopfner KP. 2007. Structural basis for DNA duplex separation by a superfamily-2 helicase. *Nat Struct Mol Biol* 14: 647-52
Cadet J, Sage E, Douki T. 2005. Ultraviolet radiation-mediated damage to cellular DNA. *Mutat Res* 571: 3-17
Cam EL, Culard F, Larquet E, Delain E, Cognet JA. 1999. DNA bending induced by the archaebacterial histone-like protein MC1. *J Mol Biol* 285: 1011-21

Cann IK, Ishino S, Hayashi I, Komori K, Toh H, et al. 1999. Functional interactions of a homolog of proliferating cell nuclear antigen with DNA polymerases in Archaea. *J Bacteriol* 181: 6591-9

Cann IK, Komori K, Toh H, Kanai S, Ishino Y. 1998. A heterodimeric DNA polymerase: evidence that members of Euryarchaeota possess a distinct DNA polymerase. *Proc Natl Acad Sci U S A* 95: 14250-5

Caruthers JM, McKay DB. 2002. Helicase structure and mechanism. *Curr Opin Struc Biol* 12: 123-33

Ceccaldi R, Liu JC, Amunugama R, Hajdu I, Primack B, et al. 2015. Homologous-recombination-deficient tumours are dependent on Poltheta-mediated repair. *Nature* 518: 258-62

Chaban B, Ng SY, Jarrell KF. 2006. Archaeal habitats--from the extreme to the ordinary. *Can J Microbiol* 52: 73-116

Chan SH, Yu AM, McVey M. 2010. Dual roles for DNA polymerase theta in alternative end-joining repair of double-strand breaks in Drosophila. *PLoS Genet* 6: e1001005

Charlebois RL, Lam WL, Cline SW, Doolittle WF. 1987. Characterization of pHV2 from Halobacterium volcanii and its use in demonstrating transformation of an archaebacterium. *Proc Natl Acad Sci U S A* 84: 8530-4

Charlebois RL, Schalkwyk LC, Hofman JD, Doolittle WF. 1991. Detailed physical map and set of overlapping clones covering the genome of the archaebacterium Haloferax volcanii DS2. *J Mol Biol* 222: 509-24

Chen CC, Walia R, Mukherjee KJ, Mahalik S, Summers DK. 2015. Indole generates quiescent and metabolically active Escherichia coli cultures. *Biotechnol J* 10: 636-46

Chen Z, Yang H, Pavletich NP. 2008. Mechanism of homologous recombination from the RecA-ssDNA/dsDNA structures. *Nature* 453: 489-4

Chimerel C, Field CM, Pinero-Fernandez S, Keyser UF, Summers DK. 2012. Indole prevents Escherichia coli cell division by modulating membrane potential. *Biochim Biophys Acta* 1818: 1590-4

Choi KH. 2012. Viral polymerases. *Adv Exp Med Biol* 726: 267-304

Chow KH, Courcelle J. 2007. RecBCD and RecJ/RecQ initiate DNA degradation on distinct substrates in UV-irradiated Escherichia coli. *Radiat Res* 168: 499-506

Christian JHB, Waltho JA. 1962. Solute Concentrations within Cells of Halophilic and Non-Halophilic Bacteria. *Biochim Biophys Acta* 65: 506-&

Ciccarelli FD, Doerks T, von Mering C, Creevey CJ, Snel B, Bork P. 2006. Toward automatic reconstruction of a highly resolved tree of life. *Science* 311: 1283-7

Cline SW, Schalkwyk LC, Doolittle WF. 1989. Transformation of the archaebacterium Halobacterium volcanii with genomic DNA. *J Bacteriol* 171: 4987-91

Coin F, Bergmann E, Tremeau-Bravard A, Egly JM. 1999. Mutations in XPB and XPD helicases found in xeroderma pigmentosum patients impair the transcription function of TFIIH. *EMBO J* 18: 1357-66

Collin F, Karkare S, Maxwell A. 2011. Exploiting bacterial DNA gyrase as a drug target: current state and perspectives. *Appl Microbiol Biotechnol* 92: 479-97

Colombet J, Robin A, Lavie L, Bettarel Y, Cauchie HM, Sime-Ngando T. 2007. Virioplankton 'pegylation': use of PEG (polyethylene glycol) to concentrate and purify viruses in pelagic ecosystems. *J Microbiol Methods* 71: 212-9

Connaris H, Chaudhuri JB, Danson MJ, Hough DW. 1999. Expression, reactivation, and purification of enzymes from Haloferax volcanii in Escherichia coli. *Biotechnol Bioeng* 64: 38-45

Constantinesco F, Forterre P, Elie C. 2002. NurA, a novel 5'-3' nuclease gene linked to rad50 and mre11 homologs of thermophilic Archaea. *EMBO Rep* 3: 537-42

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

Constantinou A, Tarsounas M, Karow JK, Brosh RM, Bohr VA, et al. 2000. Werner's syndrome protein (WRN) migrates Holliday junctions and co-localizes with RPA upon replication arrest. *EMBO Rep* 1: 80-4

Corbett KD, Shultzaberger RK, Berger JM. 2004. The C-terminal domain of DNA gyrase A adopts a DNA-bending beta-pinwheel fold. *Proc Natl Acad Sci U S A* 101: 7293-8

Cordin O, Tanner NK, Doere M, Linder P, Banroques J. 2004. The newly discovered Q motif of DEAD-box RNA helicases regulates RNA-binding and helicase activity. *Embo Journal* 23: 2478-87

Costa RM, Chigancas V, Galhardo Rda S, Carvalho H, Menck CF. 2003. The eukaryotic nucleotide excision repair pathway. *Biochimie* 85: 1083-99

Cox CJ, Foster PG, Hirt RP, Harris SR, Embley TM. 2008. The archaebacterial origin of eukaryotes. *Proc Natl Acad Sci U S A* 105: 20356-61

Craig JM, Laszlo AH, Derrington IM, Ross BC, Brinkerhoff H, et al. 2015. Direct Detection of Unnatural DNA Nucleotides dNaM and d5SICS using the MspA Nanopore. *PLoS One* 10: e0143253

Crameri A, Whitehorn EA, Tate E, Stemmer WP. 1996. Improved green fluorescent protein by molecular evolution using DNA shuffling. *Nat Biotechnol* 14: 315-9

Crowley DJ, Boubriak I, Berquist BR, Clark M, Richard E, et al. 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, White MF. 2005. DNA damage detection by an archaeal single-stranded DNA-binding protein. *J Mol Biol* 353: 507-16

D'Amours D, Jackson SP. 2002. The Mre11 complex: at the crossroads of dna repair and checkpoint signalling. *Nat Rev Mol Cell Biol* 3: 317-27

DasSarma P, Coker J, Huse V, DasSarma S. 2010. Halophiles, Industrial Applications. In *Encyclopedia of Industrial Biotechnology: Bioprocess, Bioseparation, and Cell Technology*, ed. MC Flickinger, pp. 1–43: John Wiley & Sons, Inc.

Davis AJ, Chen DJ. 2013. DNA double strand break repair via non-homologous end-joining. *Transl Cancer Res* 2: 130-43

de Boer J, Hoeijmakers JH. 2000. Nucleotide excision repair and human syndromes. *Carcinogenesis* 21: 453-60

de Boer P, Crossley R, Rothfield L. 1992. The essential bacterial cell-division protein FtsZ is a GTPase. *Nature* 359: 254-6

de Jager M, van Noort J, van Gent DC, Dekker C, Kanaar R, Wyman C. 2001. Human Rad50/Mre11 is a flexible complex that can tether DNA ends. *Mol Cell* 8: 1129-35

de Laat WL, Jaspers NG, Hoeijmakers JH. 1999. Molecular mechanism of nucleotide excision repair. *Genes Dev* 13: 768-85

Deaconescu AM, Chambers AL, Smith AJ, Nickels BE, Hochschild A, et al. 2006. Structural basis for bacterial transcription-coupled DNA repair. *Cell* 124: 507-20
Deamer D. 2010. Nanopore analysis of nucleic acids bound to exonucleases and polymerases. *Annu Rev Biophys* 39: 79-90
Della M, Palmbos PL, Tseng HM, Tonkin LM, Daley JM, et al. 2004. Mycobacterial Ku and ligase proteins constitute a two-component NHEJ repair machine. *Science* 306: 683-5
Delmas S, Duggin IG, Allers T. 2013. DNA damage induces nucleoid compaction via the Mre11-Rad50 complex in the archaeon Haloferax volcanii. *Mol Microbiol* 87: 168-79
Delmas S, Shunburne L, Ngo HP, Allers T. 2009. Mre11-Rad50 promotes rapid repair of DNA damage in the polyploid archaeon Haloferax volcanii by restraining homologous recombination. *PLoS Genet* 5: e1000552
DeLong EF, Pace NR. 2001. Environmental diversity of bacteria and archaea. *Syst Biol* 50: 470-8
Deng X, Habel JE, Kabaleeswaran V, Snell EH, Wold MS, Borgstahl GE. 2007. Structure of the full-length human RPA14/32 complex gives insights into the mechanism of DNA binding and complex formation. *J Mol Biol* 374: 865-76
Derrington IM, Butler TZ, Collins MD, Manrao E, Pavlenok M, et al. 2010. Nanopore DNA sequencing with MspA. *Proc Natl Acad Sci U S A* 107: 16060-5
Di Giulio M. 2007. The tree of life might be rooted in the branch leading to Nanoarchaeota. *Gene* 401: 108-13
Dixon DA, Kowalczykowski SC. 1993. The recombination hotspot chi is a regulatory sequence that acts by attenuating the nuclease activity of the E. coli RecBCD enzyme. *Cell* 73: 87-96
Dogliotti E, Fortini P, Pascucci B, Parlanti E. 2001. The mechanism of switching among multiple BER pathways. *Prog Nucleic Acid Res Mol Biol* 68: 3-27
Doherty AJ, Jackson SP, Weller GR. 2001. Identification of bacterial homologues of the Ku DNA repair proteins. *FEBS Lett* 500: 186-8
Duggin IG, Wake RG, Bell SD, Hill TM. 2008. The replication fork trap and termination of chromosome replication. *Mol Microbiol* 70: 1323-33
Dyall-Smith M, Tang SL, Bath C. 2003. Haloarchaeal viruses: how diverse are they? *Res Microbiol* 154: 309-13
Eisen JA, Hanawalt PC. 1999. A phylogenomic study of DNA repair genes, proteins, and processes. *Mutat Res* 435: 171-213
Elder JF, Jr., Turner BJ. 1995. Concerted evolution of repetitive DNA sequences in eukaryotes. *Q Rev Biol* 70: 297-320
Elkins JG, Podar M, Graham DE, Makarova KS, Wolf Y, et al. 2008. A korarchaeal genome reveals insights into the evolution of the Archaea. *Proc Natl Acad Sci U S A* 105: 8102-7
Eme L, Doolittle WF. 2015. Archaea. *Curr Biol* 25: R851-5
Emerson JB, Thomas BC, Andrade K, Allen EE, Heidelberg KB, Banfield JF. 2012. Dynamic viral populations in hypersaline systems as revealed by metagenomic assembly. *Appl Environ Microbiol* 78: 6309-20
Enemark EJ, Joshua-Tor L. 2008. On helicases and other motor proteins. *Curr Opin Struct Biol* 18: 243-57

Erzberger JP, Mott ML, Berger JM. 2006. Structural basis for ATP-dependent DnaA assembly and replication-origin remodeling. *Nat Struct Mol Biol* 13: 676-83

Evans MD, Dizdaroglu M, Cooke MS. 2004. Oxidative DNA damage and disease: induction, repair and significance. *Mutat Res* 567: 1-61

Eydmann T, Sommariva E, Inagawa T, Mian S, Klar AJ, Dalgaard JZ. 2008. Rtf1-mediated eukaryotic site-specific replication termination. *Genetics* 180: 27-39

Falk-Krzesinski HJ, Wolfe AJ. 1998. Genetic analysis of the nuo locus, which encodes the proton-translocating NADH dehydrogenase in Escherichia coli. *J Bacteriol* 180: 1174-84

Fang L, Davey MJ, O'Donnell M. 1999. Replisome assembly at oriC, the replication origin of E. coli, reveals an explanation for initiation sites outside an origin. *Mol Cell* 4: 541-53

Fasching CL, Cejka P, Kowalczykowski SC, Heyer WD. 2015. Top3-Rmi1 dissolve Rad51-mediated D loops by a topoisomerase-based mechanism. *Mol Cell* 57: 595-606

Fawaz MV, Topper ME, Firestine SM. 2011. The ATP-grasp enzymes. *Bioorg Chem* 39: 185-91

Feng J, Chen Y, Pu J, Yang X, Zhang C, et al. 2011. An improved malachite green assay of phosphate: mechanism and application. *Anal Biochem* 409: 144-9

Foster PG, Cox CJ, Embley TM. 2009. The primary divisions of life: a phylogenomic approach employing composition-heterogeneous methods. *Philos Trans R Soc Lond B Biol Sci* 364: 2197-207

Fotheringham IG, Dacey SA, Taylor PP, Smith TJ, Hunter MG, et al. 1986. The cloning and sequence analysis of the aspC and tyrB genes from Escherichia coli K12. Comparison of the primary structures of the aspartate aminotransferase and aromatic aminotransferase of E. coli with those of the pig aspartate aminotransferase isoenzymes. *Biochem J* 234: 593-604

Frick DN, Richardson CC. 2001. DNA primases. *Annu Rev Biochem* 70: 39-80

Frols S, White MF, Schleper C. 2009. Reactions to UV damage in the model archaeon Sulfolobus solfataricus. *Biochem Soc Trans* 37: 36-41

Fujii R, Yoshida H, Fukusumi S, Habata Y, Hosoya M, et al. 2002. Identification of a neuropeptide modified with bromine as an endogenous ligand for GPR7. *The Journal of biological chemistry* 277: 34010-6

Fujikane R, Komori K, Shinagawa H, Ishino Y. 2005. Identification of a novel helicase activity unwinding branched DNAs from the hyperthermophilic archaeon, Pyrococcus furiosus. *J Biol Chem* 280: 12351-8

Fujikane R, Shinagawa H, 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

Fukui K. 2010. DNA mismatch repair in eukaryotes and bacteria. *J Nucleic Acids* 2010

Fuss JO, Tainer JA. 2011. XPB and XPD helicases in TFIIH orchestrate DNA duplex opening and damage verification to coordinate repair with transcription and cell cycle via CAK kinase. *DNA Repair (Amst)* 10: 697-713

Gabaldon T, Huynen MA. 2004. Prediction of protein function and pathways in the genome era. *Cell Mol Life Sci* 61: 930-44

Gabbai CB, Marians KJ. 2010. Recruitment to stalled replication forks of the PriA DNA

helicase and replisome-loading activities is essential for survival. *DNA Repair (Amst)* 9: 202-9
Garcia-Heredia I, Martin-Cuadrado AB, Mojica FJ, Santos F, Mira A, et al. 2012. Reconstructing viral genomes from the environment using fosmid clones: the case of haloviruses. *PLoS One* 7: e33802
Gari K, Decaillet C, Stasiak AZ, Stasiak A, Constantinou A. 2008. The Fanconi anemia protein FANCM can promote branch migration of Holliday junctions and replication forks. *Mol Cell* 29: 141-8
Gasteiger E, Gattiker A, Hoogland C, Ivanyi I, Appel RD, Bairoch A. 2003. ExPASy: The proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res* 31: 3784-8
Ge XQ, Jackson DA, Blow JJ. 2007. Dormant origins licensed by excess Mcm2-7 are required for human cells to survive replicative stress. *Genes Dev* 21: 3331-41
Gelfand DH, Steinberg RA. 1977. Escherichia coli mutants deficient in the aspartate and aromatic amino acid aminotransferases. *J Bacteriol* 130: 429-40
Genschel J, Modrich P. 2009. Functions of MutLalpha, replication protein A (RPA), and HMGB1 in 5'-directed mismatch repair. *J Biol Chem* 284: 21536-44
Gerhardt S, Schott AK, Kairies N, Cushman M, Illarionov B, et al. 2002. Studies on the reaction mechanism of riboflavin synthase: X-ray crystal structure of a complex with 6-carboxyethyl-7-oxo-8-ribityllumazine. *Structure* 10: 1371-81
German J. 1993. Bloom syndrome: a mendelian prototype of somatic mutational disease. *Medicine (Baltimore)* 72: 393-406
Gietl A, Holzmeister P, Blombach F, Schulz S, von Voithenberg LV, et al. 2014. Eukaryotic and archaeal TBP and TFB/TF(II)B follow different promoter DNA bending pathways. *Nucleic Acids Res* 42: 6219-31
Giroux X, MacNeill SA. 2015. Inhibiting NAD+-dependent DNA ligase activity with 2-(cyclopentyloxy)-5'-deoxyadenosine (CPOdA) offers a new tool for DNA replication and repair studies in the model archaeon Haloferax volcanii. *FEMS Microbiol Lett* 362
Gonzalez-Tello P, Camacho F, Blazquez G. 1994. Density and Viscosity of Concentrated Aqueous Solutions of Polyethylene Glycol. *Journal of Chemical & Engineering Data* 39: 611-14
Goodwin S, Gurtowski J, Ethe-Sayers S, Deshpande P, Schatz MC, McCombie WR. 2015. Oxford Nanopore sequencing, hybrid error correction, and de novo assembly of a eukaryotic genome. *Genome Res* 25: 1750-6
Goosen N, Moolenaar GF. 2008. Repair of UV damage in bacteria. *DNA Repair (Amst)* 7: 353-79
Gorbalenya AE, Koonin EV. 1993. Helicases - Amino-Acid-Sequence Comparisons and Structure-Function-Relationships. *Curr Opin Struc Biol* 3: 419-29
Gouaux JE, Braha O, Hobaugh MR, Song L, Cheley S, et al. 1994. Subunit stoichiometry of staphylococcal alpha-hemolysin in crystals and on membranes: a heptameric transmembrane pore. *Proc Natl Acad Sci U S A* 91: 12828-31
Grabarek Z, Gergely J. 1990. Zero-length crosslinking procedure with the use of active esters. *Anal Biochem* 185: 131-5
Grabowski B, Kelman Z. 2003. Archeal DNA replication: eukaryal proteins in a bacterial context. *Annu Rev Microbiol* 57: 487-516

Gray MW. 1989. The evolutionary origins of organelles. *Trends Genet* 5: 294-9
Grayling RA, Sandman K, Reeve JN. 1996. DNA stability and DNA binding proteins. *Adv Protein Chem* 48: 437-67
Graziano G, Merlino A. 2014. Molecular bases of protein halotolerance. *Biochim Biophys Acta* 1844: 850-8
Gribaldo S, Poole AM, Daubin V, Forterre P, Brochier-Armanet C. 2010. The origin of eukaryotes and their relationship with the Archaea: are we at a phylogenomic impasse? *Nat Rev Microbiol* 8: 743-52
Grogan DW. 2000. The question of DNA repair in hyperthermophilic archaea. *Trends Microbiol* 8: 180-5
Grogan DW. 2004. Stability and repair of DNA in hyperthermophilic Archaea. *Curr Issues Mol Biol* 6: 137-44
Grzenia D, Carlson J, Ranil Wickramasinghea S. 2008. Tangential flow filtration for virus purification. *Journal of Membrane Science* 321: 373-80
Guixa-Boixereu N, Lysnes K, Pedros-Alio C. 1999. Viral lysis and bacterivory during a phytoplankton bloom in a coastal water microcosm. *Appl Environ Microbiol* 65: 1949-58
Gupta S, Yeeles JT, Marians KJ. 2014. Regression of replication forks stalled by leading-strand template damage: II. Regression by RecA is inhibited by SSB. *J Biol Chem* 289: 28388-98
Guy CP, Bolt EL. 2005. Archaeal Hel308 helicase targets replication forks in vivo and in vitro and unwinds lagging strands. *Nucleic acids research* 33: 3678-90
Guy CP, Haldenby S, Brindley A, Walsh DA, Briggs GS, et al. 2006. Interactions of RadB, a DNA repair protein in archaea, with DNA and ATP. *J Mol Biol* 358: 46-56
Haldenby S. 2007. *Genetic analysis of RadB, a paralogue of the archaeal Rad51/RecA homologue, RadA*. University of Nottingham
Haldenby S, White MF, Allers T. 2009. RecA family proteins in archaea: RadA and its cousins. *Biochem Soc Trans* 37: 102-7
Hall MC, Matson SW. 1999. Helicase motifs: the engine that powers DNA unwinding. *Mol Microbiol* 34: 867-77
Hall RA, Bates S, Lenardon MD, Maccallum DM, Wagener J, et al. 2013. The Mnn2 mannosyltransferase family modulates mannoprotein fibril length, immune recognition and virulence of Candida albicans. *PLoS Pathog* 9: e1003276
Han ES, Cooper DL, Persky NS, Sutera VA, Jr., Whitaker RD, et al. 2006. RecJ exonuclease: substrates, products and interaction with SSB. *Nucleic Acids Res* 34: 1084-91
Hanawalt PC. 1994. Transcription-coupled repair and human disease. *Science* 266: 1957-8
Hanawalt PC, Spivak G. 2008. Transcription-coupled DNA repair: two decades of progress and surprises. *Nat Rev Mol Cell Biol* 9: 958-70
Harris JK, Kelley ST, Spiegelman GB, Pace NR. 2003. The genetic core of the universal ancestor. *Genome Res* 13: 407-12
Hartman AL, Norais C, Badger JH, Delmas S, Haldenby S, et al. 2010. The complete genome sequence of Haloferax volcanii DS2, a model archaeon. *PloS one* 5: e9605
Hawkins ER. 1973. A comparison of yeast mannan mutants by electron microscopy. *J Biol Chem* 248: 4671-3

Hawkins M, Malla S, Blythe MJ, Nieduszynski CA, Allers T. 2013. Accelerated growth in the absence of DNA replication origins. *Nature* 503: 544-7
Hefferin ML, Tomkinson AE. 2005. Mechanism of DNA double-strand break repair by non-homologous end joining. *DNA Repair (Amst)* 4: 639-48
Helleday T, Eshtad S, Nik-Zainal S. 2014. Mechanisms underlying mutational signatures in human cancers. *Nat Rev Genet* 15: 585-98
Heller RC, Marians KJ. 2005. The disposition of nascent strands at stalled replication forks dictates the pathway of replisome loading during restart. *Mol Cell* 17: 733-43
Heller RC, Marians KJ. 2005. Unwinding of the nascent lagging strand by Rep and PriA enables the direct restart of stalled replication forks. *J Biol Chem* 280: 34143-51
Hirata A, Murakami KS. 2009. Archaeal RNA polymerase. *Curr Opin Struc Biol* 19: 724-31
Hishida T, Iwasaki H, Yagi T, Shinagawa H. 1999. Role of walker motif A of RuvB protein in promoting branch migration of holliday junctions. Walker motif a mutations affect Atp binding, Atp hydrolyzing, and DNA binding activities of Ruvb. *J Biol Chem* 274: 25335-42
Holmes ML, Dyall-Smith ML. 2000. Sequence and expression of a halobacterial beta-galactosidase gene. *Mol Microbiol* 36: 114-22
Holmes ML, Nuttall SD, Dyall-Smith ML. 1991. Construction and use of halobacterial shuttle vectors and further studies on Haloferax DNA gyrase. *J Bacteriol* 173: 3807-13
Hong Y, Chu M, Li Y, Ni J, Sheng D, et al. 2012. Dissection of the functional domains of an archaeal Holliday junction helicase. *DNA Repair (Amst)* 11: 102-11
Hopfner KP, Craig L, Moncalian G, Zinkel RA, Usui T, et al. 2002. The Rad50 zinc-hook is a structure joining Mre11 complexes in DNA recombination and repair. *Nature* 418: 562-6
Hornblower B, Coombs A, Whitaker RD, Kolomeisky A, Picone SJ, et al. 2007. Single-molecule analysis of DNA-protein complexes using nanopores. *Nat Methods* 4: 315-17
Howley PM, Israel MA, Law MF, Martin MA. 1979. A rapid method for detecting and mapping homology between heterologous DNAs. Evaluation of polyomavirus genomes. *J Biol Chem* 254: 4876-83
Hu M, Zhang C, Mu Y, Shen Q, Feng Y. 2010. Indole affects biofilm formation in bacteria. *Indian J Microbiol* 50: 362-8
Hu Y, Raynard S, Sehorn MG, Lu X, Bussen W, et al. 2007. RECQL5/Recql5 helicase regulates homologous recombination and suppresses tumor formation via disruption of Rad51 presynaptic filaments. *Genes Dev* 21: 3073-84
Huber H, Hohn MJ, Rachel R, Fuchs T, Wimmer VC, Stetter KO. 2002. A new phylum of Archaea represented by a nanosized hyperthermophilic symbiont. *Nature* 417: 63-7
Huber H, Hohn MJ, Stetter KO, Rachel R. 2003. The phylum Nanoarchaeota: present knowledge and future perspectives of a unique form of life. *Res Microbiol* 154: 165-71
Hughes AL. 1994. The evolution of functionally novel proteins after gene duplication. *Proc Biol Sci* 256: 119-24
Humbard MA, Miranda HV, Lim JM, Krause DJ, Pritz JR, et al. 2010. Ubiquitin-like

small archaeal modifier proteins (SAMPs) in Haloferax volcanii. *Nature* 463: 54-60

Humbard MA, Zhou G, Maupin-Furlow JA. 2009. The N-terminal penultimate residue of 20S proteasome alpha1 influences its N(alpha) acetylation and protein levels as well as growth rate and stress responses of Haloferax volcanii. *J Bacteriol* 191: 3794-803

Ibarra A, Schwob E, Mendez J. 2008. Excess MCM proteins protect human cells from replicative stress by licensing backup origins of replication. *Proc Natl Acad Sci U S A* 105: 8956-61

Ip CL, Loose M, Tyson JR, de Cesare M, Brown BL, et al. 2015. MinION Analysis and Reference Consortium: Phase 1 data release and analysis. *F1000Res* 4: 1075

Ira G, Malkova A, Liberi G, Foiani M, Haber JE. 2003. Srs2 and Sgs1-Top3 suppress crossovers during double-strand break repair in yeast. *Cell* 115: 401-11

Ira G, Pellicioli A, Balijja A, Wang X, Fiorani S, et al. 2004. DNA end resection, homologous recombination and DNA damage checkpoint activation require CDK1. *Nature* 431: 1011-7

Ishibashi M, Sakashita K, Tokunaga H, Arakawa T, Tokunaga M. 2003. Activation of halophilic nucleoside diphosphate kinase by a non-ionic osmolyte, trimethylamine N-oxide. *J Protein Chem* 22: 345-51

Ishino S, Nishi Y, Oda S, Uemori T, Sagara T, et al. 2016. Identification of a mismatch-specific endonuclease in hyperthermophilic Archaea. *Nucleic Acids Res* 44: 2977-86

Jaenicke R. 2000. Stability and stabilization of globular proteins in solution. *J Biotechnol* 79: 193-203

Jasin M, Rothstein R. 2013. Repair of strand breaks by homologous recombination. *Cold Spring Harb Perspect Biol* 5: a012740

Jensen RB, Carreira A, Kowalczykowski SC. 2010. Purified human BRCA2 stimulates RAD51-mediated recombination. *Nature* 467: 678-83

Jiricny J. 2006. The multifaceted mismatch-repair system. *Nat Rev Mol Cell Biol* 7: 335-46

John SG, Mendez CB, Deng L, Poulos B, Kauffman AK, et al. 2011. A simple and efficient method for concentration of ocean viruses by chemical flocculation. *Environ Microbiol Rep* 3: 195-202

Jolley KA, Rapaport E, Hough DW, Danson MJ, Woods WG, Dyall-Smith ML. 1996. Dihydrolipoamide dehydrogenase from the halophilic archaeon Haloferax volcanii: homologous overexpression of the cloned gene. *J Bacteriol* 178: 3044-8

Jun SH, Reichlen MJ, Tajiri M, Murakami KS. 2011. Archaeal RNA polymerase and transcription regulation. *Crit Rev Biochem Mol* 46: 27-40

Kadyrov FA, Dzantiev L, Constantin N, Modrich P. 2006. Endonucleolytic function of MutLalpha in human mismatch repair. *Cell* 126: 297-308

Kaplan DL, O'Donnell M. 2003. Rho factor: transcription termination in four steps. *Curr Biol* 13: R714-6

Karanja KK, Cox SW, Duxin JP, Stewart SA, Campbell JL. 2012. DNA2 and EXO1 in replication-coupled, homology-directed repair and in the interplay between HDR and the FA/BRCA network. *Cell Cycle* 11: 3983-96

Karow JK, Constantinou A, Li JL, West SC, Hickson ID. 2000. The Bloom's syndrome

gene product promotes branch migration of holliday junctions. *Proc Natl Acad Sci U S A* 97: 6504-8
Kawate H, Sakumi K, Tsuzuki T, Nakatsuru Y, Ishikawa T, et al. 1998. Separation of killing and tumorigenic effects of an alkylating agent in mice defective in two of the DNA repair genes. *Proc Natl Acad Sci U S A* 95: 5116-20
Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJ. 2015. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat Protoc* 10: 845-58
Kelman LM, Kelman Z. 2004. Multiple origins of replication in archaea. *Trends Microbiol* 12: 399-401
Kelman Z, O'Donnell M. 1995. DNA polymerase III holoenzyme: structure and function of a chromosomal replicating machine. *Annu Rev Biochem* 64: 171-200
Kelman Z, White MF. 2005. Archaeal DNA replication and repair. *Curr Opin Microbiol* 8: 669-76
Kil YV, Baitin DM, Masui R, Bonch-Osmolovskaya EA, Kuramitsu S, Lanzov VA. 2000. Efficient strand transfer by the RadA recombinase from the hyperthermophilic archaeon Desulfurococcus amylolyticus. *J Bacteriol* 182: 130-4
Kimple ME, Brill AL, Pasker RL. 2013. Overview of affinity tags for protein purification. *Curr Protoc Protein Sci* 73: Unit 9 9
Klapstein K, Chou T, Bruinsma R. 2004. Physics of RecA-mediated homologous recognition. *Biophys J* 87: 1466-77
Kobayashi J, Tauchi H, Sakamoto S, Nakamura A, Morishima K, et al. 2002. NBS1 localizes to gamma-H2AX foci through interaction with the FHA/BRCT domain. *Curr Biol* 12: 1846-51
Komori K, Fujikane R, Shinagawa H, Ishino Y. 2002. Novel endonuclease in Archaea cleaving DNA with various branched structure. *Genes Genet Syst* 77: 227-41
Komori K, Hidaka M, Horiuchi T, Fujikane R, Shinagawa H, Ishino Y. 2004. Cooperation of the N-terminal Helicase and C-terminal endonuclease activities of Archaeal Hef protein in processing stalled replication forks. *J Biol Chem* 279: 53175-85
Komori K, Miyata T, DiRuggiero J, Holley-Shanks R, Hayashi I, et al. 2000. Both RadA and RadB are involved in homologous recombination in Pyrococcus furiosus. *J Biol Chem* 275: 33782-90
Koonin EV, Wolf YI. 2008. Genomics of bacteria and archaea: the emerging dynamic view of the prokaryotic world. *Nucleic Acids Res* 36: 6688-719
Koonin EV, Yutin N. 2014. The dispersed archaeal eukaryome and the complex archaeal ancestor of eukaryotes. *Cold Spring Harb Perspect Biol* 6: a016188
Korbel JO, Jensen LJ, von Mering C, Bork P. 2004. Analysis of genomic context: prediction of functional associations from conserved bidirectionally transcribed gene pairs. *Nat Biotechnol* 22: 911-7
Kowalczykowski SC. 2000. Initiation of genetic recombination and recombination-dependent replication. *Trends Biochem Sci* 25: 156-65
Krejci L, Altmannova V, Spirek M, Zhao X. 2012. Homologous recombination and its regulation. *Nucleic Acids Res* 40: 5795-818
Krejci L, Song B, Bussen W, Rothstein R, Mortensen UH, Sung P. 2002. Interaction with Rad51 is indispensable for recombination mediator function of Rad52. *J Biol Chem* 277: 40132-41
Krejci L, Van Komen S, Li Y, Villemain J, Reddy MS, et al. 2003. DNA helicase Srs2

disrupts the Rad51 presynaptic filament. *Nature* 423: 305-9
Kuchta RD, Stengel G. 2010. Mechanism and evolution of DNA primases. *Biochim Biophys Acta* 1804: 1180-9
Kukkaro P, Bamford DH. 2009. Virus-host interactions in environments with a wide range of ionic strengths. *Environ Microbiol Rep* 1: 71-7
Kuriyan J, O'Donnell M. 1993. Sliding clamps of DNA polymerases. *J Mol Biol* 234: 915-25
Kurokawa Y, Murayama Y, Haruta-Takahashi N, Urabe I, Iwasaki H. 2008. Reconstitution of DNA strand exchange mediated by Rhp51 recombinase and two mediators. *PLoS Biol* 6: e88
Kvaratskhelia M, Wardleworth BN, White MF. 2001. Multiple Holliday junction resolving enzyme activities in the Crenarchaeota and Euryarchaeota. *FEBS Lett* 491: 243-6
Lake JA. 1989. Origin of the eukaryotic nucleus: eukaryotes and eocytes are genotypically related. *Can J Microbiol* 35: 109-18
Lam WL, Doolittle WF. 1989. Shuttle vectors for the archaebacterium Halobacterium volcanii. *Proc Natl Acad Sci U S A* 86: 5478-82
Langer D, Hain J, Thuriaux P, Zillig W. 1995. Transcription in archaea: similarity to that in eucarya. *Proc Natl Acad Sci U S A* 92: 5768-72
Lankinen MH, Vilpo LM, Vilpo JA. 1996. UV- and gamma-irradiation-induced DNA single-strand breaks and their repair in human blood granulocytes and lymphocytes. *Mutat Res* 352: 31-8
Lanyi JK. 1974. Salt-dependent properties of proteins from extremely halophilic bacteria. *Bacteriol Rev* 38: 272-90
Large A, Stamme C, Lange C, Duan Z, Allers T, et al. 2007. Characterization of a tightly controlled promoter of the halophilic archaeon Haloferax volcanii and its use in the analysis of the essential cct1 gene. *Mol Microbiol* 66: 1092-106
Laver T, Harrison J, O'Neill PA, Moore K, Farbos A, et al. 2015. Assessing the performance of the Oxford Nanopore Technologies MinION. *Biomol Detect Quantif* 3: 1-8
Lestini R, Duan Z, Allers T. 2010. The archaeal Xpf/Mus81/FANCM homolog Hef and the Holliday junction resolvase Hjc define alternative pathways that are essential for cell viability in Haloferax volcanii. *DNA Repair (Amst)* 9: 994-1002
Lestini R, Laptenok SP, Kuhn J, Hink MA, Schanne-Klein MC, et al. 2013. Intracellular dynamics of archaeal FANCM homologue Hef in response to halted DNA replication. *Nucleic Acids Res* 41: 10358-70
Li C, Kappock TJ, Stubbe J, Weaver TM, Ealick SE. 1999. X-ray crystal structure of aminoimidazole ribonucleotide synthetase (PurM), from the Escherichia coli purine biosynthetic pathway at 2.5 A resolution. *Structure* 7: 1155-66
Li G, Young KD. 2013. Indole production by the tryptophanase TnaA in Escherichia coli is determined by the amount of exogenous tryptophan. *Microbiology* 159: 402-10
Li GM. 2008. Mechanisms and functions of DNA mismatch repair. *Cell Res* 18: 85-98
Li WQ, Hu N, Hyland PL, Gao Y, Wang ZM, et al. 2013. Genetic variants in DNA repair pathway genes and risk of esophageal squamous cell carcinoma and gastric adenocarcinoma in a Chinese population. *Carcinogenesis* 34: 1536-42
Li YF, Heelis PF, Sancar A. 1991. Active site of DNA photolyase: tryptophan-306 is the

intrinsic hydrogen atom donor essential for flavin radical photoreduction and DNA repair in vitro. *Biochemistry* 30: 6322-9
Li Z, Lu S, Hou G, Ma X, Sheng D, et al. 2008. Hjm/Hel308A DNA helicase from Sulfolobus tokodaii promotes replication fork regression and interacts with Hjc endonuclease in vitro. *J Bacteriol* 190: 3006-17
Liang C, Marsit CJ, Houseman EA, Butler R, Nelson HH, et al. 2012. Gene-environment interactions of novel variants associated with head and neck cancer. *Head Neck* 34: 1111-8
Lin Z, Nei M, Ma H. 2007. The origins and early evolution of DNA mismatch repair genes--multiple horizontal gene transfers and co-evolution. *Nucleic Acids Res* 35: 7591-603
Lindahl T. 1993. Instability and decay of the primary structure of DNA. *Nature* 362: 709-15
Lindas AC, Karlsson EA, Lindgren MT, Ettema TJ, Bernander R. 2008. A unique cell division machinery in the Archaea. *Proc Natl Acad Sci U S A* 105: 18942-6
Lio YC, Mazin AV, Kowalczykowski SC, Chen DJ. 2003. Complex formation by the human Rad51B and Rad51C DNA repair proteins and their activities in vitro. *J Biol Chem* 278: 2469-78
Lis JT, Schleif R. 1975. Size fractionation of double-stranded DNA by precipitation with polyethylene glycol. *Nucleic Acids Res* 2: 383-9
Little JW, Mount DW. 1982. The SOS regulatory system of Escherichia coli. *Cell* 29: 11-22
Liu J, Renault L, Veaute X, Fabre F, Stahlberg H, Heyer WD. 2011. Rad51 paralogues Rad55-Rad57 balance the antirecombinase Srs2 in Rad51 filament formation. *Nature* 479: 245-8
Liu Y, Schroder J, Schmidt B. 2013. Musket: a multistage k-mer spectrum-based error corrector for Illumina sequence data. *Bioinformatics* 29: 308-15
Lohman TM, Bjornson KP. 1996. Mechanisms of helicase-catalyzed DNA unwinding. *Annu Rev Biochem* 65: 169-214
Londei P. 2005. Evolution of translational initiation: new insights from the archaea. *FEMS Microbiol Rev* 29: 185-200
Luebben SW, Kawabata T, Akre MK, Lee WL, Johnson CS, et al. 2013. Helq acts in parallel to Fancc to suppress replication-associated genome instability. *Nucleic Acids Res* 41: 10283-97
Luk AW, Williams TJ, Erdmann S, Papke RT, Cavicchioli R. 2014. Viruses of haloarchaea. *Life (Basel)* 4: 681-715
Lundgren M, Andersson A, Chen L, Nilsson P, Bernander R. 2004. Three replication origins in Sulfolobus species: synchronous initiation of chromosome replication and asynchronous termination. *Proc Natl Acad Sci U S A* 101: 7046-51
Lutsenko E, Bhagwat AS. 1999. Principal causes of hot spots for cytosine to thymine mutations at sites of cytosine methylation in growing cells. A model, its experimental support and implications. *Mutat Res* 437: 11-20
Machwe A, Xiao L, Groden J, Orren DK. 2006. The Werner and Bloom syndrome proteins catalyze regression of a model replication fork. *Biochemistry* 45: 13939-46
Mack M, van Loon AP, Hohmann HP. 1998. Regulation of riboflavin biosynthesis in

Bacillus subtilis is affected by the activity of the flavokinase/flavin adenine dinucleotide synthetase encoded by ribC. *J Bacteriol* 180: 950-5

Mackintosh SG, Raney KD. 2006. DNA unwinding and protein displacement by superfamily 1 and superfamily 2 helicases. *Nucleic acids research* 34: 4154-9

MacNeill SA. 2010. Structure and function of the GINS complex, a key component of the eukaryotic replisome. *Biochem J* 425: 489-500

Magoc T, Salzberg SL. 2011. FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* 27: 2957-63

Maitra RD, Kim J, Dunbar WB. 2012. Recent advances in nanopore sequencing. *Electrophoresis* 33: 3418-28

Makarova KS, Koonin EV, Kelman Z. 2012. The CMG (CDC45/RecJ, MCM, GINS) complex is a conserved component of the DNA replication system in all archaea and eukaryotes. *Biol Direct* 7: 7

Makarova KS, Wolf YI, van der Oost J, Koonin EV. 2009. Prokaryotic homologs of Argonaute proteins are predicted to function as key components of a novel system of defense against mobile genetic elements. *Biol Direct* 4: 29

Makarova KS, Yutin N, Bell SD, Koonin EV. 2010. Evolution of diverse cell division and vesicle formation systems in Archaea. *Nat Rev Microbiol* 8: 731-41

Mankouri HW, Craig TJ, Morgan A. 2002. SGS1 is a multicopy suppressor of srs2: functional overlap between DNA helicases. *Nucleic Acids Res* 30: 1103-13

Mankouri HW, Ngo HP, Hickson ID. 2007. Shu proteins promote the formation of homologous recombination intermediates that are processed by Sgs1-Rmi1-Top3. *Mol Biol Cell* 18: 4062-73

Manosas M, Perumal SK, Bianco PR, Ritort F, Benkovic SJ, Croquette V. 2013. RecG and UvsW catalyse robust DNA rewinding critical for stalled DNA replication fork rescue. *Nat Commun* 4: 2368

Manrao EA, Derrington IM, Laszlo AH, Langford KW, Hopper MK, et al. 2012. Reading DNA at single-nucleotide resolution with a mutant MspA nanopore and phi29 DNA polymerase. *Nat Biotechnol* 30: 349-53

Marechal A, Zou L. 2013. DNA damage sensing by the ATM and ATR kinases. *Cold Spring Harb Perspect Biol* 5

Marini F, Kim N, Schuffert A, Wood RD. 2003. POLN, a nuclear PolA family DNA polymerase homologous to the DNA cross-link sensitivity protein Mus308. *J Biol Chem* 278: 32014-9

Marini F, Wood RD. 2002. A human DNA helicase homologous to the DNA cross-link sensitivity protein Mus308. *J Biol Chem* 277: 8716-23

Marintcheva B, Weller SK. 2003. Helicase motif la is involved in single-strand DNA-binding and helicase activities of the herpes simplex virus type 1 origin-binding protein, UL9. *J Virol* 77: 2477-88

Matos J, Blanco MG, Maslen S, Skehel JM, West SC. 2011. Regulatory control of the resolution of DNA recombination intermediates during meiosis and mitosis. *Cell* 147: 158-72

Matson SW, Bean DW, George JW. 1994. DNA helicases: enzymes with essential roles in all aspects of DNA metabolism. *Bioessays* 16: 13-22

Matsumoto Y, Kim K. 1995. Excision of deoxyribose phosphate residues by DNA polymerase beta during DNA repair. *Science* 269: 699-702

Matsuoka S, Ballif BA, Smogorzewska A, McDonald ER, 3rd, Hurov KE, et al. 2007. ATM and ATR substrate analysis reveals extensive protein networks responsive to DNA damage. *Science* 316: 1160-6

Mayle R, Campbell IM, Beck CR, Yu Y, Wilson M, et al. 2015. DNA REPAIR. Mus81 and converging forks limit the mutagenicity of replication fork breakage. *Science* 349: 742-7

Mazin AV, Mazina OM, Bugreev DV, Rossi MJ. 2010. Rad54, the motor of homologous recombination. *DNA Repair (Amst)* 9: 286-302

McClendon TB, Sullivan M, Bernstein K, Yanowitz J. 2016. Promotion of Homologous Recombination by SWS-1 in Complex with RAD-51 Paralogs in Caenorhabditis elegans. *Genetics*

McCready S, Muller JA, Boubriak I, Berquist BR, Ng WL, DasSarma S. 2005. UV irradiation induces homologous recombination genes in the model archaeon, Halobacterium sp. NRC-1. *Saline Systems* 1: 3

McEntee K, Weinstock GM, Lehman IR. 1979. Initiation of general recombination catalyzed in vitro by the recA protein of Escherichia coli. *Proc Natl Acad Sci U S A* 76: 2615-9

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

McGlynn P, Lloyd RG. 2002. Recombinational repair and restart of damaged replication forks. *Nat Rev Mol Cell Biol* 3: 859-70

McGlynn P, Lloyd RG. 2002. Genome stability and the processing of damaged replication forks by RecG. *Trends Genet* 18: 413-9

McIlwraith MJ, West SC. 2008. DNA repair synthesis facilitates RAD52-mediated second-end capture during DSB repair. *Mol Cell* 29: 510-6

Meetei AR, Medhurst AL, Ling C, Xue Y, Singh TR, et al. 2005. A human ortholog of archaeal DNA repair protein Hef is defective in Fanconi anemia complementation group M. *Nat Genet* 37: 958-63

Mevarech M, Frolow F, Gloss LM. 2000. Halophilic enzymes: proteins with a grain of salt. *Biophys Chem* 86: 155-64

Michel B, Flores MJ, Viguera E, Grompone G, Seigneur M, Bidnenko V. 2001. Rescue of arrested replication forks by homologous recombination. *Proc Natl Acad Sci U S A* 98: 8181-8

Mimitou EP, Symington LS. 2008. Sae2, Exo1 and Sgs1 collaborate in DNA double-strand break processing. *Nature* 455: 770-4

Modrich P. 2006. Mechanisms in eukaryotic mismatch repair. *J Biol Chem* 281: 30305-9

Mohaghegh P, Karow JK, Brosh RM, Jr., Bohr VA, Hickson ID. 2001. The Bloom's and Werner's syndrome proteins are DNA structure-specific helicases. *Nucleic Acids Res* 29: 2843-9

Moldovan GL, Pfander B, Jentsch S. 2007. PCNA, the maestro of the replication fork. *Cell* 129: 665-79

Morimatsu K, Kowalczykowski SC. 2003. RecFOR proteins load RecA protein onto gapped DNA to accelerate DNA strand exchange: a universal step of recombinational repair. *Mol Cell* 11: 1337-47

Morimatsu K, Kowalczykowski SC. 2014. RecQ helicase and RecJ nuclease provide complementary functions to resect DNA for homologous recombination. *Proc Natl*

Acad Sci U S A 111: E5133-42
Mullakhanbhai MF, Larsen H. 1975. Halobacterium volcanii spec. nov., a Dead Sea halobacterium with a moderate salt requirement. *Arch Microbiol* 104: 207-14
Muzzini DM, Plevani P, Boulton SJ, Cassata G, Marini F. 2008. Caenorhabditis elegans POLQ-1 and HEL-308 function in two distinct DNA interstrand cross-link repair pathways. *DNA Repair (Amst)* 7: 941-50
Nair PA, Smith P, Shuman S. 2010. Structure of bacterial LigD 3'-phosphoesterase unveils a DNA repair superfamily. *Proc Natl Acad Sci U S A* 107: 12822-7
Nakamura A, Fujihashi M, Aono R, Sato T, Nishiba Y, et al. 2012. Dynamic, ligand-dependent conformational change triggers reaction of ribose-1,5-bisphosphate isomerase from Thermococcus kodakarensis KOD1. *J Biol Chem* 287: 20784-96
Neale MJ, Keeney S. 2006. Clarifying the mechanics of DNA strand exchange in meiotic recombination. *Nature* 442: 153-8
Neuwald AF, Aravind L, Spouge JL, Koonin EV. 1999. AAA+: A class of chaperone-like ATPases associated with the assembly, operation, and disassembly of protein complexes. *Genome Res* 9: 27-43
Newman TJ, Mamun MA, Nieduszynski CA, Blow JJ. 2013. Replisome stall events have shaped the distribution of replication origins in the genomes of yeasts. *Nucleic Acids Res* 41: 9705-18
Neylon C, Kralicek AV, Hill TM, Dixon NE. 2005. Replication termination in Escherichia coli: structure and antihelicase activity of the Tus-Ter complex. *Microbiol Mol Biol Rev* 69: 501-26
Nguyen B, Sokoloski J, Galletto R, Elson EL, Wold MS, Lohman TM. 2014. Diffusion of human replication protein A along single-stranded DNA. *J Mol Biol* 426: 3246-61
Nishino T, Komori K, Ishino Y, Morikawa K. 2005. Structural and functional analyses of an archaeal XPF/Rad1/Mus81 nuclease: asymmetric DNA binding and cleavage mechanisms. *Structure* 13: 1183-92
Norais C, Hawkins M, Hartman AL, Eisen JA, Myllykallio H, Allers T. 2007. Genetic and physical mapping of DNA replication origins in Haloferax volcanii. *PLoS Genet* 3: e77
Nowak MA, Boerlijst MC, Cooke J, Smith JM. 1997. Evolution of genetic redundancy. *Nature* 388: 167-71
Nunoura T, Takaki Y, Kakuta J, Nishi S, Sugahara J, et al. 2011. Insights into the evolution of Archaea and eukaryotic protein modifier systems revealed by the genome of a novel archaeal group. *Nucleic Acids Res* 39: 3204-23
Nuttall SD, Dyallsmith ML. 1993. Hf1 and Hf2 - Novel Bacteriophages of Halophilic Archaea. *Virology* 197: 678-84
O'Donnell M, Langston L, Stillman B. 2013. Principles and concepts of DNA replication in bacteria, archaea, and eukarya. *Cold Spring Harb Perspect Biol* 5
Oh SD, Lao JP, Hwang PY, Taylor AF, Smith GR, Hunter N. 2007. BLM ortholog, Sgs1, prevents aberrant crossing-over by suppressing formation of multichromatid joint molecules. *Cell* 130: 259-72
Oksenych V, Coin F. 2010. The long unwinding road: XPB and XPD helicases in damaged DNA opening. *Cell Cycle* 9: 90-6
Olsen GJ, Woese CR. 1997. Archaeal genomics: an overview. *Cell* 89: 991-4
Olsson M, Lindahl T. 1980. Repair of alkylated DNA in Escherichia coli. Methyl group

transfer from O6-methylguanine to a protein cysteine residue. *J Biol Chem* 255: 10569-71
Onesti S, MacNeill SA. 2013. Structure and evolutionary origins of the CMG complex. *Chromosoma* 122: 47-53
Oprian DD, Coon MJ. 1982. Oxidation-reduction states of FMN and FAD in NADPH-cytochrome P-450 reductase during reduction by NADPH. *J Biol Chem* 257: 8935-44
Oren A. 1999. Bioenergetic aspects of halophilism. *Microbiol Mol Biol Rev* 63: 334-48
Oren A. 2008. Microbial life at high salt concentrations: phylogenetic and metabolic diversity. *Saline systems* 4: 2
Oren A, Heldal M, Norland S, Galinski EA. 2002. Intracellular ion and organic solute concentrations of the extremely halophilic bacterium Salinibacter ruber. *Extremophiles* 6: 491-8
Ortenberg R, Rozenblatt-Rosen O, Mevarech M. 2000. The extremely halophilic archaeon Haloferax volcanii has two very different dihydrofolate reductases. *Mol Microbiol* 35: 1493-505
Oyama T, Oka H, Mayanagi K, Shirai T, Matoba K, et al. 2009. Atomic structures and functional implications of the archaeal RecQ-like helicase Hjm. *BMC Struct Biol* 9: 2
Ozaki S, Katayama T. 2009. DnaA structure, function, and dynamics in the initiation at the chromosomal origin. *Plasmid* 62: 71-82
Pace NR. 1997. A molecular view of microbial diversity and the biosphere. *Science* 276: 734-40
Pages V, Fuchs RP. 2003. Uncoupling of leading- and lagging-strand DNA replication during lesion bypass in vivo. *Science* 300: 1300-3
Paques F, Haber JE. 1999. Multiple pathways of recombination induced by double-strand breaks in Saccharomyces cerevisiae. *Microbiol Mol Biol Rev* 63: 349-404
Parmar K, D'Andrea A, Niedernhofer LJ. 2009. Mouse models of Fanconi anemia. *Mutat Res* 668: 133-40
Patel SS, Picha KM. 2000. Structure and function of hexameric helicases. *Annu Rev Biochem* 69: 651-97
Pauling C. 1982. Bacteriophages of Halobacterium-Halobium - Isolation from Fermented Fish Sauce and Primary Characterization. *Can J Microbiol* 28: 916-21
Pauling C. 1982. Bacteriophages of Halobacterium halobium: isolated from fermented fish sauce and primary characterization. *Can J Microbiol* 28: 916-21
Paull TT, Gellert M. 2000. A mechanistic basis for Mre11-directed DNA joining at microhomologies. *Proc Natl Acad Sci U S A* 97: 6409-14
Pause A, Sonenberg N. 1992. Mutational Analysis of a Dead Box Rna Helicase - the Mammalian Translation Initiation-Factor Eif-4a. *Embo Journal* 11: 2643-54
Peeters E, Charlier D. 2010. The Lrp family of transcription regulators in archaea. *Archaea* 2010: 750457
Pelve EA, Lindas AC, Martens-Habbena W, de la Torre JR, Stahl DA, Bernander R. 2011. Cdv-based cell division and cell cycle organization in the thaumarchaeon Nitrosopumilus maritimus. *Mol Microbiol* 82: 555-66
Pereira SL, Reeve JN. 1998. Histones and nucleosomes in Archaea and Eukarya: a comparative analysis. *Extremophiles* 2: 141-8

Pester M, Schleper C, Wagner M. 2011. The Thaumarchaeota: an emerging view of their phylogeny and ecophysiology. *Curr Opin Microbiol* 14: 300-6

Petermann E, Helleday T. 2010. Pathways of mammalian replication fork restart. *Nat Rev Mol Cell Biol* 11: 683-7

Petit C, Sancar A. 1999. Nucleotide excision repair: from E. coli to man. *Biochimie* 81: 15-25

Piersen CE, McCullough AK, Lloyd RS. 2000. AP lyases and dRPases: commonality of mechanism. *Mutat Res* 459: 43-53

Pietila MK, Demina TA, Atanasova NS, Oksanen HM, Bamford DH. 2014. Archaeal viruses and bacteriophages: comparisons and contrasts. *Trends Microbiol* 22: 334-44

Pietila MK, Laurinmaki P, Russell DA, Ko CC, Jacobs-Sera D, et al. 2013. Insights into head-tailed viruses infecting extremely halophilic archaea. *J Virol* 87: 3248-60

Pietila MK, Roine E, Paulin L, Kalkkinen N, Bamford DH. 2009. An ssDNA virus infecting archaea: a new lineage of viruses with a membrane envelope. *Mol Microbiol* 72: 307-19

Pisani D, Cotton JA, McInerney JO. 2007. Supertrees disentangle the chimerical origin of eukaryotic genomes. *Mol Biol Evol* 24: 1752-60

Pitcher RS, Brissett NC, Doherty AJ. 2007. Nonhomologous end-joining in bacteria: a microbial perspective. *Annu Rev Microbiol* 61: 259-82

Pitcher RS, Tonkin LM, Green AJ, Doherty AJ. 2005. Domain structure of a NHEJ DNA repair ligase from Mycobacterium tuberculosis. *J Mol Biol* 351: 531-44

Prakash R, Satory D, Dray E, Papusha A, Scheller J, et al. 2009. Yeast Mph1 helicase dissociates Rad51-made D-loops: implications for crossover control in mitotic recombination. *Genes Dev* 23: 67-79

Ragunathan K, Liu C, Ha T. 2012. RecA filament sliding on DNA facilitates homology search. *Elife* 1: e00067

Ralf C, Hickson ID, Wu L. 2006. The Bloom's syndrome helicase can promote the regression of a model replication fork. *J Biol Chem* 281: 22839-46

Ramakrishnan C, Dani VS, Ramasarma T. 2002. A conformational analysis of Walker motif A [GXXXXGKT (S)] in nucleotide-binding and other proteins. *Protein Eng* 15: 783-98

Rastogi S, Liberles DA. 2005. Subfunctionalization of duplicated genes as a transition state to neofunctionalization. *BMC Evol Biol* 5: 28

Ravanat JL, Douki T, Cadet J. 2001. Direct and indirect effects of UV radiation on DNA and its components. *J Photochem Photobiol B* 63: 88-102

Reeve JN, Sandman K, Daniels CJ. 1997. Archaeal histones, nucleosomes, and transcription initiation. *Cell* 89: 999-1002

Ren B, Kuhn J, Meslet-Cladiere L, Briffotaux J, Norais C, et al. 2009. Structure and function of a novel endonuclease acting on branched DNA substrates. *EMBO J* 28: 2479-89

Reuter CJ, Maupin-Furlow JA. 2004. Analysis of proteasome-dependent proteolysis in Haloferax volcanii cells, using short-lived green fluorescent proteins. *Appl Environ Microbiol* 70: 7530-8

Reyes-Lamothe R, Nicolas E, Sherratt DJ. 2012. Chromosome replication and segregation in bacteria. *Annu Rev Genet* 46: 121-43

Rhodes ME, Spear JR, Oren A, House CH. 2011. Differences in lateral gene transfer in hypersaline versus thermal environments. *BMC Evol Biol* 11: 199

Richards JD, Cubeddu L, Roberts J, Liu H, White MF. 2008. The archaeal XPB protein is a ssDNA-dependent ATPase with a novel partner. *J Mol Biol* 376: 634-44

Richards JD, Johnson KA, Liu H, McRobbie AM, McMahon S, et al. 2008. Structure of the DNA repair helicase hel308 reveals DNA binding and autoinhibitory domains. *The Journal of biological chemistry* 283: 5118-26

Ritsert K, Huber R, Turk D, Ladenstein R, Schmidt-Base K, Bacher A. 1995. Studies on the lumazine synthase/riboflavin synthase complex of Bacillus subtilis: crystal structure analysis of reconstituted, icosahedral beta-subunit capsids with bound substrate analogue inhibitor at 2.4 A resolution. *J Mol Biol* 253: 151-67

Rivera MC, Lake JA. 2004. The ring of life provides evidence for a genome fusion origin of eukaryotes. *Nature* 431: 152-5

Roberts JA, Bell SD, White MF. 2003. An archaeal XPF repair endonuclease dependent on a heterotrimeric PCNA. *Mol Microbiol* 48: 361-71

Rocak S, Linder P. 2004. Dead-box proteins: The driving forces behind RNA metabolism. *Nat Rev Mol Cell Bio* 5: 232-41

Rocha EP, Cornet E, Michel B. 2005. Comparative and evolutionary analysis of the bacterial homologous recombination systems. *PLoS Genet* 1: e15

Saini N, Ramakrishnan S, Elango R, Ayyar S, Zhang Y, et al. 2013. Migrating bubble during break-induced replication drives conservative DNA synthesis. *Nature* 502: 389-92

Saintigny Y, Makienko K, Swanson C, Emond MJ, Monnat RJ, Jr. 2002. Homologous recombination resolution defect in werner syndrome. *Mol Cell Biol* 22: 6971-8

Sakofsky CJ, Ayyar S, Malkova A. 2012. Break-induced replication and genome stability. *Biomolecules* 2: 483-504

Samson RY, Obita T, Freund SM, Williams RL, Bell SD. 2008. A role for the ESCRT system in cell division in archaea. *Science* 322: 1710-3

San Filippo J, Sung P, Klein H. 2008. Mechanism of eukaryotic homologous recombination. *Annu Rev Biochem* 77: 229-57

Sancar A, Lindsey-Boltz LA, Unsal-Kacmaz K, Linn S. 2004. Molecular mechanisms of mammalian DNA repair and the DNA damage checkpoints. *Annu Rev Biochem* 73: 39-85

Sandman K, Reeve JN. 2005. Archaeal chromatin proteins: different structures but common function? *Curr Opin Microbiol* 8: 656-61

Sanger F, Nicklen S, Coulson AR. 1977. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A* 74: 5463-7

Sanjuan R, Nebot MR, Chirico N, Mansky LM, Belshaw R. 2010. Viral mutation rates. *J Virol* 84: 9733-48

Santos F, Moreno-Paz M, Meseguer I, Lopez C, Rossello-Mora R, et al. 2011. Metatranscriptomic analysis of extremely halophilic viral communities. *ISME J* 5: 1621-33

Santos F, Yarza P, Parro V, Briones C, Anton J. 2010. The metavirome of a hypersaline environment. *Environ Microbiol* 12: 2965-76

Sartori AA, Jiricny J. 2003. Enzymology of base excision repair in the hyperthermophilic archaeon Pyrobaculum aerophilum. *J Biol Chem* 278: 24563-76

Sato T, Atomi H, Imanaka T. 2007. Archaeal type III RuBisCOs function in a pathway for AMP metabolism. *Science* 315: 1003-6

Schaaper RM. 1993. Base selection, proofreading, and mismatch repair during DNA replication in Escherichia coli. *J Biol Chem* 268: 23762-5

Scharer OD. 2013. Nucleotide excision repair in eukaryotes. *Cold Spring Harb Perspect Biol* 5: a012609

Schroder E, Jonsson T, Poole L. 2003. Hydroxyapatite chromatography: altering the phosphate-dependent elution profile of protein as a function of pH. *Anal Biochem* 313: 176-8

Schwartz EK, Heyer WD. 2011. Processing of joint molecule intermediates by structure-selective endonucleases during homologous recombination in eukaryotes. *Chromosoma* 120: 109-27

Sclafani RA, Holzen TM. 2007. Cell cycle regulation of DNA replication. *Annu Rev Genet* 41: 237-80

Sehorn MG, Sigurdsson S, Bussen W, Unger VM, Sung P. 2004. Human meiotic recombinase Dmc1 promotes ATP-dependent homologous DNA strand exchange. *Nature* 429: 433-7

Seigneur M, Bidnenko V, Ehrlich SD, Michel B. 1998. RuvAB acts at arrested replication forks. *Cell* 95: 419-30

Sencilo A, Jacobs-Sera D, Russell DA, Ko CC, Bowman CA, et al. 2013. Snapshot of haloarchaeal tailed virus genomes. *RNA Biol* 10: 803-16

Serata M, Iino T, Yasuda E, Sako T. 2012. Roles of thioredoxin and thioredoxin reductase in the resistance to oxidative stress in Lactobacillus casei. *Microbiology* 158: 953-62

Sfeir A, Symington LS. 2015. Microhomology-Mediated End Joining: A Back-up Survival Mechanism or Dedicated Pathway? *Trends Biochem Sci* 40: 701-14

Sharples GJ, Ingleston SM, Lloyd RG. 1999. Holliday junction processing in bacteria: insights from the evolutionary conservation of RuvABC, RecG, and RusA. *J Bacteriol* 181: 5543-50

Shen JC, Loeb LA. 2000. The Werner syndrome gene: the molecular basis of RecQ helicase-deficiency diseases. *Trends Genet* 16: 213-20

Shendure J, Ji HL. 2008. Next-generation DNA sequencing. *Nat Biotechnol* 26: 1135-45

Shereda RD, Kozlov AG, Lohman TM, Cox MM, Keck JL. 2008. SSB as an organizer/mobilizer of genome maintenance complexes. *Crit Rev Biochem Mol Biol* 43: 289-318

Shimmin LC, Dennis PP. 1996. Conserved sequence elements involved in regulation of ribosomal protein gene expression in halophilic archaea. *J Bacteriol* 178: 4737-41

Shin DS, Pellegrini L, Daniels DS, Yelent B, Craig L, et al. 2003. Full-length archaeal Rad51 structure and mutants: mechanisms for RAD51 assembly and control by BRCA2. *EMBO J* 22: 4566-76

Sigurdsson S, Van Komen S, Bussen W, Schild D, Albala JS, Sung P. 2001. Mediator function of the human Rad51B-Rad51C complex in Rad51/RPA-catalyzed DNA strand exchange. *Genes Dev* 15: 3308-18

Singleton MR, Dillingham MS, Gaudier M, Kowalczykowski SC, Wigley DB. 2004. Crystal structure of RecBCD enzyme reveals a machine for processing DNA breaks. *Nature* 432: 187-93

Singleton MR, Dillingham MS, Wigley DB. 2007. Structure and mechanism of helicases and nucleic acid translocases. *Annu Rev Biochem* 76: 23-50
Singleton MR, Scaife S, Wigley DB. 2001. Structural analysis of DNA replication fork reversal by RecG. *Cell* 107: 79-89
Singleton MR, Wigley DB. 2002. Modularity and specialization in superfamily 1 and 2 helicases. *J Bacteriol* 184: 1819-26
Sinha N, Nussinov R. 2001. Point mutations and sequence variability in proteins: redistributions of preexisting populations. *Proc Natl Acad Sci U S A* 98: 3139-44
Sinha PK, Castro-Guerrero N, Patki G, Sato M, Torres-Bacete J, et al. 2015. Conserved amino acid residues of the NuoD segment important for structure and function of Escherichia coli NDH-1 (complex I). *Biochemistry* 54: 753-64
Sinha RP, Hader DP. 2002. UV-induced DNA damage and repair: a review. *Photochem Photobiol Sci* 1: 225-36
Smith J, Modrich P. 1996. Mutation detection with MutH, MutL, and MutS mismatch repair proteins. *Proc Natl Acad Sci U S A* 93: 4374-9
Smith P, Nair PA, Das U, Zhu H, Shuman S. 2011. Structures and activities of archaeal members of the LigD 3'-phosphoesterase DNA repair enzyme superfamily. *Nucleic Acids Res* 39: 3310-20
Snider J, Thibault G, Houry WA. 2008. The AAA+ superfamily of functionally diverse proteins. *Genome Biol* 9: 216
Sommers JA, Rawtani N, Gupta R, Bugreev DV, Mazin AV, et al. 2009. FANCJ uses its motor ATPase to destabilize protein-DNA complexes, unwind triplexes, and inhibit RAD51 strand exchange. *J Biol Chem* 284: 7505-17
Soultanas P, Wigley DB. 2001. Unwinding the 'Gordian knot' of helicase action. *Trends Biochem Sci* 26: 47-54
Stantial N, Dumpe J, Pietrosimone K, Baltazar F, Crowley DJ. 2016. Transcription-coupled repair of UV damage in the halophilic archaea. *DNA Repair (Amst)* 41: 63-68
Stark JM, Pierce AJ, Oh J, Pastink A, Jasin M. 2004. Genetic steps of mammalian homologous repair with distinct mutagenic consequences. *Mol Cell Biol* 24: 9305-16
Strillinger E, Grotzinger SW, Allers T, Eppinger J, Weuster-Botz D. 2016. Production of halophilic proteins using Haloferax volcanii H1895 in a stirred-tank bioreactor. *Appl Microbiol Biotechnol* 100: 1183-95
Stroud A, Liddell S, Allers T. 2012. Genetic and Biochemical Identification of a Novel Single-Stranded DNA-Binding Complex in Haloferax volcanii. *Front Microbiol* 3: 224
Sugasawa K, Masutani C, Uchida A, Maekawa T, van der Spek PJ, et al. 1996. HHR23B, a human Rad23 homolog, stimulates XPC protein in nucleotide excision repair in vitro. *Mol Cell Biol* 16: 4852-61
Sung P. 1997. Yeast Rad55 and Rad57 proteins form a heterodimer that functions with replication protein A to promote DNA strand exchange by Rad51 recombinase. *Genes Dev* 11: 1111-21
Sung P, Krejci L, Van Komen S, Sehorn MG. 2003. Rad51 recombinase and recombination mediators. *J Biol Chem* 278: 42729-32
Svejstrup JQ. 2002. Mechanisms of transcription-coupled DNA repair. *Nat Rev Mol Cell*

Biol 3: 21-9
Svendsen JM, Smogorzewska A, Sowa ME, O'Connell BC, Gygi SP, et al. 2009. Mammalian BTBD12/SLX4 assembles a Holliday junction resolvase and is required for DNA repair. *Cell* 138: 63-77
Symington LS, Gautier J. 2011. Double-strand break end resection and repair pathway choice. *Annu Rev Genet* 45: 247-71
Tafel AA, Wu L, McHugh PJ. 2011. Human HEL308 localizes to damaged replication forks and unwinds lagging strand structures. *The Journal of biological chemistry* 286: 15832-40
Takata K, Reh S, Tomida J, Person MD, Wood RD. 2013. Human DNA helicase HELQ participates in DNA interstrand crosslink tolerance with ATR and RAD51 paralogs. *Nat Commun* 4: 2338
Takata M, Yamamoto K, Matsushita N, Kitao H, Hirano S, Ishiai M. 2006. The Fanconi anemia pathway promotes homologous recombination repair in DT40 cell line. *Subcell Biochem* 40: 295-311
Tan C, Tomkins J. 2015. Information Processing Differences Between Archaea and Eukaraya—Implications for Homologs and the Myth of Eukaryogenesis. *Answers Research Journal* 8: 121-41
Tang SL, Nuttall S, Dyall-Smith M. 2004. Haloviruses HF1 and HF2: evidence for a recent and large recombination event. *J Bacteriol* 186: 2810-7
Tang SL, Nuttall S, Ngui K, Fisher C, Lopez P, Dyall-Smith M. 2002. HF2: a double-stranded DNA tailed haloarchaeal virus with a mosaic genome. *Mol Microbiol* 44: 283-96
Taniguchi T, Garcia-Higuera I, Andreassen PR, Gregory RC, Grompe M, D'Andrea AD. 2002. S-phase-specific interaction of the Fanconi anemia protein, FANCD2, with BRCA1 and RAD51. *Blood* 100: 2414-20
Tapias A, Auriol J, Forget D, Enzlin JH, Scharer OD, et al. 2004. Ordered conformational changes in damaged DNA induced by nucleotide excision repair factors. *J Biol Chem* 279: 19074-83
Taylor MR, Spirek M, Chaurasiya KR, Ward JD, Carzaniga R, et al. 2015. Rad51 Paralogs Remodel Pre-synaptic Rad51 Filaments to Stimulate Homologous Recombination. *Cell* 162: 271-86
Teng G, Papavasiliou FN. 2007. Immunoglobulin somatic hypermutation. *Annu Rev Genet* 41: 107-20
Terpe K. 2003. Overview of tag protein fusions: from molecular and biochemical fundamentals to commercial systems. *Appl Microbiol Biotechnol* 60: 523-33
Thangavel S, Berti M, Levikova M, Pinto C, Gomathinayagam S, et al. 2015. DNA2 drives processing and restart of reversed replication forks in human cells. *J Cell Biol* 208: 545-62
Thoden JB, Kappock TJ, Stubbe J, Holden HM. 1999. Three-dimensional structure of N5-carboxyaminoimidazole ribonucleotide synthetase: a member of the ATP grasp protein superfamily. *Biochemistry* 38: 15480-92
Tomasz M. 1995. Mitomycin C: small, fast and deadly (but very selective). *Chem Biol* 2: 575-9
Tomasz M, Lipman R, Chowdary D, Pawlak J, Verdine GL, Nakanishi K. 1987. Isolation and structure of a covalent cross-link adduct between mitomycin C and DNA.

Science 235: 1204-8
Toulme F, Le Cam E, Teyssier C, Delain E, Sautiere P, et al. 1995. Conformational changes of DNA minicircles upon the binding of the archaebacterial histone-like protein MC1. *J Biol Chem* 270: 6286-91
Turner J, Hingorani MM, Kelman Z, O'Donnell M. 1999. The internal workings of a DNA polymerase clamp-loading machine. *EMBO J* 18: 771-83
Tuteja N, Tuteja R. 2004. Unraveling DNA helicases - Motif, structure, mechanism and function. *Eur J Biochem* 271: 1849-63
Veaute X, Jeusset J, Soustelle C, Kowalczykowski SC, Le Cam E, Fabre F. 2003. The Srs2 helicase prevents recombination by disrupting Rad51 nucleoprotein filaments. *Nature* 423: 309-12
Velankar SS, Soultanas P, Dillingham MS, Subramanya HS, Wigley DB. 1999. Crystal structures of complexes of PcrA DNA helicase with a DNA substrate indicate an inchworm mechanism. *Cell* 97: 75-84
Venkitaraman AR. 2002. Cancer susceptibility and the functions of BRCA1 and BRCA2. *Cell* 108: 171-82
Wackett L. 2012. Industrial applications of microbial salt-tolerant enzymes. *Microbial Biotechnology* 5: 668-69
Wadsworth RI, White MF. 2001. Identification and properties of the crenarchaeal single-stranded DNA binding protein from Sulfolobus solfataricus. *Nucleic Acids Res* 29: 914-20
Walker JR, Corpina RA, Goldberg J. 2001. Structure of the Ku heterodimer bound to DNA and its implications for double-strand break repair. *Nature* 412: 607-14
Wang X, Lutkenhaus J. 1996. FtsZ ring: the eubacterial division apparatus conserved in archaebacteria. *Mol Microbiol* 21: 313-9
Ward JD, Muzzini DM, Petalcorin MI, Martinez-Perez E, Martin JS, et al. 2010. Overlapping mechanisms promote postsynaptic RAD-51 filament disassembly during meiotic double-strand break repair. *Mol Cell* 37: 259-72
Wardell K. 2013. *Regulation of Homologous Recombination in the Archaeon Haloferax volcanii*. University of Nottingham
Weiner AM, Deininger PL, Efstratiadis A. 1986. Nonviral retroposons: genes, pseudogenes, and transposable elements generated by the reverse flow of genetic information. *Annu Rev Biochem* 55: 631-61
Weller GR, Kysela B, Roy R, Tonkin LM, Scanlan E, et al. 2002. Identification of a DNA nonhomologous end-joining complex in bacteria. *Science* 297: 1686-9
Wendoloski D, Ferrer C, Dyall-Smith ML. 2001. A new simvastatin (mevinolin)-resistance marker from Haloarcula hispanica and a new Haloferax volcanii strain cured of plasmid pHV2. *Microbiology* 147: 959-64
West SC. 1997. Processing of recombination intermediates by the RuvABC proteins. *Annu Rev Genet* 31: 213-44
Whitby MC. 2010. The FANCM family of DNA helicases/translocases. *DNA Repair (Amst)* 9: 224-36
White MF. 2003. Archaeal DNA repair: paradigms and puzzles. *Biochem Soc Trans* 31: 690-3
White MF. 2011. Homologous recombination in the archaea: the means justify the ends. *Biochem Soc Trans* 39: 15-9

White MF, Bell SD. 2002. Holding it together: chromatin in the Archaea. *Trends Genet* 18: 621-6

Wickham-Smith C. 2015. *Identifying the protein interactions of Mre11-Rad50 in Haloferax volcanii during double-strand break repair*. University of Nottingham

Wilson DM, 3rd, Bohr VA. 2007. The mechanics of base excision repair, and its relationship to aging and disease. *DNA Repair (Amst)* 6: 544-59

Wilson MA, Kwon Y, Xu Y, Chung WH, Chi P, et al. 2013. Pif1 helicase and Poldelta promote recombination-coupled DNA synthesis via bubble migration. *Nature* 502: 393-6

Winker S, Woese CR. 1991. A definition of the domains Archaea, Bacteria and Eucarya in terms of small subunit ribosomal RNA characteristics. *Syst Appl Microbiol* 14: 305-10

Woese CR, Fox GE. 1977. Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proc Natl Acad Sci U S A* 74: 5088-90

Woese CR, Kandler O, Wheelis ML. 1990. Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proc Natl Acad Sci U S A* 87: 4576-9

Wolf YI, Rogozin IB, Kondrashov AS, Koonin EV. 2001. Genome alignment, evolution of prokaryotic genome organization, and prediction of gene function using genomic context. *Genome Res* 11: 356-72

Woodman IL, Bolt EL. 2009. Molecular biology of Hel308 helicase in archaea. *Biochem Soc Trans* 37: 74-8

Woodman IL, Bolt EL. 2011. Winged helix domains with unknown function in Hel308 and related helicases. *Biochem Soc Trans* 39: 140-4

Woodman IL, Brammer K, Bolt EL. 2011. Physical interaction between archaeal DNA repair helicase Hel308 and Replication Protein A (RPA). *DNA Repair (Amst)* 10: 306-13

Woodman IL, Briggs GS, Bolt EL. 2007. Archaeal Hel308 domain V couples DNA binding to ATP hydrolysis and positions DNA for unwinding over the helicase ratchet. *J Mol Biol* 374: 1139-44

Woods WG, Dyall-Smith ML. 1997. Construction and analysis of a recombination-deficient (radA) mutant of Haloferax volcanii. *Mol Microbiol* 23: 791-7

Wu L, Hickson ID. 2003. The Bloom's syndrome helicase suppresses crossing over during homologous recombination. *Nature* 426: 870-4

Wu Y, Lee SH, Williamson EA, Reinert BL, Cho JH, et al. 2015. EEPD1 Rescues Stressed Replication Forks and Maintains Genome Stability by Promoting End Resection and Homologous Recombination Repair. *PLoS Genet* 11: e1005675

Wu Z, Liu J, Yang H, Xiang H. 2014. DNA replication origins in archaea. *Front Microbiol* 5: 179

Zhang C, Tian B, Li S, Ao X, Dalgaard K, et al. 2013. Genetic manipulation in Sulfolobus islandicus and functional analysis of DNA repair genes. *Biochem Soc Trans* 41: 405-10

Zhang J. 2003. Evolution by gene duplication: an update. *RENDS in Ecology and Evolution* 18

Zhao A, Gray FC, MacNeill SA. 2006. ATP- and NAD+-dependent DNA ligases share an essential function in the halophilic archaeon Haloferax volcanii. *Mol Microbiol* 59:

743-52
Zhao X, Li G, Liang S. 2013. Several Affinity Tags Commonly Used in Chromatographic Purification. *Journal of Analytical Methods in Chemistry* 2013: 8
Zheng L, Shen B. 2011. Okazaki fragment maturation: nucleases take centre stage. *J Mol Cell Biol* 3: 23-30
Zou Y, Van Houten B. 1999. Strand opening by the UvrA(2)B complex allows dynamic recognition of DNA damage. *EMBO J* 18: 4889-901