Genetic and Biochemical Analysis of the Hel308 Helicase in the Archaeon *Haloferax volcanii*

Rebecca Janet Lever, MSci (Hons)

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

September 2019

Contents

Abstract v
Acknowledgements vi
Abbreviations vii

Chapter 1: Introduction **1**
1.1. Archaea 1
1.1.1. The discovery of Archaea 1
1.1.2. The Characteristics of Archaea 4
1.1.3. Halophiles 5
1.2. *Haloferax volcanii* as a model organism 5
1.2.1. *H. volcanii* Genetic toolbox 7
1.2.2. *H. volcanii* Biochemical toolbox 10
1.3. DNA replication 10
1.3.1. Initiation 10
1.3.2. Elongation 14
1.3.3. Termination 16
1.3.4. Lesions in the DNA template 17
1.4. DNA repair 17
1.4.1. DNA Damage 17
1.4.1.1. Endogenous 18
1.4.1.2. Exogenous 19
1.4.2. DNA Repair Pathways 21
1.4.2.1. Direct Repair 21
1.4.2.2. Excision Repair 21
1.4.2.3. Double Strand Break Repair 28
1.5. Homologous recombination 32
1.5.1. Pre-synapsis 34
1.5.2. Synapsis 37
1.5.3. Post-synapsis 38
1.5.3.1. Synthesis-dependent strand invasion 38
1.5.3.2. Double-strand break repair 40
1.5.4. Regulation of Homologous Recombination 45
1.6. Replication Fork Restart 46
1.6.1. Stalled Fork Processing 47
1.6.2. Replisome Reassembly 50
1.7. CRISPR-Cas adaptive immunity 51
1.8. DNA Helicases 53
1.8.1. Helicase Structure 54
1.8.2. Mechanism of Helicase Unwinding 55
1.8.3. Hel308 58

1.8.3.1. Structure of Hel308 58
1.8.3.2. Biochemical analysis of Hel308 60
1.8.3.3. Hel308 protein:protein Interactions 61
1.8.3.4. Genetic analysis of Hel308 62

Chapter 2: Materials and Methods 65
2.1. Materials 65
2.1.1. Strains 65
2.1.2. Plasmids 67
2.1.3. Oligonucleotides 71
2.1.4. Chemicals and Enzymes 74
2.1.4.1. Media 74
2.1.4.2. Other Solutions 77
2.2. Methods 77
2.2.1. *E. coli* Microbiology 77
2.2.2. *H. volcanii* Microbiology 78
2.2.3. DNA Extraction from Cells 80
2.2.4. Nucleic Acid Manipulation 81
2.2.5. Genetic Manipulation of *Haloferax volcanii* 87
2.2.6. Genotype Screening 92
2.2.7. Phenotyping of *Haloferax volcanii* 93
2.2.8. Protein Purification 98

Chapter 3: Separation of function in Hel308 103
3.1. Background 103
3.2. Aims and Objectives 113
3.3. Materials 113
3.3.1. Strains 113
3.3.2. Plasmids 114
3.3.3. Oligonucleotides 122
3.4. Results 124
3.4.1. Hel308 as an anti-recombinase 124
3.4.1.1. Development of an improved *radA in trans* expression construct for deletion of *radA* in a strain deleted for *radB* 124
3.4.1.2. Generation of *hel308, radA,* and *radB* deleted strains 126
3.4.1.3. Characterisation of *hel308* and *radA* deleted strains 130
3.4.2. Separation of function Mutations in Hel308 133
3.4.2.1. Generation of ATPase-null and hyper-recombinant mutant *hel308* strains 133

3.4.2.2. Characterisation of ATPase-null and hyper-recombinant mutant *hel308* strains 148
3.5. Discussion 155
3.6. Future Perspectives 161
3.7. Conclusion 163

Chapter 4: Biochemical Analysis of Hel308 **165**
4.1. Background 165
4.2. Aims and Objectives 174
4.3. Materials 174
4.3.1. Strains 174
4.3.2. Plasmids 175
4.3.3. Oligonucleotides 182
4.4. Results 183
4.4.1. Development of Chromosomally-Tagged Expression strains 183
4.4.2. Characterisation of Chromosomally-Tagged Expression strains 200
4.4.3. Protein purification of natively expressed proteins 203
4.4.4. Protein:Protein Interactions *in vivo* 204
4.4.5. Development of episomal overexpression constructs 216
4.4.6. Protein overexpression and purification 223
4.5. Discussion 227
4.5.1. Hel308 interacting partners *in vivo* 227
4.5.2. Development of Methods 234
4.6. Future Perspectives 236
4.7. Conclusion 236

Chapter 5: Nucleases and Helicases **237**
5.1. Background 237
5.2. Aims and Objectives 240
5.3. Materials 240
5.3.1. Strains 240
5.3.2. Plasmids 242
5.3.3. Oligonucleotides 247
5.4. Results 247
5.4.1. Generation of *hel308, hef, hjc,* and *pinA* deleted strains 247
5.4.1.1. Plasmid construction 248
5.4.1.2. Generation of *pinA* deleted strains 251
5.4.1.3. Generation of Δ*pinA* in combination with *hel308, hef,* and *hjc* deletions 253

5.4.2. Characterisation of *hel308, hef, hjc,* and *pinA* deleted strains 256
5.4.3. Generation of *hel308* mutant and *cas3* deleted strains 261
5.4.4. Characterisation of *hel308* mutant and *cas3* deleted strains 268
5.5. Discussion 271
5.6. Future Perspectives 272
5.7. Conclusion 274

Chapter 6: Genetic analysis of the *recJ* genes 275
6.1. Background 275
6.1.1. The CMG complex and DNA replication 275
6.1.2. *H. volcanii* RecJ 275
6.2. Aims and Objectives 277
6.3. Materials 277
6.3.1. Strains 277
6.3.2. Plasmids 278
6.3.3. Oligonucleotides 279
6.4. Results 282
6.4.1. Phylogenetic analysis of Archaeal *recJ* 282
6.4.2. Generation of *recJ* and *ginS* Deletion Strains 288
6.4.3. Characterisation of *recJ* Deletion Strains 312
6.4.4. Development of Chromosomally-Tagged Expression Constructs 314
6.5. Discussion 322
6.6. Future Perspectives 325
6.7. Conclusion 327

Chapter 7: Conclusion and Future Perspectives 329

References 335
Appendix 1 363
Appendix 2 375

Abstract

Hel308 is a monomeric 3' to 5' DNA helicase conserved throughout archaea and metazoans, but is absent from bacteria and fungi. Hel308 acts analogously to RecQ-family helicases, which serve as guardians of genome stability. Hel308 has been implicated in a variety of overlapping DNA repair pathways, including homologous recombination. However, the exact role of Hel308 remains elusive. *Haloferax volcanii* was used as a model archaeon to study cellular role of Hel308.

Deletion mutants of *hel308*, or point mutants that affect the ATP-dependent helicase activity, are slow-growing and are sensitive to DNA cross-linking agents, which are potent blocks to DNA replication. By contrast, point mutants that exhibit hyper-recombinant phenotypes are not deficient in DNA repair or normal growth. These hyper-recombinant point mutations identify a separation of function of Hel308, where the ability to bind D-loop intermediates of recombination is affected, but the helicase activity involved in DNA repair is not.

Mutants deleted for *radB*, a recombination mediator, exhibit a decrease in recombination to approximately 5% of that in wild type cells. However, a double deletion of *hel308* with *radB* restores, in part, homologous recombination. In this study, strains deleted for *hel308* exhibited a 5-fold increase in recombination compared with wild-type, which suggests an anti-recombinase function for Hel308 and fits with its role as a guardian of genome stability.

Strains deleted for *radA* usually have a severe growth defect and are very sensitive to both UV irradiation and MMC treatment. Strains deleted for *hel308* exhibit similar phenotypes, but not to the same severity as the Δ*radA* strains. Most interestingly, the strain deleted for both *hel308* and *radA* generated in this study was healthier than either single deletion. These data suggest that in the absence of Hel308-mediated repair and homologous recombination, a third distinct DNA-repair pathway is activated, which is ordinarily supressed by both Hel308-mediated repair and homologous recombination. An alternative explanation is that in deleting both the *hel308* and *radA* genes, a suppressor mutation has arisen. Only a singular clone with the Δ*hel308*Δ*radA* phenotype was generated, which supports the suppressor-mutation theory. However, a suppressor mutation or mutations would need the ability to bypass two distinct repair pathways in order to restore wildtype-like fitness, which seems unlikely.

Acknowledgements

First and foremost, I would like to thank Thorsten for the opportunity to work on this project and for his confidence in me as a student. I would also like to thank him for his guidance, support, and anecdotes over the four years, there was never a dull moment.

Secondly, I would like to thank past and present members of the lab for their help, friendship, and endless supply of cake through the years. I would especially like to thank Becca for passing the 'Hel308 baton' on to me and for answering my endless questions in the beginning.

I would also like to thank my amazing family for their support and encouragement, particularly my Dad who (temporarily) gave up his study so that I may write my thesis in there, and my Mum who was the best soundboard when I couldn't find the words I needed! I thank my sisters, Gaby and Ellie, for providing comic relief when I was taking the world too seriously and I thank my Nan for always being so proud of me. Thank you to my fiancé Harry, who kept me well stocked with chocolate throughout as well as being my biggest cheerleader. Finally, I would like to thank my dogs; Belle, Daisy, and Lola for keeping me company during the many hours of writing and providing fur-apy when I needed it – they give the best cuddles!

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
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
SW	Salt water
TAE	Tris/Acetic acid/EDTA
TBE	Tris/Borate/EDT A
TE	Tris/EDT A
TEMED	Tetramethylethylenediamine
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 The discovery of Archaea

Until the mid-1970s, the widely accepted two domain tree of life was constructed based upon physiological, morphological, and pathological similarities. On one branch of the tree, prokaryotic organisms were grouped together, this encompassed bacteria and archaea though the two were yet to be formally distinguished. The second branch was composed of eukaryotic organisms. In 1977 Carl Woese and George Fox saw to revolutionise this logical two domain tree via the use of nucleotide sequence analysis (Woese and Fox, 1977). Focussing on the small subunit ribosomal RNAs (16S rRNA), which is common to all living organisms and has a high sequence conservation, Woese and Fox proposed a three domain tree comprising bacteria and eukaryotes as before, and the new "third domain of life"; archaebacteria. Analysis via this sophisticated phylogenetic approach later revealed that archaebacteria, despite being prokaryotic, were actually more closely related to eukaryotes than bacteria and thus were renamed archaea (Woese, Kandler and Wheelis, 1990) (Figure 1.1).

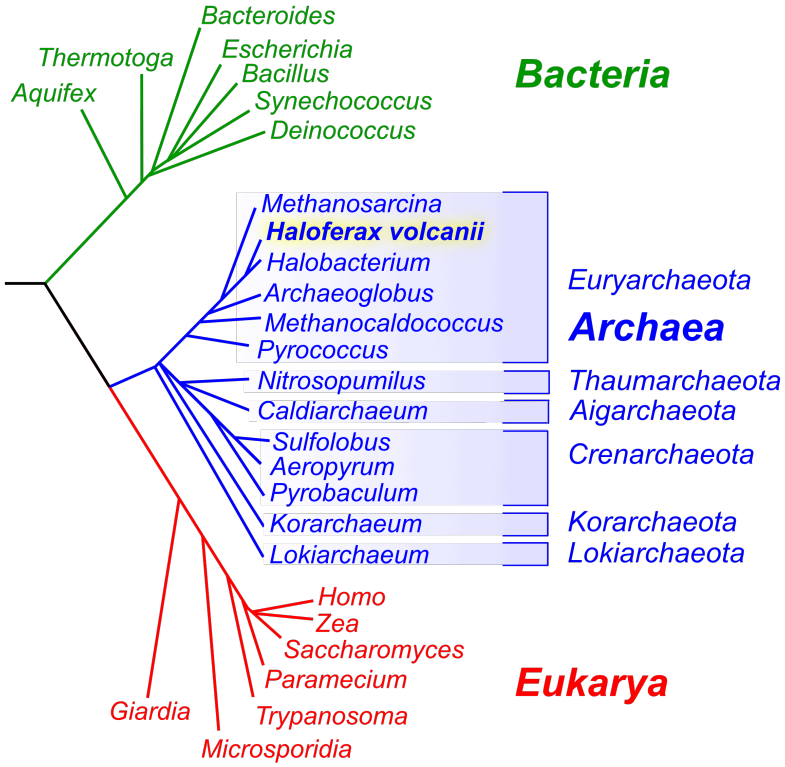


Figure 1.1: **Phylogenetic tree constructed via 16S rRNA sequencing**. The three domains of life consist of Bacteria, Eukarya, and Archaea, only select representative phyla and genera are shown. Figure adapted from (Allers and Mevarech, 2005).

It was proposed that archaea and eukaryotes are derived from a common ancestor, subsequent to the split from bacteria, and so were thought as sister groups (Pace, 1997). However, since Woese's original proposal, the phylogenetic tree has progressed considerably having seen the addition, loss, and reclassification of several phyla. Deeper analysis of archaeal rRNA nucleotide sequences suggested two main phyla – Crenarchaeota and Euryarchaeota (Winker and Woese, 1991) – though this has since been revised with the discovery of an additional four phyla. A superphylum within the archaea, known as TACK, has been recognised as a cluster of four main phyla: Thaumarchaeota, originally thought as mesophilic Crenarchaeota (Brochier-Armanet *et al.*, 2008; Pester, Schleper and Wagner, 2011); Aigarchaeota (Nunoura *et al.*, 2011); Crenarchaeota; and Korarchaeota, an ancient division believed to have deviated prior to the Crenarchaeota and Euryarchaeota separation (Auchtung, Takacs-Vesbach and Cavanaugh, 2006; Elkins *et al.*, 2008). Other phyla include the Nanoarchaeota renowned for having very few genes and a small cell size (Huber et al. 2003; Munsen-McGee et al. 2015), the Bathyarchaeota, and the Geoarchaeota. Little is known about the latter two phyla (Koonin, 2015). Nanoarchaeota belong to the DPANN superphylum, which was first proposed in 2013 and named for the first five groups discovered; Diapherotrites, Parvarchaeota, Aenigmarchaeota, Nanoarchaeota and Nanohaloarchaeota (Rinke *et al.*, 2013). Later additions to the superphylum include Woesearchaeota, Pacearchaeota and Altiarchaeota (Castelle *et al.*, 2015; Spang, Caceres and Ettema, 2017). The most current, though not wholly accepted, phylogenetic tree displays eukaryotes not as a sister domain to archaea but instead as direct descendants from the TACK superphylum (Koonin and Yutin, 2014) (Figure 1.2).

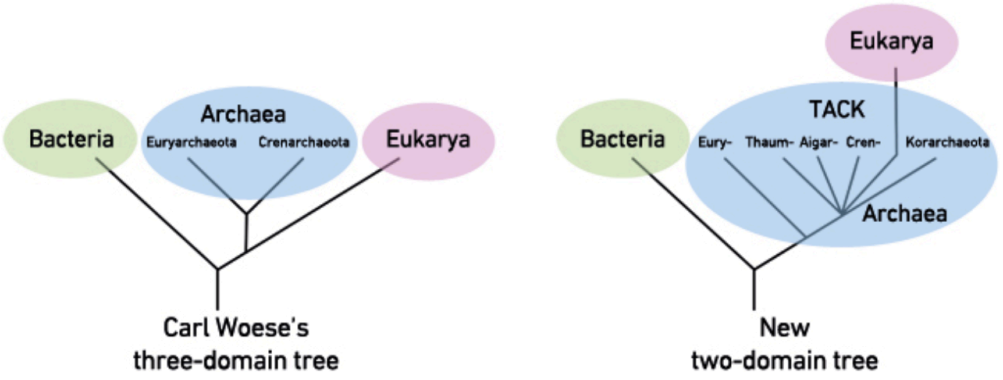


Figure 1.2: **Phylogenetic relationship between Eukarya and Archaea**. Woese's classical three-domain tree of life (left) where Archaea and Eukarya are sister lineages, compared with the new two-domain tree (right), placing Eukarya as a secondary domain that arose from within the Archaea. Figure from (Marriott and Allers, 2016).

Evidence to support this hypothesis includes; Sampylation, an ubiquitin-like conjugation system closely related to the eukaryotic ubiquitin proteasome system (Maupin-Furlow, 2013); and RNA interference, a eukaryotic system involved in the regulation of gene expression and antivirus defence which has components derived from both bacteria and the Euryarchaeota of archaea (Makarova *et al.*, 2009). Additionally, within the *Sulfolobales* of the Crenarchaeota and *Nitrosopumilus maritimus* of the Thaumarchaeota, a distinct cell division system homologous to the ESCRT-III membrane remodelling complex has been characterised (Lindgren *et al.*, 2008; Samson *et al.*, 2008; Makarova *et al.*, 2010; Martens-Habbena *et al.*, 2011). Ubiquitin signalling, RNA interference, actin and tubulin based cytoskeletons are all examples of typical eukaryotic components that are also seen dispersed across the archaea, this provides evidence for and supports the hypothesis that eukaryotes evolved from archaea (Figure 1.3). More recently the discovery of the Lokiarchaeota, named for the hydrothermal vent in the Arctic Ocean near Loki's Castle where the species was first found, are believed to be the closest related archaea to eukaryotes. This supports the two domain tree comprising bacteria and archaea, with eukaryotes as direct descendants of archaea (Williams *et al.*, 2013; Embley and Williams, 2015; Spang *et al.*, 2015). Furthermore, biological methane, most likely produced by methanogenic euryarchaea, has been found in rocks that are 3.4 billion years old. This predates the earliest known eukaryotic microfossils, which are around 1.8 billion years old, providing additional evidence that archaea are an ancient form of life that predate eukaryotes (Jaakkola *et al.*, 2016).

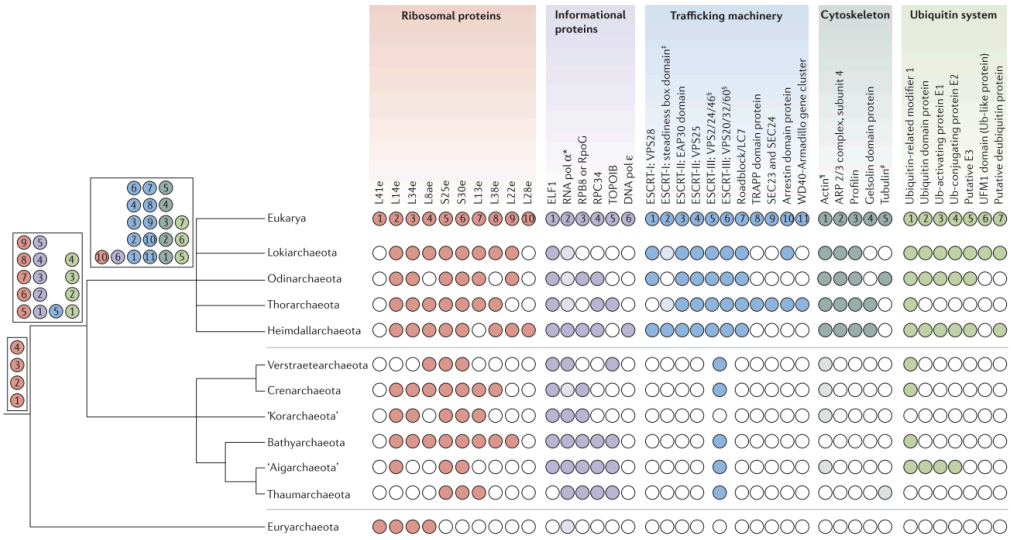


Figure 1.3: **Origins of eukaryotic signature proteins present in the first eukaryotic common ancestor (FECA).** This presence of eukaryotic signature proteins (ESPs) homologues in various archaeal lineages (filled circles), as well as their inferred time of emergence along the schematic tree of Archaea. The origin of each ESP is indicated on a schematic tree of life (left-hand side). Figure adapted from (Eme *et al.*, 2017).

Despite this evidence, the debate between a two-domain vs a three-domain tree continues (Zhou *et al.*, 2018). Several large scale phylogenetic studies exploring the relationship between the archaea and eukarya have been conducted and although progress has been made, no individual study is definitive. Several studies support the two-domain hypothesis of the Eukarya being a part of an archaeal lineage (Rivera and Lake, 2004; Pisani, Cotton and McInerney, 2007; Embley *et al.*, 2008; Foster, Cox and Martin Embley, 2009; Williams *et al.*, 2013; Embley and Williams, 2015; Spang *et al.*, 2015). There are recent studies however, that support and defend the theory that the Archaea and Eukarya are independent monophyletic lineages gives rise to a three-domain tree (Harris *et al.*, 2003; Ciccarelli, 2006; Koonin and Yutin, 2014; Van Der Gulik, Hoff and Speijer, 2017).

1.1.2 The Characteristics of Archaea

Archaea are highly diverse organisms found throughout many different environments including very extreme and harsh habitats such as hypersaline brines, hydrothermal vents, and acidic/alkaline waters. For this reason, archaea are often labelled “extremophiles”. Though it is true that the majority of extremophiles tend to be archaea, not all archaea are extremophiles. Representing upwards of 20% of all microbial cells on Earth, archaea are widely found in moderate habitats such as fresh water sediments, soil, and oceans (DeLong and Pace, 2001; Chaban, Ng and Jarrell, 2006).

Though archaea possess features in common with both bacteria and eukaryotes, archaea are morphologically more similar to bacteria. However, when considering DNA replication, transcription and translation, the archaeal mechanisms that carry out these fundamental processes resemble a less-complex version of the equivalent eukaryotic mechanisms. (Grabowski & Kelman 2003, Barry & Bell 2006).

As one might expect, archaea also possess unique qualities. One such quality is the tendency to have a higher gene density than bacterial or eukaryotic genomes (Koonin and Wolf, 2008). Similar to bacteria and some eukaryotes, archaeal genomes are composed of circular chromosomes and arrange genes in operons (Olsen and Woese, 1997). Archaea can possess a single origin of replication, a typical characteristic of bacteria, or multiple origins of replication similar to eukaryotes. Some archaea also possess the ability to initiate DNA replication in the complete absence of replication origins, this will be discussed in further detail in 1.3 DNA Replication.

1.1.3 Halophiles

Halophilic archaea, or halophiles, are found in hypersaline conditions such as the Dead Sea. To survive in such high salt conditions (1 - 3.5 M) either a 'salt out' or a 'salt in' mechanism must be employed by the organism. 'Salt out' is the choice mechanism adopted by the various halophilic bacteria and eukaryotes present in these conditions. Here, an osmotic balance is maintained via two techniques. Firstly, the pumping out of salts from the cell and secondly, cytosolic packing with organic solutes such as glycerol (Christian and Waltho, 1962; Oren, 1999, 2008). Halophilic archaea and a few halophilic bacteria however, prefer the 'salt in' mechanism whereby an osmotic balance is maintained by the accumulation of high levels of salt in the cytosol (Oren *et al.*, 2002). The proteins within these organisms have over time adapted to function in this high salt to low water ratio.

1.2 *Haloferax volcanii* as a model organism

Haloferax volcanii was originally isolated from the Dead Sea in 1975 (Mullakhanbhai and Larsen, 1975; Arahal *et al.*, 1996). *H. volcanii's* discovery, alongside many other halophilic organisms, was considered ground breaking as it opposed the previous conception that no living entity could survive in the Dead Sea, hence the name. *H. volcanii* belongs to Euryarchaeota phylum of archaea. The organism is coccoid, appears red in colour due to the presence of carotenoids, and is surrounded by a glycoprotein S layer (Hartman *et al.*, 2010). *H. volcanii* is easily cultured and possesses a large genetic toolbox, thus is an ideal candidate to study archaeal genetics. The choice culture conditions of *H. volcanii* is an aerobic atmosphere, at an optimum growth temperature of 45°C, and in the presence of 1.7 – 2.5 M sodium chloride. *H. volcanii* generally has a generation time of 2-3 hours in liquid media.

The *H. volcanii* genome is highly polyploid with a genome copy number of up to 20 copies per cell, additionally the *H. volcanii* genome is particularly GC rich with a %GC content of approximately 65% (Breuert *et al.*, 2006). The complete 4.2 Mb genome of wild-type *H. volcanii* (DS2 isolate) is composed of a 2.85 Mb main chromosome and three mini chromosomes: pHV1, pHV3, and pHV4; at 86 kb, 442 kb, and 690 kb respectively (Charlebois *et al.*, 1991). Also present in the wildtype isolate DS2, is the 6 kb plasmid pHV2, which was intentionally lost from the laboratory strain; this consequently triggered the integration of the pHV4 mini chromosome onto the main chromosome (Wendoloski, Ferrer and Dyall-Smith, 2001; Hawkins *et al.*, 2013). A full genome sequence is available for *H. volcanii* (Hartman *et al.*, 2010).

Concurrent rounds of replication occur within the *H. volcanii* cell resulting in an origin:terminus copy number maximum: minimum ratio of >2:1. These DNA replication profiles, generated via deep sequencing, are evidence against the presence of a defined cell cycle, including a defined S-phase, within *H. volcanii* (Hawkins *et al.*, 2013) (Figure 1.4).

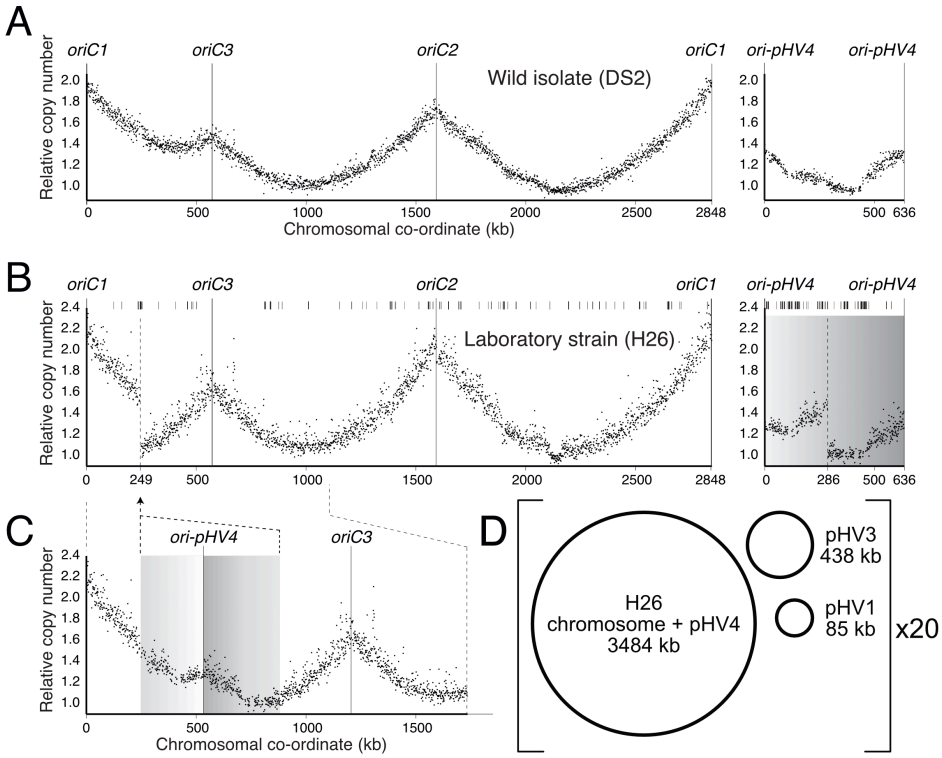


Figure 1.4 **Replication profiles for *H. volcanii* wild isolate and laboratory strain.** A) Relative copy number plotted against chromosomal coordinate for the main chromosome and pHV4 of wild isolate DS2. Circular chromosomes are shown linearized at position 0, vertical lines mark replication origins. B) Relative copy number plotted against chromosomal coordinate for the main chromosome and pHV4 of H26 mapped to the DS2 genome. pHV4 shading reflects chromosomal coordinate. C) Reconstructed assembly of the main chromosome with pHV4 integrated at approximately 249 kb. Grey shading from B indicates the orientation of pHV4 integration. D) *H. volcanii* origin:terminus copy number ~20. Figure adapted from (Hawkins *et al.*, 2013).

1.2.1 *H. volcanii* Genetic toolbox

Selectable markers

Several gene deletions from the *H. volcanii* laboratory strain have been made to allow the use of auxotrophic selectable markers. Though antibiotic selectable markers are available in *H. volcanii,* the auxotrophic selection methods were preferred as these bypass the problem of homologous recombination. Antibiotic selectable markers were originally isolated as mutant alleles of the essential genes *gyrB* and *hmgA*, and thus demonstrate near-complete homology. Consequently, the host strain could acquire antibiotic resistance via homologous recombination between the chromosomal alleles and the plasmid borne resistance markers. Mutated *gyrB* gives resistance against novobiocin, an inhibitor of DNA gyrase. Likewise, the mutated *hmgA* gives resistance against mevinolin which is an HMG-CoA reductase inhibitor (Lam and Doolittle, 1989; Holmes, Nuttall and Dyall-smith, 1991).

The auxotrophic selectable markers ordinarily used are *pyrE2, trypA. leuB*, and *hdrB*, which encode orotate phosphoribosyl transferase, tryptophan synthase, 3-isopropylmalate dehydrogenase, and dihydrofolate reductase respectively. These proteins are involved in the corresponding essential pathways: uracil biosynthesis (Bitan-banin *et al.*, 2003), tryptophan biosynthesis, leucine biosynthesis (Allers *et al.*, 2004), and thymidine biosynthesis (Ortenberg, Rozenblatt-Rosen and Mevarech, 2000).

Reporter genes

The ß-galactosidase gene is the first of two reporter genes widely used in *H. volcanii*. The gene is often employed in growth competition assays and recombination assays (Delmas *et al.*, 2009), as it enables the detection of blue colonies upon X-gal treatment of the strains (Holmes and Dyall-Smith, 2000). The second widely used reporter gene is a modified green fluorescent protein (GFP), comprising three amino acid substitutions enabling the adaptation of the protein to the high salt conditions required for *H. volcanii* (Crameri *et al.*, 1996; Reuter and Maupin-furlow, 2004).

Transformation

By first removing the S-layer via EDTA treatment, *H. volcanii* can be easily transformed with plasmid DNA (*dam*-). *H. volcanii* strains deleted for *mrr* are unable to recognise and degrade *dam* methylated DNA and therefore, these strains can be directly transformed with *dam*+ plasmid DNA. It has been found that the addition of polyethylene glycol facilitates the uptake of DNA (Cline,

Schalkwyk and Doolittle, 1989). Though the use of linear DNA for transformation is not unheard of, the transformation efficiency using linear DNA is approximately 100-fold less than that using plasmid DNA. This decrease in efficiency is a consequence of the Mre11-Rad50 restraint on homologous recombination in *H. volcanii.* Mre11-Rad50 is crucial in the repair of double strand breaks and appears to function in an anti-recombinase fashion. Recognising linear DNA as damaged in this way, Mre11-Rad50 actively prevents the recombination of linear DNA onto the chromosome. This was not expected based on studies carried out in other organisms including *Saccharomyces cerevisiae, Deinococcus radiodurans,* and *Bacillus subtilis* (Delmas *et al.*, 2009).

Gene Deletion/Replacement

A gene deletion and knock out method has been developed in *H. volcanii* by utilising the *pyrE2* marker. Strains deleted for *pyrE2* (Δ*pyrE2*) are transformed with a plasmid consisting of a deletion construct for the desired gene, including typically a *trpA* marker in place of the gene, and the *pyrE2* marker. Successful transformants, known as “pop-ins”, will grow on media lacking uracil. By then removing the uracil selection, the pop-out step is able to occur. Successful pop-outs lose the *pyrE2* marker and so are selected for on Hv-Ca + 5-FOA media, since 5-fluoroorotic acid (5-FOA) is a fatal inhibitor for strains retaining the *pyrE2* marker. The inclusion of the *trpA* marker in place of the gene enables direct selection for the deletion mutants (Figure 1.5).

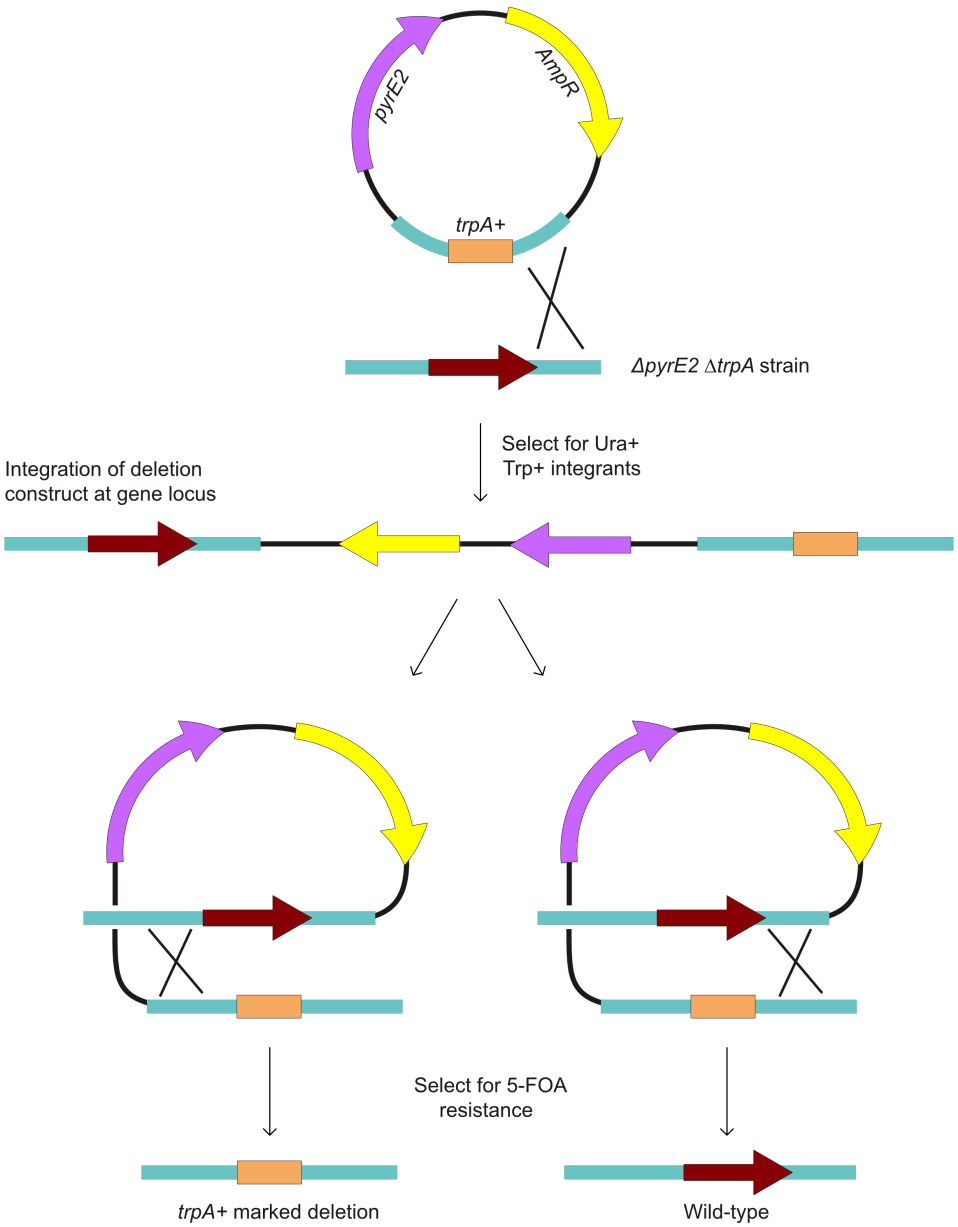


Figure 1.5: **Gene deletion by pop-in/pop-out.** Δ*pyrE2* strains are transformed with a *pyrE2*+ deletion construct. Pop-ins are selected for by their ability to grow on media lacking uracil. Relieving the uracil selection allows for pop-out to occur. Recombination between regions of homology can either be upstream or downstream of the gene. Pop-out is selected for by plating on 5-FOA. The resultant gene allele will either be a deletion or wild-type. Replcing the gene with the *trpA*+ marker in Δ*trpA* strains enables direct selection for deletion mutants.

Gene Expression

It is possible to overexpress genes in *H. volcanii* by the introduction of an episomal plasmid expressing the desired gene downstream of the tryptophan-inducible promotor *p.tnaA*. The *p.tnaA* promotor is tightly controlled in *H. volcanii* cells, having originated from the *H. volcanii* tryptophanase gene *tnaA*. The overexpression of halophilic proteins in a native environment has been achieved utilising this promotor (Allers *et al.*, 2010).

1.2.2 *H. volcanii* Biochemical toolbox

Since *H. volcanii* has adopted the salt-in method to survive in the organism's hypersaline habitat, native proteins have subsequently adapted to fold correctly in high salt conditions. If expression of these proteins is attempted in heterologous hosts such as *E. coli*, these halophilic proteins typically misfold or are insoluble in the low salt conditions. Thus this traditional biochemical method of expression in bacteria is not recommended.

Biochemical tools have been developed to enable the expression and purification of hexahistidine and strep tagged proteins in *H. volcanii*. For the histidine tag to be practical, several genetic modifications to *H. volcanii* were required to minimize co-purification of the recombinant protein with naturally histidine rich proteins. Host strains were firstly deleted for the Mrr restriction enzyme; this deletion enables direct transformation of *dam*+ protein expression plasmids into *H. volcanii*. The removal of the histidine rich linker from PitA and similarly the histidine rich C terminus from Cdc48d were also required (Stroud, Liddell and Allers, 2012).

1.3 DNA replication

Complete and accurate DNA replication is fundamental to the proliferation of life in all domains. The replication of DNA is tightly regulated and must occur before cell division in order to pass on genetic information to the next generation. There are three distinct stages of DNA replication seen in all three domains of life: initiation, elongation, and termination.

1.3.1 Initiation

Origin-dependent replication

DNA replication is normally initiated at defined regions on the genome known as "origins of replication". Bacteria have circular chromosomes and typically

undergo concurrent rounds of replication initiated at a single origin of replication called *oriC* (O'Donnell, Langston and Stillman, 2013). The *E.coli oriC* comprises a collection of 9 bp repeat elements, known as "DnaA boxes", within the DNA unwinding element (DUE) and spans approximately 250 bp in length. DnaA boxes are sequence-specific binding sites for the initiator protein DnaA, an AAA+ family ATPase, which forms a nucleoprotein filament at the *oriC* (Ozaki and Katayama, 2009). The loading of DnaA at the *oriC* results in the stretching, and consequently the melting, of the DNA duplex within the DUE. DnaB, the bacterial replicative helicase, is then able to load onto the resultant single stranded DNA with the assistance of the AAA+ family ATPase helicase loader, DnaC (Erzberger, Mott and Berger, 2006). DnaB is a hexameric helicase, which is broken by DnaC in order to load onto the ssDNA (Arias-Palomo *et al.*, 2013). DnaB unwinds the dsDNA using ATP hydrolysis, the resultant ssDNA is subsequently bound by the ssDNA-binding protein (SSB) forming a nucleofilament protein in order to prevent secondary structure formation and degradation by nucleases. The *E. coli* SSB is a homotetramer that binds ssDNA via an oligosaccharide binding fold (OB-fold) and forms protein:protein interactions via the C-terminal domain (Shereda *et al.*, 2008).

Eukaryotes have multiple origins of replication across a linear genome (O'Donnell, Langston and Stillman, 2013). Initiation at origins of replication is the responsibility of the origin recognition complex (ORC), a heteromeric protein complex composed of the proteins Orc1-6 (Bell and Dutta, 2002). ORC recruits the replication factor Cdc6, an AAA+ family ATPase, and Cdt1, which function as a helicase loader and in turn recruits the eukaryotic replicative helicase, MCM. MCM is a heterohexameric helicase composed of subunits MCM2-7. Together, ORC, Cdc6, Cdt1, and MCM form the pre-replicative complex and give license to the origin for replication. MCM forms part of the CMG complex along with Cdc45, and GINS (Makarova, Koonin and Kelman, 2012).The formation of the CMG complex converts the pre-replicative complex into the pre-initiation complex, which is tightly regulated by phosphorylation (Sclafani and Holzen, 2007). Similar to bacteria, the ssDNA generated in the unwinding of the CMG complex is bound by replication protein A (RPA), which is the eukaryotic equivalent to the bacterial SSB protein (Nguyen *et al.*, 2014).

Archaea have either one or multiple origins of replication. *Pyrococcus abyssi* contains a single origin of replication, in contrast to *H. volcanii* which has three origins of replication on the main chromosome known as *oriC-1, 2,* and *3.* The laboratory strain, H26, has an additional chromosomal origin of replication, *ori-pHV4*, as a result of the integration of the pHV4 mini chromosome onto the main chromosome (Kelman and Kelman, 2004; Hawkins *et al.*, 2013) (Figure 1.6).

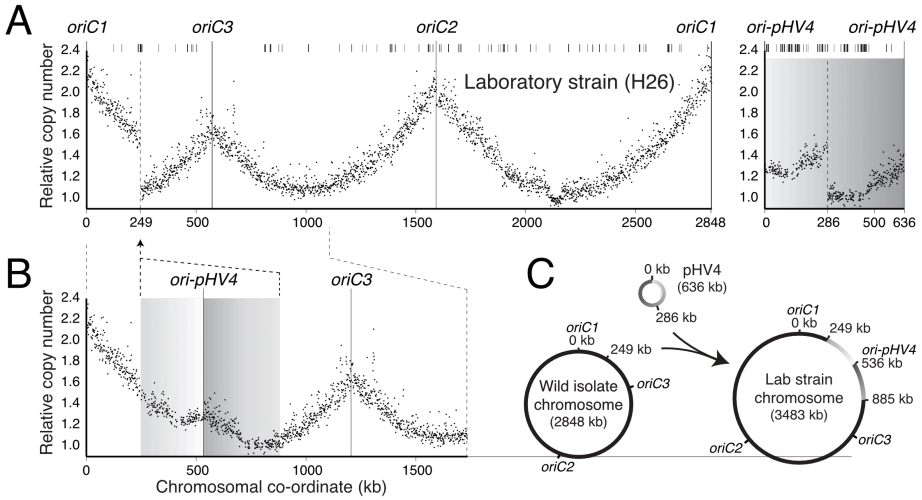


Figure 1.6 **Replication profiles for *H. volcanii* laboratory strain and integration of pHV4 into the main chromosome.** A) Relative copy number plotted against chromosomal coordinate for the main chromosome and pHV4 of H26 mapped to the DS2 genome. pHV4 shading reflects chromosomal coordinate. B) Reconstructed assembly of the main chromosome with pHV4 integrated at approximately 249 kb. Grey shading from A indicates the orientation of pHV4 integration. C) Integration of pHV4 into the main chromosome of H26. Figure adapted from (Hawkins *et al.*, 2013).

All archaeal species contain at least one homologue of the eukaryotic Orc1 or Cdc6 initiator proteins, though many archaeal species contain multiple Orc1/Cdc6 proteins (Barry and Bell, 2006). For example, *Pyrococcus abyssi* has only one Orc1/Cdc6 homologue, *Sulfolobus acidocaldarius* and *Sulfolobus solfataricus* each contain three Orc1/Cdc6 homologues, and *H.volcanii* has a total of sixteen *orc1/cdc6* genes (Nilsson *et al.*, 2004; Hartman *et al.*, 2010). Not all of the *H. volcanii orc1/cdc6* genes are located adjacent to an origin of replication and at least two are known not to be involved in DNA replication (Norais *et al.*, 2007). Deletion studies in *H. volcanii* have revealed that many of the *orc1/cdc6* genes have overlapping functions. The Orc1 proteins bind DNA at the origin recognition boxes (ORBs) and in turn recruit MCM. MCM is the replicative helicase that forms part of the CMG complex along with Cdc45, and GINS (Makarova, Koonin and Kelman, 2012). Archaeal replication resembles its eukaryotic counterpart, for which the CMG complex is essential. Whilst *H. volcanii* MCM and GINS homologues are known, a Cdc45 homologue has yet to be identified. Archaeal RecJ, known as GINS-associated nuclease (GAN) in several archaeal species, is proposed to be the Cdc45 homologue. However, few archaeal RecJ-MCM-GINS complexes have been isolated biochemically. Like eukaryotic MCM proteins and bacterial DnaB, archaeal MCM is an AAA+ family ATPase. In contrast to eukaryotes, archaeal genomes have only one *mcm* homologue, which forms a homohexamer with the exception of *Methanococcus maripaludis*. The *M. maripaludis* genome contains four *mcm* homologues, MCMA, MCMB, MCMC, and MCMD. Whilst only MCMA is essential,

deletion of the other three *mcm* genes result in proliferation defects suggesting that *M. maripaludis* MCM proteins form a heterohexamer (Walters and Chong, 2017). Furthermore, no archaeal homologues to bacterial or eukaryotic helicase loading factors have been found, suggesting that MCM interacts directly with the Orc proteins at the origins of replication (Barry and Bell, 2006). Similar to bacteria and eukaryotes, the ssDNA generated in the unwinding of the CMG complex is bound by a binding protein in order to prevent secondary structure formation and degradation by nucleases. Depending on the species, the archaeal binding protein is either eukaryotic RPA-like or bacterial SSB-like. For example, the single stranded binding protein from *Sulfolobus solfataricus* (Crenarchaea) has structural similarities to the bacterial SSB (Wadsworth, 2002). Conversely, the euryarchaea have a single stranded binding protein that more closely resembles the eukaryotic RPA. The formation of the RPA nucleoprotein filament varies between archaeal species. In *H. volcanii*, there are three RPA homologues, RPA1-3. RPA1 and RPA3 interact with the archaea specific RPA-associated proteins RPAP1 and RPAP3 however, none of these proteins are essential. Conversely, the RPA2 protein acts alone and is essential (Stroud, Liddell and Allers, 2012).

After the initiator protein, replicative helicase, and replication factors have been recruited, the bi directional synthesis of DNA is initiated at the origin of replication. Shown below is a replication fork annotated with replication factors from all three domains of life (Figure 1.7).

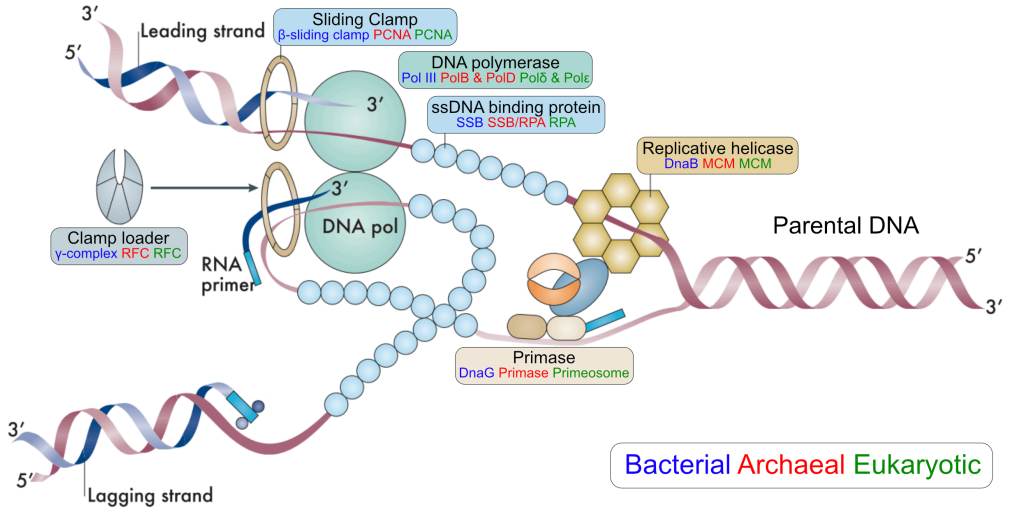


Figure 1.7 **Components of a replication fork.** A replication fork showing equivalent components from all three domains of lif, bacteria (blue), archaea (red), and eukaryotes (green). Figure adapted from (Lindås and Bernander, 2013).

Recombination-dependent Replication

DNA replication is typically origin-dependent. However, a few species of archaea that can initiate DNA replication in the absence of origins of replication have been identified. RadA is essential in strains lacking origins of replication, suggesting that homologous recombination is the alternative mechanism for initiation of DNA replication (Figure 1.8) (Michel and Bernander, 2014). Recombination-dependent replication is an alternative pathway possible in *Haloferax volcanii* due to the organisms polyploidy status. Not only is life possible, but recombination-dependent replication allows *H. volcanii* to grow faster when all of the main chromosomal origins are deleted, suggesting that origins of replication are selfish genetic-elements that ensure their own replication (Hawkins *et al.*, 2013).

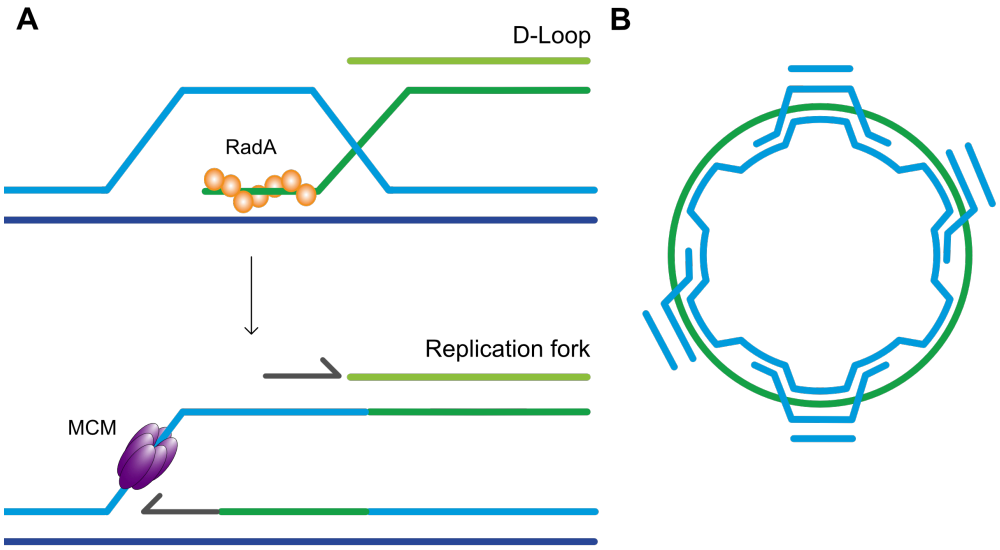


Figure 1.8 **Recombination dependent replication.** A) Remodelling of the D-loop into a replication fork. B) Recombination dependent replication can be initiated at any location on the chromosome. Multiple replication forks can be initiated at any given time Figure adapted from (Ausiannikava and Allers, 2017).

1.3.2 Elongation

The elongation stage of DNA replication can be further divided into two phases, priming and DNA synthesis. Both phases are discussed below.

Priming

DNA polymerases require a short primer from which to extend in order to initiate DNA synthesis. DNA primase synthesises RNA primers on both the leading and

lagging strands in order to initiate DNA synthesis (Le Breton *et al.*, 2007). DNA polymerase is then able to elongate the RNA primers and carry out DNA synthesis. DNA synthesis is continuous on the leading DNA strand (5'-3'), and discontinuous on the lagging strand (3'-5'), instead forming Okazaki fragments.

The bacterial primase is DnaG, which requires interaction with DnaB for initiation. The generation of the primer occurs in the 3'-5' direction, the opposite direction to the helicase unwinding activity of DnaB (Fang, Davey and O'Donnell, 1999). RNA primers synthesised by DnaG can vary in length from 10 to 60 nucleotides, though are typically around 11 nucleotides in length (Frick and Richardson, 2001).

The eukaryotic primase is a heterodimer of PriS and PriL subunits, which in turn associates with Pol α and the accessory B subunit to form the Pol α/primase complex. The Pol α/primase complex acts to synthesise and elongate RNA primers, generating an RNA-DNA hybrid. Primase synthesises RNA primers of 8 to 12 nucleotides. Pol α then elongates the primers to around 30 nucleotides before handing the resultant RNA-DNA hybrid over to the replicative polymerase (Arezi and Kuchta, 2000; Frick and Richardson, 2001; Kuchta and Stengel, 2011).

Archaeal species contain homologues of the eukaryotic PriS and PriL primase subunits but are lacking Pol α and the accessory B subunits. Archaeal primases typically synthesise RNA primers of 7 to 10 nucleotides long (Barry and Bell, 2006).

DNA synthesis

RNA primers, synthesised by primases, are then extended by DNA polymerases. In bacteria, DNA is synthesised by the C-family DNA polymerase III holoenzyme comprising two DNA polymerases in complex with a further nine subunits (Kelman and O'donnell, 1995). The polymerase core, responsible for DNA polymerisation and proof-reading, comprises polymerases α and ε, a 3'-5' exonuclease subunit, and the θ subunit. The γ-complex loads the β-sliding clamp, which is able to encircle and slide along the DNA. The β-sliding clamp in turn loads DNA polymerase III onto the DNA and increases the processivity of DNA polymerase III by stabilising the complex (Kuriyan and O'Donnell, 1993; Turner *et al.*, 1999).

Eukaryotes contain two distinct B-family DNA polymerases, known as Polε and Polδ. Polε and Polδ perform leading and lagging strand DNA synthesis respectively (O'Donnell, Langston and Stillman, 2013). The proliferating cell nuclear antigen (PCNA) is a sliding clamp with structural homology to the bacterial β-sliding clamp. PCNA aids Polε and Polδ in DNA replication by

tethering the replicative enzymes to the replication fork. Like the bacterial β-sliding clamp, PCNA requires a clamp loader known as replication factor C (RFC) in eukaryotes. RFC has a similar structure to the γ-complex (Moldovan, Pfander and Jentsch, 2007).

All archaea contain a B-family DNA polymerase called PolB. Euryarchaea and Thaumarchaeota have an additional DNA polymerase belonging to the D-family of polymerases, known as PolD. PolD is composed of the DP1 and DP2 subunits. DP1 is related to the eukaryotic Polδ and exhibits both 3'-5' exonuclease activity and proofreading ability (Cann *et al.*, 1998; Barry and Bell, 2006). Like eukaryotes, archaeal DNA polymerases are assisted by the PCNA sliding clamp, which is loaded onto to DNA by RFC (Cann *et al.*, 1999).

Lagging strand DNA synthesis is discontinuous, leading to the formation of Okazaki fragment flap structures alternating with RNA primers. To generate a continuous nascent strand of DNA, primer removal, gap filling, and ligation are necessary. In bacteria, the RNase H endonuclease is responsible for removal of the Okazaki fragment flaps before ligation (Reyes-Lamothe, Nicolas and Sherratt, 2012). In eukaryotes, PCNA coordinates Polδ, RNase H, flap endonuclease 1 (FEN1), Dna2, and DNA ligase I to carry out Okazaki fragment maturation (Zheng and Shen, 2011). RNA primers are removed and degraded by FEN1 and RNase H2 in archaea. DNA ligase LigA is then recruited to the nick site by PCNA to join the gap in the DNA (Barry and Bell, 2006). *H. volcanii* has an additional DNA ligase, LigN, which has been proposed to have an overlapping function in Okazaki fragment maturation (Zhao, Gray and MacNeill, 2006; Giroux and Macneill, 2015).

1.3.3 Termination

Bacteria have circular genomes and termination occurs opposite to the *oriC* when the bidirectional replication forks meet at the *Ter* sequences. *Ter* sequences are oppositely orientated from the oncoming replication fork, forming a replication trap in order to block replication forks from traveling through them (Duggin *et al.*, 2008). This occurs when the *Ter* sites are bound by the terminator protein Tus. The Tus-*Ter* complex blocks the action of the replicative helicase, DnaB, the mechanism of which is yet to be determined (Neylon *et al.*, 2005).

Eukaryotes, unlike bacteria, do not have defined termination sites. Instead termination occurs when two replication forks meet and the nascent strands are ligated together. This is a random event, which occurs between two origins of replication. At highly transcribed regions such as ribosomal operons, site specific fork barriers prevent the collision of the replication and transcription machinery and consequently terminate replication (Eydmann *et al.*, 2008)

Similarly, archaea do not appear to have defined termination sites. Termination is most likely to occur as a result of colliding replication forks from multiple origins of replication. Replication profiles in *Sulfolobus spp.* reveal that termination occurs asynchronously (Lundgren *et al.*, 2004). Replication profiles in *H. volcanii* reveal that termination happens over broad regions of the genome indicated by smooth valleys as opposed to sharp troughs (Hawkins *et al.*, 2013) (Figure 1.4 and Figure 1.6).

1.3.4 Lesions in the DNA template

Undamaged DNA is replicated accurately by the replication machinery. However, the presence of DNA lesions prevents the replication machinery from making faithful copies of the template DNA. Lesions on the lagging strand can be bypassed by a new priming event and repaired at a later time due to the discontinuous synthesis. However, lesions on the leading strand can result in replication fork stalling or collapse (Pages and Fuchs, 2003). Genomic mutations or wide-scale genome aberrations can arise if DNA lesions remain unrepaired.

1.4 DNA Repair

1.4.1 DNA Damage

Direct chemical damage to DNA can come from both endogenous (spontaneous) and exogenous (environmental) sources. Damage to DNA is largely unavoidable and can have serious consequences, including mutation of the genome or even death of the cell. Low levels of mutation may be tolerated however, high levels of mutation is incompatible with life. Therefore, efficient repair of DNA damage is essential. Archaea often inhabit extreme environments and therefore, would be expected to have robust DNA repair pathways (Figure 1.9).

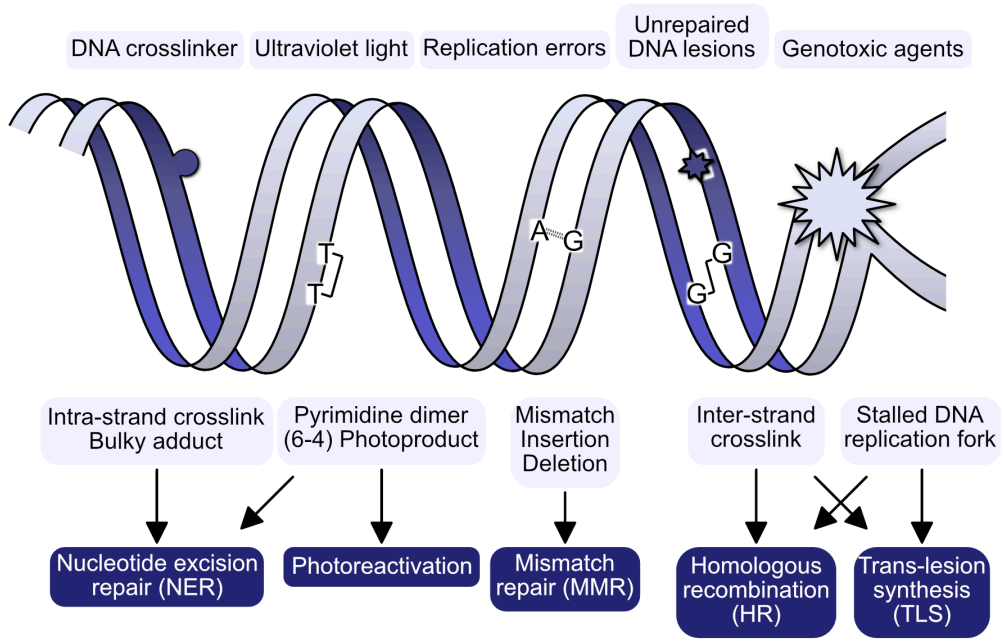


Figure 1.9 **Types of DNA damage.** Common types of DNA damage caused by exogenous and endogenous sources. Figure adapted from (White and Allers, 2018).

1.4.1.1 Endogenous

Normal metabolic processes within the cell generate chemical species with the potential to cause endogenous DNA damage.

Oxidative DNA lesions

Free radicals are by-products of normal cellular metabolism (Helleday, Eshtad and Nik-Zainal, 2014). Free radicals, such as reactive oxygen species and nitrogen oxide species, have the potential to cause over 25 different types of oxidative base lesions. One of the best understood oxidative base lesions is 8-oxo-2'-deoxyguanosine, which forms hydrogen bonds with adenine resulting in G·C to T·A base pair transitions (Evans, Dizdaroglu and Cooke, 2004). Breakages in the DNA phosphodiester backbone can also be a consequence of oxidative DNA damage.

Deamination

Deamination occurs spontaneously in all DNA bases that contain primary amines. Common reactions include hydrolytic deamination of 5-methylcytosine to thymine resulting in a G·C to T·A base pair transition (Lutsenko and Bhagwat, 1999); deamination of adenine to hypoxanthine resulting in a T·A to G·C base pair transition (Lindahl, 1993); and spontaneous or catalysed deamination of

cytosine to uracil. In higher eukaryotes, the latter is performed by members of the cytidine deaminase family such as Activation-induced cytidine deaminase (AID), which normally converts cytidine to uridine initiating the hypermutation process during immunoglobulin maturation (Teng and Papavasiliou, 2007). The principal repair mechanism for deamination is base excision repair (BER).

1.4.1.2 Exogenous DNA damage

Exogenous DNA damage is caused by chemical or physical agents external to the cell. Relevant examples for this thesis are described below.

Ultraviolet light

Ultraviolet (UV) radiation is a major source of DNA damage, which causes a variety of different types of DNA lesions including pyrimidine dimers, 6-4 photoproducts, and both single and double strand DNA breaks. Upon excitation by high-energy UV radiation, the C5 and C6 double bonds on neighbouring pyrimidine bases can become covalently crosslinked generating a cyclobutane pyrimidine dimer (CPD). Similarly, this excitation can induce the formation of a single bond between the C4 and C6 on neighbouring pyrimidine bases generating a 6-4 photoproduct (Ravanat, Douki and Cadet, 2001; Sinha and Häder, 2002) (Figure 1.10). Both CPDs and 6-4 photoproducts are 'bulky' and distort the phosphodiester backbone of DNA. Unrepaired 'bulky' lesions can result in G·C to T·A or GG·CC to TT·AA transitions during the next round of DNA replication (Helleday, Eshtad and Nik-Zainal, 2014). Excluding placental mammals, organisms can directly repair CPDs and 6-4 photoproducts using photolyases, this process is called photoreactivation. Nucleotide excision repair (NER) is another DNA repair mechanism with the ability to remove bulky DNA lesions. During NER, the section of DNA containing the DNA lesion is removed, and the resultant gap is filled in and ligated (Sinha and Häder, 2002). In humans, NER is the primary repair mechanism for repairing UV-induced lesions. The photoreactivation and nucleotide excision repair mechanisms will be discussed in further detail in sections 1.4.2.1 and 1.4.2.2.

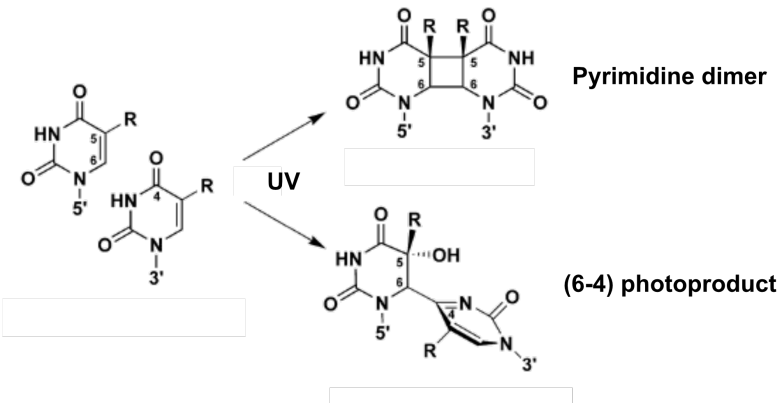


Figure 1.10 **UV-irradiation results in the formation of pyrimidine dimers and (6-4) photoproducts.** UV light is absorbed by pyrimidine C=C double bonds, exciting the electrons and allowing them to react with the neighbouring molecule. Covalent crosslinking between the C=C double bonds of adjacent thymadines form pyrimidine dimers. Covalent crosslinking between C6 and C4 of adjacent pyrimidines form (6-4) photoproducts. Figure adapted from (Li *et al.*, 2006).

Reactive oxygen species can also arise as a result of UV radiation. These free radicals can then cause single stranded DNA breaks (Lankinen, Vilpo and Vilpo, 1996; Cadet, Sage and Douki, 2005). If two single stranded DNA breaks occur in close proximity of one another or if a replication fork encounters a single stranded DNA break on the leading strand, a double strand DNA break (DSB) can arise. The repair of DSBs can be carried out by either homologous recombination (HR), non-homologous end joining (NHEJ), or microhomology-mediated end joining (MHEJ). These three repair mechanisms will be discussed in further detail in section 1.4.2.3.

DNA Crosslinking agents

There is a variety of chemical agents that have the ability to form crosslinks within DNA molecules. Due to the relevance in this thesis, only mitomycin C (MMC) will be discussed here. MMC is a naturally synthesised chemical mutagen produced by *Streptomyces caespitosus,* and is often used as a chemotherapeutic anti-tumour agent (Tomasz, 1995). A monoadduct between MMC and DNA forms when MMC alkylates a guanine base. A second alkylation event involving another guanine base generates a bisadduct. Depending on the location of the second guanine, the resultant crosslinks the DNA can be intra-strand, within the same strand of DNA, or inter-strand, between different strands of DNA. Both NER and homologous recombination can repair inter-strand crosslinks. Replication forks are unable to pass through DNA crosslinks and therefore, unrepaired DNA crosslinks can stall DNA replication.

1.4.2 DNA Repair Pathways

1.4.2.1 Direct Repair

The majority of organisms contain mechanisms to repair DNA via the direct reversal of the chemical change.

Photoreactivation

Photoreactivation is the reversal of UV-induced cyclobutane pyrimidine dimers and 6-4 photoproducts using lesion specific photolyases. In *E. coli*, photolyases are monomeric and contain the chromophore cofactor folate, which absorbs violet/blue light photons and consequently excites a second cofactor, flavin (Li, Sancar and Heelis, 1991). The excited flavin transfers an electron to the cyclobutane pyrimidine dimer to generate a dimer radical anion, which in turn splits the two pyrimidine bases, thus reversing the damage (Sancar *et al.*, 2004). Placental mammals lack photolyases and therefore are unable to directly repair UV-induced cyclobutene pyrimidine dimers and 6-4 photoproducts via photoreactivation.

Conversion of O^6-methylguanine to Guanine

The removal of the O^6-methyl group from O^6-methylguanine (O^6MeGua) bases by methylguanine DNA methyltransferases is another mechanistic example of direct reversal of DNA damage. The methylguanine DNA methyltransferase flips out the O^6MeGua DNA base and removes the methyl group. A cysteine located in the enzyme's active site accepts the methyl group causing the enzyme to become inactive after only one catalytic event and to dissociate from the DNA (Olsson and Lindahl, 1980). Thereupon, the methyltransferase is activated as a transcription factor, leading to the expression of other genes acting in the repair of alkylated DNA lesions. Mice lacking the *MGMT* gene, which encodes methylguanine DNA methyltransferase, are highly susceptible to tumourigenesis by DNA alkylating agents (Kawate *et al.*, 1998).

1.4.2.2 Excision Repair

Excision repair involves cutting the DNA strand either side of the damaged bases or nucleotides, and removing this damaged section. The undamaged strand acts as a template for DNA synthesis to fill in the gap. Defects in excision repair pathways result in a higher incidence of cancer susceptibility.

Base Excision Repair

Base excision repair (BER) rectifies small DNA lesion damage that arise from oxidation, deamination, methylation, and alkylation. Small DNA lesions do not significantly distort the DNA helix. Misincorporated uracil is also removed and replaced via the BER pathway. The main components of BER are conserved throughout the three domains of life (Sartori and Jiricny, 2003). Defects in BER have been implicated in age-related diseases such as neurodegeneration and cancer (Wilson and Bohr, 2007).

Lesion-specific DNA glycosylases initiate the BER pathway by flipping out the damage base and cleaving the N-glycosylic bond between the damaged base and the deoxyribose. The removal of the base leaves an apurinic or apyrimidinic (AP) site. An AP endonuclease generates a single strand break by cutting the sugar-phosphate backbone at the 5' side of the AP site. The downstream fragment is terminated with a baseless deoxyribose-phosphate (dRP) residue and the upstream fragment is terminated with a free 3' hydroxyl group. The removal of the baseless dRP varies between organisms. In *E. coli* the baseless dRP is removed by a 5' phosphodiesterase such as RecJ (Piersen, McCullough and Lloyd, 2000). In some organisms, the DNA glycosylases display additional beta lyase activity that cleaves the phosphodiester bond 3' to the AP site. In mammalian cells, processing of the AP sites can occur via two different pathways. These pathways are termed 'short-patch' and 'long-patch' BER, which result in the replacement of a single nucleotide and 2-6 nucleotides respectively. The 'short-patch' BER method is typically preferred (Dogliotti *et al.*, 2001; Sartori and Jiricny, 2003). DNA polymerase β fills in the nucleotide gap and consequently removes the dRP moiety by elimination. The remaining nick is then ligated (Matsumoto and Kim, 1995). 'Long-patch' BER generates a longer 'flap' oligonucleotide during the gap filling process. The overhanging flap is excised by the flap-endonuclease (FEN1), which is present in eukaryotes and archaea. DNA ligase then seals the resultant nick (Lindahl, 1993) (Figure 1.11).

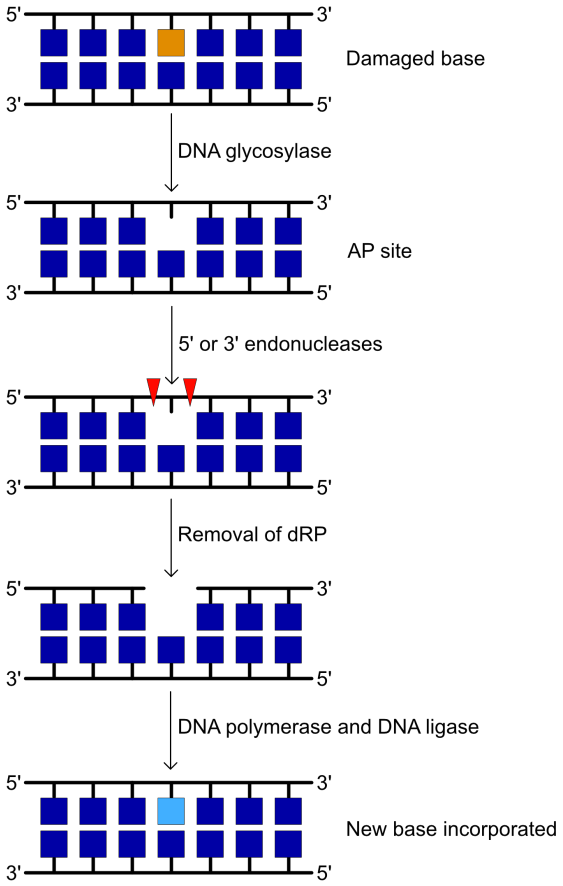


Figure 1.11 **Base excision repair.** Glycosylases remove damaged bases resulting in an apurinic or apyrimidinic (AP) site. AP endonuclease cleave the sugar-phosphate backbone to remove the baseless deoxyribose-phosphate (dRP). DNA polymerase uses the complementary strand of DNA as a template to fill in the gap and DNA ligase seals the nick.

Nucleotide Excision Repair

Nucleotide excision repair (NER) is a highly versatile DNA damage removal pathway and is considered to be relatively 'error free'. Lesions repaired by NER tend to cause significant distortion to the DNA helix (de Laat, Jaspers and Hoeijmakers, 1999). Defects in NER results in extreme photosensitivity and predisposition to cancer as seen in patients with xeroderma pigmentosum (XP), Cockayne syndrome (CS), and the photosensitive form of trichothiodystrophy (TTD) (De Boer and Hoeijmakers, 2000).

The stages of NER are lesion recognition, 3' and 5' sequential dual incisions, unwinding, patch repair synthesis, and ligation. These general stages are conserved throughout all three domains of life. However, the bacterial and eukaryotic proteins responsible show very little homology and mammalian NER is overall more complex than in other species (Figure 1.12).

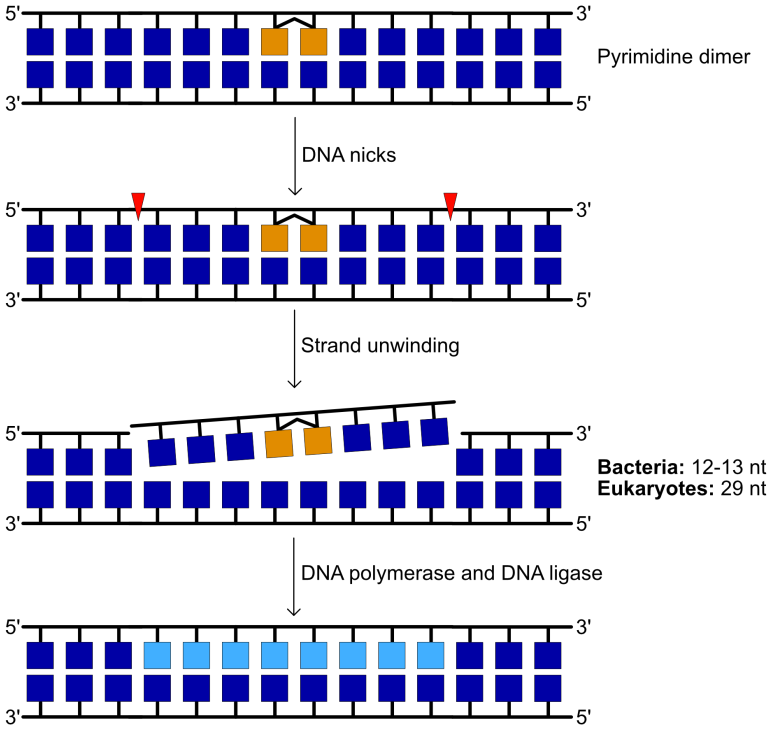


Figure 1.12 **Nucleotide excision repair.** Endonucleases make incisions 3' and 5' to the DNA lesion, shown here as a pyrimidine dimer. Helicases unwind the oligonucleotide. DNA polymerase uses the complementary strand of DNA as a template to fill in the gap and DNA ligase seals the nick.

The UvrA-UvrB protein complex, with an A_2B_1 stoichiometry, identifies the DNA lesion in *E. coli.* The UvrA subunit binds the DNA lesion and the UvrB helicase locally kinks the DNA in order to unwind approximately 5bp around the damaged DNA. UvrB then binds tightly to the ssDNA lesion via a beta hairpin and isomerises, causing the UvrA subunit to dissociate from the complex. The UvrB-DNA complex is then bound by UvrC, which makes a 3' followed by a 5' incision in the DNA, generating a 12-13 nucleotide oligomer. UvrD, also known as helicase II, is then recruited to the nick sites to remove the oligomer and cause dissociation of UvrC. DNA polymerase I is able to access the 5' site hydroxyl residue in order to synthesise DNA across the gap. DNA ligase then seals the remaining nicks (Petit and Sancar, 1999).

In eukaryotes, NER follows the same steps as the bacterial pathway but a more complex set of proteins is involved, which has evolved independently. Global genomic repair (GGR) and transcription coupled repair (TCR) are sub pathways both found in eukaryotes. GGR is used throughout the genome, whereas TCR is specific to the repair of lesions that arrest the RNA polymerase in the transcribed

strand of expressed genes. The TCR sub pathway is also found in bacteria (Hanawalt, 1994).

The human UvrABC counterpart comprises 16-17 proteins organised into six repair factors known as XPA, RPA, TFIIH, XPC-RAD23B, XPG, and ERCC1-XPF (Scharer, 2013). A stalled RNA polymerase upstream (5') of the DNA lesion is thought to be recognised by the CSA and CSB proteins, which are mutated in Cockayne syndrome. CSA and CSB together initiate TCR by recruiting the other NER factors to the DNA lesion site (Costa *et al.*, 2003). In contrast, the XPC subunit of the XPC-RAD23B repair factor recognises thermodynamically destabilised duplex DNA to initiate GGR. XPC is also recruited by ubiquitin ligase activated by DDB1/DDB2 upon recognition of UV damaged DNA in heterochromatin. RAD23B acts to stabilise XPC and dissociates from XPC upon DNA damage recognition (Sugasawa *et al.*, 1996, 2001). XPC-RAD23B recruits the TFIIH repair factor to the lesion site. TFIIH consists of a core (XPB, p52, p62, p34, p44, and TTDA) associated to the Cdk-activating kinase (CAK) sub-complex via the XPD helicase. XPB and XPD are 3'-5' and 5'-3' DNA helicase respectively, and are responsible for unwinding the DNA around the lesion (Tapias *et al.*, 2004; Oksenych and Coin, 2010). Assembly of the pre-incision complex is triggered by the association of XPD with the DNA lesion. The pre-incision complex consists of XPA, which is thought to be a regulatory factor that maintains the proteins within the complex; RPA, a single-strand DNA binding protein that co-ordinates excision and repair synthesis events; and XPG, an exonuclease responsible for making a 3' incision to generate a hydroxyl group. The hydroxyl group is used to initiate repair synthesis. Following the 3' incision by XPG, ERCC1-XPF makes a 5' incision generating a 24-32 oligonucleotide containing the damaged DNA, which then disassociates. DNA polymerases δ and ε, assisted by the PCNA sliding clamp and the clamp loader RFC, fill in the gap and the resultant nick is then sealed by DNA ligase I.

Most archaea contain homologues of the eukaryotic NER proteins such as XPF, XPG, XPB, and XPD (Grogan, 2000; White, 2003). However, some archaeal species are lacking either one or both of the XPB and XPD helicases, suggesting that this helicase cooperation is not essential during archaeal NER. Sequence identity between archaeal and eukaryotic XBP proteins is approximately 25-30%, and most crenarchaea possess multiple copies (Richards, Cubeddu, *et al.*, 2008). The archaeal nuclease Bax1 co-transcribes with XPB and is thought to be the archaeal equivalent of the eukaryotic XPG. Long and short version of archaeal XPF have been identified. The long form, Hef, is found exclusively in euryarchaea, and contains nuclease and helicase domains; it is homologous to the FANCM protein that acts in inter-strand crosslink repair in higher eukaryotes (Meetei *et al.*, 2005). The short version is found in crenarchaea and thaumarchaea, and contains a nuclease domain only (Horiuchi *et al.*, 2004;

Nishino, Komori, Tsuchiya, *et al.*, 2005). The crenarchaeal XPF catalytic activity is dependent on the PCNA sliding clamp (Roberts, Bell and White, 2003).

Orthologues of the bacterial UvrABC system have been found in some mesophilic methanogens and halophiles, for example *Methanothermobacter thermoautotrophicus* and *H. volcanii* respectively (Costa *et al.*, 2003; Lestini, Duan and Allers, 2010). Some archaea, such as *Methanosarcina mazei,* have a mixture of bacterial UvrABC and eukaryotic XPF orthologues (White, 2003). It is possible that in some species the bacterial-like NER pathway has replaced the original eukaryotic-like NER pathway via lateral gene transfer (Kelman and White, 2005). The bacterial and eukaryotic pathways could act in tandem within the cell, alternatively the archaeal NER proteins may possess additional functions.

No XPC-RAD23B homologues have been identified in archaea therefore, it remains unclear how DNA damage is detected in archaea lacking the bacterial UvrABC system. The *Halobacterium* RPA is upregulated following UV irradiation and the *S. solfataricus* SSB protein is capable of recognising DNA damage such as mismatches, photoproducts, and bulky lesions *in vitro*, suggesting that the single stranded binding protein SSB/RPA could act as the DNA damage recognition factor (Cubeddu and White, 2005; McCready *et al.*, 2005).

Mismatch Repair

DNA Mismatch repair (MMR) corrects mismatched base pairs that arise primarily from replication errors. This is an important process for maintaining genome stability. The MMR proteins, and the overall mechanism, is highly conserved across most species of bacteria and eukaryotes. However, such MMR proteins are found in only a limited set of archaeal species. The ‘long-patch’ mechanism is characteristic of post-replicative MMR and involves the exclusion of a long oligonucleotide. MMR increases the accuracy of DNA replication between 20- and 400-fold in *E. coli* (Schaaper, 1993). Mutations and epigenetic splicing of the MMR proteins have been implicated in hereditary non-polyposis colorectal cancers in humans (Jiricny, 2006; Hsieh and Yamane, 2009). There are two known types of mechanisms for MMR, an *E. coli* (and closely related species) specific mechanism, and a more widely employed mechanism found in eukaryotes and most bacteria.

In the *E. coli* (and closely related species) specific mechanism, mismatched DNA is guided by the absence of methylation at GATC sequences, this also serves to restrict DNA repair to the newly synthesised strand of DNA known as the daughter strand (Fukui, 2010). The daughter strand remains unmethylated

for up to 2 minutes following synthesis, due to the slow rate of methylation by Dam methylase. This allows time for the MutS homodimer to recognise mismatched DNA. Upon recognition, MutS recruits MutL to activate the restriction endonuclease MutH. MutH generates an entry point for the excision reaction to occur by nicking the daughter strand (unmethylated) at a hemimethylated GATC site (Smith and Modrich, 1996). The daughter strand is then nicked either upstream (5') or downstream (3') of the DNA damage by exonuclease ExoVII/RecJ or ExoI/ExoX respectively. The UvrD helicase then removes the oligonucleotide (Li, 2008). DNA polymerase III fills in the gap and the resultant nick is then sealed by DNA ligase. MutS and MutL homologues are found in most bacteria and eukaryotes however, homologues of *E. coli* MutH are sparse.

The more widely employed mechanism, found in eukaryotes and most bacteria, recognises mismatched base pairs via strand discontinuities. The eukaryotic MutSα sliding clamp recognises base-base mismatches and insertions/deletions of up to two nucleotide in length. Larger nucleotide insertions/deletions are recognised by the MutSβ sliding clamp (Jiricny, 2006; Hsieh and Yamane, 2009). The discontinuous DNA strand is nicked either upstream (5') or downstream (3') of the DNA damage by MutLα (Kadyrov *et al.*, 2006; Modrich, 2006). The damaged DNA segment is then excised by the exonuclease EXOI in concert with the single stranded DNA binding protein, RPA. DNA polymerase δ, assisted by the PCNA sliding clamp and the clamp loader RFC, fill in the gap and the resultant nick is then sealed by DNA ligase I (Genschel and Modrich, 2009). Similarly in bacteria lacking MutH, mismatched base pairs and short insertion/deletion loops are recognised by MutS and nicked by MutL. The damaged DNA segment is then excised by the exonucleases such as ExoI and RecJ, and a helicase such as UvrD, both in concert with the single stranded DNA binding protein, SSB. DNA polymerase III fills in the gap and the resultant nick is then sealed by DNA ligase.

Hyperthermophiles, as a whole, lack any MMR machinery with homology to the MutS and MutL system employed by eukaryotes and bacteria. However, some hyperthermophilic archaea such as *Sulfolobus spp.* exhibit effective mutation-avoidance mechanisms in response to DNA damaging agents, suggesting that an alternative repair pathway is available to these archaea (Grogan, 2004). Phylogenetic analysis has detected MutS and MutL orthologues in only nine halophilic and methanogenic archaeal species, which appear to have arisen by horizontal gene transfer from bacteria (Lin, Nei and Ma, 2007). EndoMS, originally known as "NucS", is a mismatch specific endonuclease originally identified in *Pyrococcus furiosus* but has also been found in other archaea such as *Thermococcus kodakarensis*, and suggests the presence of a novel archaeal mismatch repair mechanism (Ishino *et al.*, 2016; Nakae *et al.*, 2016). The *H. volcanii* genome contains two homologues of MutL and four of MutS. Only the

two MutS protein that are encoded with the MutL proteins have been shown to function in the *H. volcanii* MMR mechanism (Hartman *et al.*, 2010).

1.4.2.3 Double Strand Break Repair

Both exogenous and endogenous sources can induce double strand breaks (DSB). DSBs are intentionally formed during the eukaryotic processes of meiosis, responsible for creating genome diversity, and V(D)J recombination, which is important in the development of T-cells and B-cells in the vertebrate immune system (Davis and Chen, 2013). Failure to repair DSBs correctly results in cellular breakage and rearrangements, leading to cellular death and a predisposition to immune system disorders and cancers (Hefferin and Tomkinson, 2005).

The repair of DSBs can be achieved by three different mechanisms: non-homologous end joining (NHEJ), microhomology mediated end joining (MMEJ), and homologous recombination (HR) (Figure 1.13). End joining is a rapid method of DSB repair however, these mechanisms are inherently error-prone. In contrast, homologous recombination is a more accurate method requiring undamaged homologous DNA to use as a template for repair but is a far slower process.

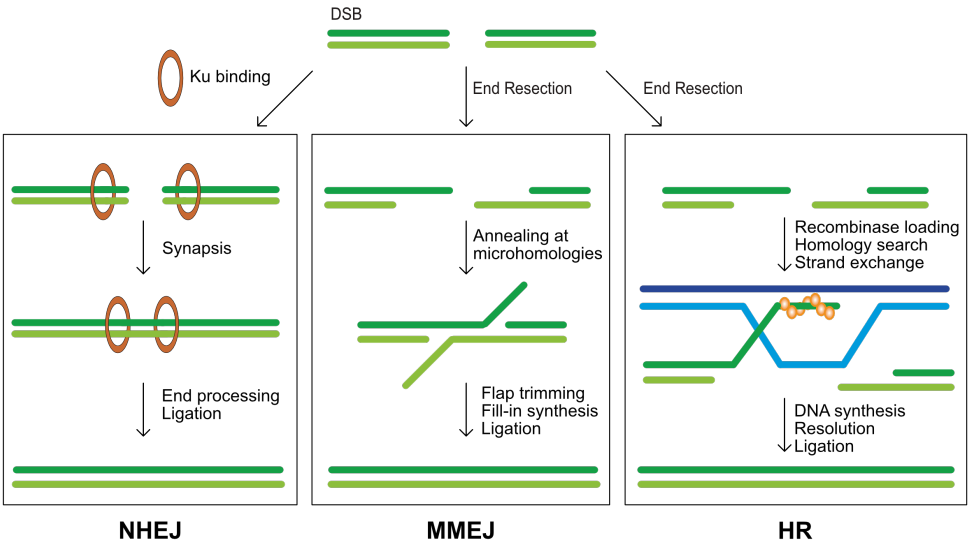


Figure 1.13 **Double-strand break repair pathways.** The repair of DSBs can be achieved by three different mechanisms. During non-homologous end joining (NHEJ)(left), Ku binds to the dsDNA ends at the break site and recruits bridging factors, polymerases and exonucleases, which process the ends. The DNA ends are then ligated to fix the DSB. During microhomology mediated end joining (MMEJ)(middle), DNA ends are processed to reveal microhomologies, which then align and anneal. Gaps are filling is carried out by translesion polymerases. DNA ligase seals the gap. During homologous recombination (HR)(right), DNA ends are resected to allow recombinases to load onto the ssDNA. The recombinase-DNA filament invades an intact homologous DNA molecule to use as a template for DNA repair.

The cell cycle largely influences the mechanism of repair used in eukaryotes. During S and G2 phases, HR is favoured due to the presence of a homologous sister chromatid that can be used as a template following DNA replication. Unlike HR and MMEJ, NHEJ does not involve end resection and thus is the favoured mechanism during the G1 phase when end resection activity is low within the cell (Ira *et al.*, 2004). Both NHEJ and MMEJ occur during all stages of the cell cycle and involve DNA tethering proteins, known as Ku and PARP1 respectively. NHEJ and has predominance over MMEJ as Ku has a higher affinity for DNA compared with PARP1.

In addition to the cell-cycle phase, the Mre11-Rad50 complex also influences the mechanism of DSB repair used. Mre11 is a single stranded DNA nuclease and Rad50 is an ATPase. The Rad50 protein contains a zinc hook motif at the tip of a long coiled domain, which allows for intermolecular interactions between other Rad50 molecules, DNA molecules, or chromosomes (Paull and Gellert, 2000; De Jager *et al.*, 2001; Hopfner *et al.*, 2002). The Mre11-Rad50 further complexes with Xrs2 in vertebrates or with Nbs1 in yeast. An Xrs2/Nbs1 homologue is missing in archaea despite the presence of the Mre11 and Rad50 homologues.

In thermophilic archaea, it is common to find Mre11 and Rad50 within an operon containing the hexameric helicase HerA and the 5' to 3' nuclease NurA (Constantinesco, Forterre and Elie, 2002; Constantinesco *et al.*, 2004; White, 2011). It is suggested that the HerA-NurA complex could fulfil the role of the Xrs2/Nbs1 protein but, the HerA and NurA proteins are not present in *H. volcanii.* The Xrs2/Nbs1 protein acts to recruit the Mre11-Rad50 complex to sites of DNA damage through interactions with the histone γH2AX and mediates protein interactions between the Mre11-Rad50 complex and other DNA repair factors (Kobayashi *et al.*, 2002). The Mre11-Rad50 complex tethers DSBs and controls the end resection step. The end resection step promotes HR and MMEJ mechanisms, whereas the NHEJ mechanism is promoted by limited end processing (Symington and Gautier, 2011). In *H. volcanii,* the Mre11-Rad50 complex restrains homologous recombination at DSB sites to allow time for the MMEJ mechanism to repair the damage instead (Delmas *et al.*, 2009).

The ATM (ataxia-telangiectasia mutated) and ATR (ATM- and Rad3-related) kinases also act alongside their downstream kinases in order to direct DNA repair pathways in mammalian cells. ATM detects DSBs and in turn recruits the Mre11-Rad50-Nsb1 complex to the site of the damage (Maréchal and Zou, 2013). ATR interacts with the ATR-interacting protein (ATRIP) to detect ssDNA present at stalled replication forks and the ssDNA generated from DSB processing. Both ATM and ATR initiate signalling cascades involving checkpoint kinases Chk1 and Chk2. Chk1 and Chk2 initiate a secondary wave of phosphorylation events in a large signalling network (Matsuoka *et al.*, 2007). One such target for phosphorylation is BRCA1, a promoter of homologous recombination (Sancar *et al.*, 2004).

Non-homologous end joining

Non-homologous end joining (NHEJ) often results in minor changes of the DNA sequence at the break site and occasionally leads to chromosomal rearrangements by the joining of previously unlinked DNA molecules (Davis and Chen, 2013).The basic mechanism of NHEJ is conserved throughout the eukaryotes, some bacteria and a very limited number of archaea. The NHEJ mechanism follows the recognition of DSBs, the bringing together of DNA broken ends, the processing of these ends, ligation, and disassembly of the NHEJ complex.

In eukaryotes, the Ku heterodimer, comprising Ku70 and Ku80 subunits, recognises and binds DSBs. The Ku70 and Ku80 subunits form a ring structure that encircles the dsDNA (Walker, Corpina and Goldberg, 2001). Ku forms a scaffold at both ends of the DSB, and in turn recruits the NHEJ factors (Davis and Chen, 2013). The DNA-dependent protein kinase catalytic subunit (DNA-PKcs) is an NHEJ factor involved in forming the synaptic complex responsible

for bringing the DNA ends together. Non-compatible single stranded DNA overhangs are either filled in by DNA polymerases μ and λ, or removed by exonucleases such as Artemis. Loss of nucleotides often occurs during this step. Ligation of the processed ends is catalysed by the ligase IV and XRCC4 complex aided by the XRCC4-like factor (XLF).

The presence of a Ku homologue typically denotes the presence of the NHEJ pathway, however not all bacteria contain a Ku gene; for example the *E. coli* K12 strain lacks a Ku gene. There is no obvious phylogenetic pattern to the distribution of the NHEJ apparatus, suggesting that NHEJ systems are acquired by horizontal gene transfer (Bowater and Doherty, 2006). Bacterial Ku proteins are typically smaller than their eukaryotic counterparts and are predominantly homodimers (Aravind and Koonin, 2001; Doherty, Jackson and Weller, 2001). Bacterial Ku genes are typically expressed within an operon alongside DNA dependent ligases such as LigD, forming a species-specific NHEJ complex (Weller *et al.*, 2002). The Ku associated ligases contain a PolX family polymerase domain and demonstrate exonuclease activity, suggesting that these ligases 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 Ku have been identified in several species of archaea including *Archaeoglobus fulgidus,* and homologues of the bacterial LigD phosphoesterase domain have been observed in several Euryarchaeota species. There is no evidence of NHEJ apparatus homologues in *H. volcanii* (Aravind and Koonin, 2001; Bowater and Doherty, 2006; Nair, Smith and Shuman, 2010; Smith *et al.*, 2011)

Microhomology-mediated end joining

Microhomology-mediated end joining (MMEJ) involves the alignment and annealing of microhomologous sequences at the broken DNA ends prior to ligation. Like NHEJ, MMEJ is error prone and can lead to deletions/insertions, and chromosomal translocations at the DNA break point.

The tethering protein in MMEJ is poly ADP-ribose polymerase 1 (PARP1). PARP1 recognises and binds DSBs, the DNA ends 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 nick the DNA strands at the site of a DSB, initiating the end resection process. The 5'-3' exonuclease Exo1 then creates a ssDNA tract (Symington and Gautier, 2011). End resectioning in mammalian cells generates a 1-18 nucleotide tract, which is maintained by the binding of RPA (ssDNA binding protein) (Sfeir and Symington, 2015). The resected ends anneal to one another via complementary base pairing and the translesion DNA polymerase

Polθ fills in the ssDNA gaps. Polθ, encoded by the *POLQ* gene, contains an N-terminal helicase domain and a C-terminal polymerase domain, which lacks proof reading capability. The helicase domain shows homology with HelQ (the human Hel308 homologue) and is critical in the regulation of DNA end joining (Marini and Wood, 2002; Chan, Yu and McVey, 2010; Ceccaldi *et al.*, 2015). Hel308 will be discussed in further detail in Section 1.7. After DNA synthesis by Polθ, the endonuclease XPF-ERCC1 complex removes the heterologous 3' flaps and DNA ligases Lig1 and Lig3 ligate the DNA ends.

The multidomain protein LigD, comprising polymerase, nuclease, and ligation domains, is responsible for the majority of the MMEJ pathway in some bacteria (Pitcher, Brissett and Doherty, 2007). The Mre11-Rad50 complex in *H. volcanii* prevents the repair of double strand breaks by homologous recombination and promotes the MMEJ pathway instead (Delmas *et al.*, 2009). MMEJ is largely understudied in prokaryotes compared to eukaryotes.

Homologous Recombination

Homologous recombination (HR) is considered to be an error-free double strand break repair mechanism. Homologous recombination requires a sequence of homology, typically a sister chromatid during S or G2 phase of the cell cycle, as a template for repair of the damaged DNA strand. HR plays a role in the repair of DNA lesions, DNA replication in the absence of origins, and maintaining genome stability. The understanding of HR is essential to this study and will be discussed in further detail below.

1.5 Homologous Recombination

Homologous recombination (HR) involves the exchange of nucleotide sequence between near-identical DNA sequences and is considered as an error-free method of DNA repair. HR is critical for the repair of a variety of DNA lesions, for example double strand breaks, single strand breaks/gaps, and interstrand DNA crosslinks. In addition to DNA repair, HR is involved in other DNA processes such as the restart of stalled/broken replication forks, chromosomal pairing, and exchange during meiosis (Michel *et al.*, 2001; Neale and Keeney, 2006). Failure to repair DNA damage and resolve DNA replication stress in a timely manner leads to genomic instability, a contributing factor to cancer development. There are several known genetic mutations that render the cell unable to fulfil this fundamental process, for example mutations in the BRCA genes, the FA genes, and RecQ helicases. Mutations in the BRCA1 and BRCA2 genes lead to a predisposition to breast and ovarian cancers (Venkitaraman, 2002). Defects in proteins such as FancD1, FancD2, and FancJ, render the cell unable to repair interstrand crosslinks causing Fanconi anemia, which is a cancer

predisposition syndrome (Kee and D'Andrea, 2010). Mutations in the RecQ helicase BLM, named for Bloom's syndrome, promotes excessive crossing over of sister chromatids, which leads to increased levels of chromosomal rearrangements, genomic instability, and a predisposition to cancer (Wu and Hickson, 2003). Mutations in the RecQ helicase WRN, named for Werner syndrome, result in defects in homologous recombination resolution and cell division. This ultimately leads to genetic instability, a predisposition to cancer, and premature aging (Saintigny *et al.*, 2002). Unrestrained HR can lead to undesirable DNA rearrangements and therefore, many regulatory mechanisms have evolved to ensure accurate and timely repair by the homologous recombination repair pathway. This will be discussed in further detail in section 1.5.4.

The overall mechanism of HR is conserved throughout the three domains of life and can be summarised in three main stages; pre-synapsis, synapsis, and post-synapsis. DNA is resected during pre-synapsis in order to generate ssDNA 3' overhangs for the sequence homology search. Once found, the ssDNA overhang invades the dsDNA homologous region forming a D-loop. During post-synapsis, the DNA complex is resolved to allow DNA repair or stalled replication fork restart. Crossover and non-crossover products arise when an exchange of genetic material between the two strands has taken place and no exchange of genetic material occurs, respectively (Figure 1.14). Each stage of homologous recombination will be discussed in further detail in sections 1.5.1 to 1.5.3.

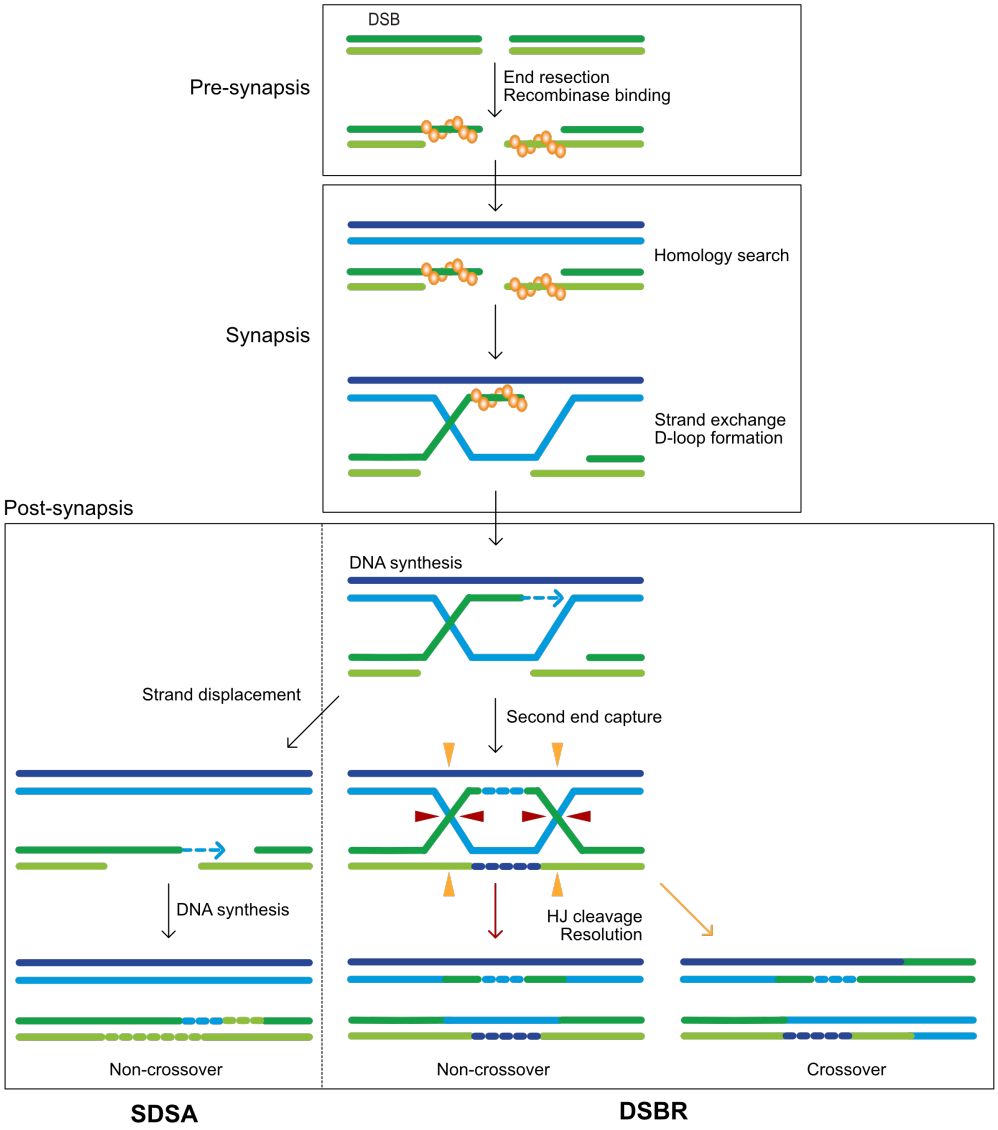


Figure 1.14 **Homologous recombination pathways.** A schematic of homologous recombination using the repair of double strand breaks as an example. Recombination occurs in three stages: pre-synapsis, synapsis, and post-synapsis. Blue DNA is the homologous template DNA, green DNA is the damaged DNA, orange spheres indicate recombinases, dashed lines indicate newly synthesised DNA, and arrow heads indicate cleavage events.

1.5.1 Pre-synapsis

During pre-synapsis, dsDNA is resected in a 5'-3' direction, generating ssDNA 3' overhangs. Recombinases are loaded onto the 3' ssDNA overhangs, displacing the single stranded binding proteins in the process. The bacterial recombinase is RecA, the eukaryotic recombinases are Rad51 (repair/fork restart) and DMC1 (during meiosis), and the recombinase in archaea is RadA (Figure 1.15).

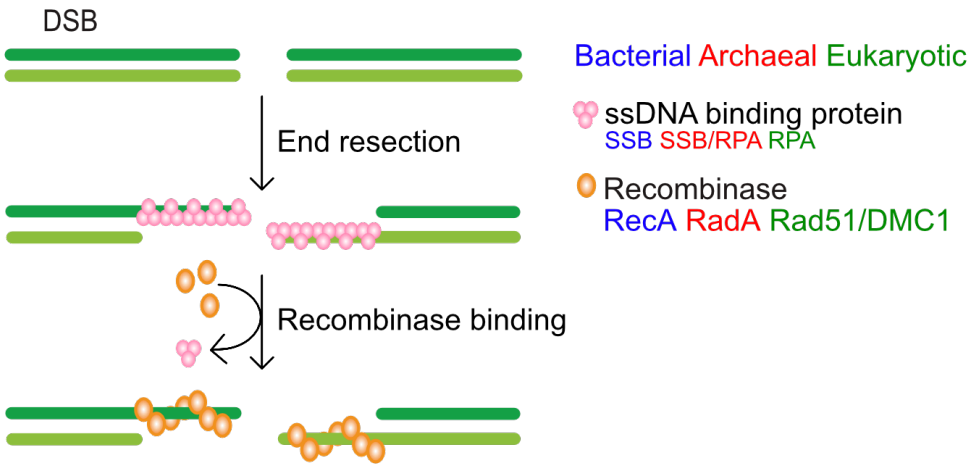


Figure 1.15 **Homologous recombination pre-synapsis.** 5' to 3' DSB end resection followed by displacement of single stranded DNA binding protein and recombinase nucleoprotein filament formation upon the 3' ssDNA tails. Recombinase filament formation is a dynamic and reversible process.

Bacterial pre-synapsis

Both the RecBCD and RecFOR pathways can initiate homologous recombination in *E. coli.* Initiation via either pathway results in ssDNA onto which RecA, the bacterial recombinase, is loaded. Once RecA is loaded, the ssDNA then invades a homologous DNA molecule forming a D-loop (Rocha, Cornet and Michel, 2005). The RecBCD holoenzyme first binds to dsDNA and demonstrates both helicase and nuclease functions that unwind and degrade dsDNA until it reaches a χ site. A χ site (5'-GCTGGTGG-3') is a regulatory DNA sequence that acts to mitigate the 3' nuclease activity of RecBCD (Dixon and Kowalczykowski, 1993). The helicase activity of RecBCD is subsequently promoted at the χ site, generating a 3' ssDNA on onto which RecA, the bacterial recombinase, is loaded (Kowalczykowski, 2000; Singleton *et al.*, 2004). In contrast, the RecFOR complex binds to gapped ssDNA, displacing the single stranded binding protein (SSB), and loads RecA onto the ssDNA (Morimatsu and Kowalczykowski, 2003). The endonuclease RecJ, acting alongside RecFOR, enlarges the ssDNA region in necessary. In both the RecBCD and RecFOR pathways, RecA forms a right-handed helical filament on the ssDNA, with six RecA molecules and 18 nucleotides per turn, thereby stretching the DNA 1.5x. During synapsis, the RecA filament is able to slide along the dsDNA molecule in search of a homologous sequence (Ragunathan, Liu and Ha, 2012).

Eukaryotic pre-synapsis

End resection is carried out by the Mre11-Rad50-Xrs2 (MRX) complex alongside Sae2 in yeast, and the Mre11-Rad50-Nbs1 (MRN) complex alongside

CtIP in mammals (Jasin and Rothstein, 2013). BRCA1 interacts with MRN and CtIP, promoting homologous recombination and single strand annealing, suggesting that BRCA1 has an end resectioning role (Stark *et al.*, 2004). If extensive end resectioning is required, either the Exo1 exonuclease or the helicase/nuclease Sgs1/Dna2 are recruited to process the DNA ends (Mimitou and Symington, 2008).

Rad51 is the eukaryotic homologue of the bacterial RecA. Rad51 is loaded onto ssDNA that either arises from perturbations in DNA replication or is generated from resecting 5' strands at DSBs. The loading of Rad51 displaces the single stranded binding protein RPA, this is achieved with the assistance of recombination mediators BRCA2 (in higher eukaryotes), Rad52, Rad54, and Swi5-Sfr2 (Benson, Baumann and West, 1998; Kurokawa *et al.*, 2008; Mazin *et al.*, 2010). The Rad55-Rad57 heterodimer in yeast is a recombination mediator, which acts to promote the stability of the Rad51 presynaptic filament (Sung, 1997; Krejci *et al.*, 2002). It is proposed that in higher eukaryotes, BRCA2 is a recombinase mediator acting to promote the assembly of Rad51 onto ssDNA in preference over dsDNA, and to aid the displacement of RPA (Jensen, Carreira and Kowalczykowski, 2010). Like bacterial RecA, Rad51 is loaded onto the ssDNA in a right-handed helical geometry, with six Rad51 molecules and 18 nucleotides per turn (Chen, Yang and Pavletich, 2008). This geometry stretches the ssDNA by 50%, which aids efficient homology search (Klapstein, Chou and Bruinsma, 2004). Dmc1 is the meiosis-specific recombinase used in place of Rad51. Dmc1 is able to displace RPA and form filaments on ssDNA. This action catalyses homologous recombination between sister chromatids at programmed DSBs during meiosis (Sehorn *et al.*, 2004).

Archaeal pre-synapsis

The eukaryotic Mre11 and Rad50 homologous recombination initiation proteins are conserved in archaea. The homologous proteins in bacteria are SbcD and SbcC. It is common to find Mre11 and Rad50 within an operon containing the hexameric helicase HerA and the 5' to 3' nuclease NurA, but only in thermophilic archaea (Constantinesco, Forterre and Elie, 2002; Constantinesco *et al.*, 2004; White, 2011). In *Pyrococcus furiosus*, Mre11-Rad50 generate short 3' overhangs that act as a platform from which the HerA-NurA complex initiates end resection (Hopfner *et al.*, 2000). Conversely, the Mre11-Rad50 complex delays the repair of DSBs by HR in *H. volcanii*, suggesting that Mre11-Rad50 acts to control the entry into the HR pathway (Delmas *et al.*, 2009). Furthermore, the HerA and NurA proteins are not present in *H. volcanii.* Like bacteria and eukaryotes, the SSB/RPA homologue binds to the 3' ssDNA overhangs after end resection occurs and must be displaced to allow binding of the archaeal recombinase, RadA. Deletion of the *radA* gene in *H. volcanii* results in severe growth, DNA repair, and recombination defects (Delmas *et al.*, 2009). RadB, a

paralogue of RadA, is a recombinase mediator protein that assists in RadA nucleoprotein filament formation on ssDNA. Most archaea contain at least one copy of the *radB* gene. Strains deleted for *radB* also show growth, DNA repair, and recombination defects in *H. volcanii*, but to a lesser extent as a *radA* deletion (Guy *et al.*, 2006). Two suppressor mutations that alleviate the Δ*radB* phenotype have been found in RadA (Wardell, 2013; Wardell *et al.*, 2017). These mutations, S101P and A196V, suggest that RadB induces conformational changes in RadA and therefore counteract the Δ*radB* phenotype by promoting the polymerisation of RadA on ssDNA during pre-synapsis (Figure 1.16). Protein:protein interaction studies have shown that RadA and RadB interact *in vivo* in *Pyrococcus furiosus* and *H. volcanii* (Komori *et al.*, 2000; Wardell, 2013).

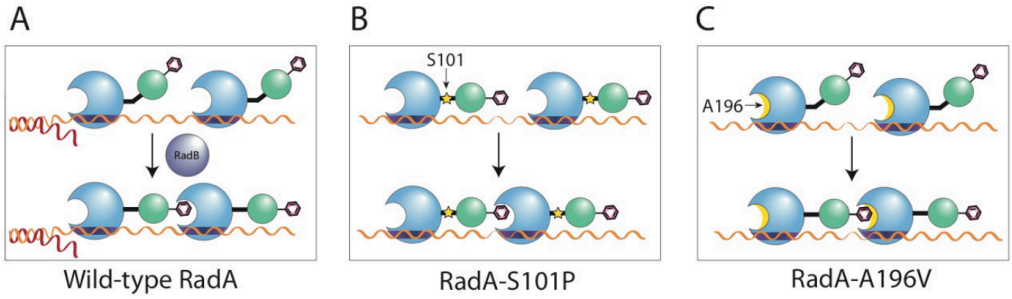


Figure 1.16 **Δ*radB* suppressor mutations in *radA, radA-S101P* and *radA-A196V*.** A) RadB induces a conformational change in wild-type RadA, which is required for efficient nucleoprotein filament formation. B) The RadA-S101P (yellow star) mutation locks RadA into a polymerisation-competent position, negating the need for RadB. C) The RadA-A196V (yellow arc) mutation increases the hydrophobicity of the RadA binding pocket resulting in stronger hydrophobic interactions between RadA monomers, negating the need for RadB. RadA core domain in blue and N-terminal domain in green. Adapted from (Wardell, 2013).

1.5.2 Synapsis

During synapsis, the core reaction of strand exchange occurs. Strand exchange is catalysed by the recombinases RecA in bacteria, Rad51 (repair/fork restart)/DMC1 (during meiosis) in eukaryotes, and RadA in archaea. The recombinase nucleoprotein filament facilitates a physical interaction between the invading ssDNA and the homologous dsDNA template. In eukaryotes, the homology search is catalysed by the Rad51 nucleoprotein filament assisted by Rad54 and Rhd54. Together they enable the sliding of the ssDNA along the homologous dsDNA template. Once homology is found, the recombinase catalyses strand invasion and D-loop formation (Kil, Baitin and Lanzov, 2000).

1.5.3 Post-synapsis

There are several different processing pathways for the recombination intermediates generated by strand exchange. Resolution of homologous recombination generates either a crossover product, where a genetic exchange has occurred, or a non-crossover product, also known as a "gene conversion".

1.5.3.1 Synthesis-dependent strand annealing

Synthesis-dependent strand annealing (SDSA) (Figure 1.17) can occur during both mitotic and meiotic DSB repair in eukaryotes (San Filippo, Sung and Klein, 2008). SDSA produces only non-crossover products, which reduces the likelihood of a genome rearrangement. Conversely, the double strand break repair (DSBR) of resolving recombination intermediates is characterised by the generation of crossover products. It is thought that HR occurs over two waves of repair, the first wave follows the SDSA pathway and the second follows the DSBR pathway.

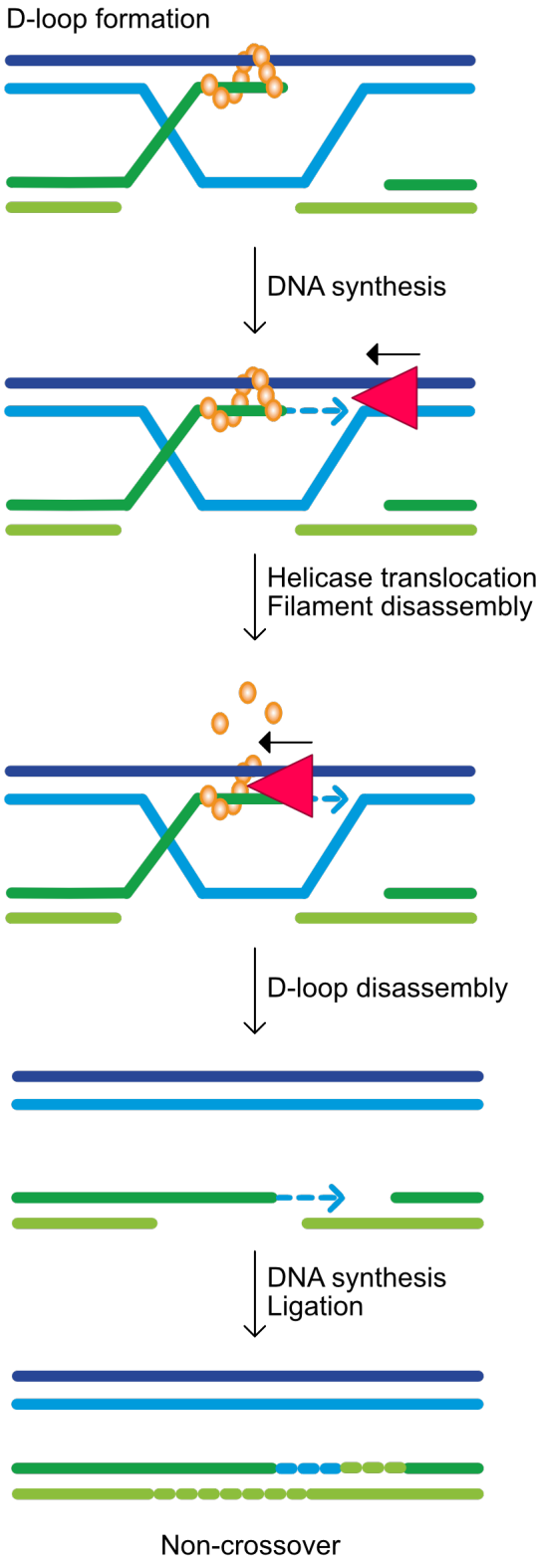


Figure 1.17 **Synthesis-dependent strand annealing (SDSA).** Following strand invasion, DNA is synthesised and helicases (pink triangle) act to disrupt the recombinase nucleoprotein filament (orange spheres) resulting in the disassembly of the D-loop. The gaps are filled and ligated generating a non-crossover product.

During SDSA, helicases act to reverse the extended D-loop; the newly synthesised DNA strand on the first end of the DSB then anneals to the resected strand of the second end of the DSB. The Srs2 helicase in yeast is thought to promote SDSA by dissociating bound Rad51 from the D-loops. It is suggested that the Srs2 helicase activity is directly stimulated by dsDNA bound Rad51 (Pâques and Haber, 1999; Ira *et al.*, 2003; Krejci *et al.*, 2003). The yeast homologue of the human FANCM translocase, Mph1, 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 nucleoprotein filaments, suggesting that these helicases promote the SDSA resolution pathway (Bugreev *et al.*, 2007; Hu *et al.*, 2007; Sommers *et al.*, 2009).

1.5.3.2 Double-strand break repair

The double strand break repair (DSBR) mechanism involves the capture of the second DSB end and the formation of a double Holliday junction (dHJ) (McIlwraith and West, 2008). In eukaryotes, the capture of the second DSB end depends on the Rad52 recombination mediator protein. A double Holliday junction is a branched structure composed of two linked DNA duplexes in a cross-shaped configuration. In order to complete DNA repair, the two DNA molecules require separating. This can be achieved either by endonuclease cleavage or by helicase unwinding, resulting in resolution or dissolution of the Holliday junction respectively.

Holliday junction resolution

Endonucleases known as “resolvases” are specialised to cleave branched DNA structures and are therefore able to resolve Holliday junctions. Cleavage of the dHJ in the same sense results in non-crossover DNA products. Conversely, cleavage of the dHJ in the opposite sense results in crossover DNA products (Figure 1.18).

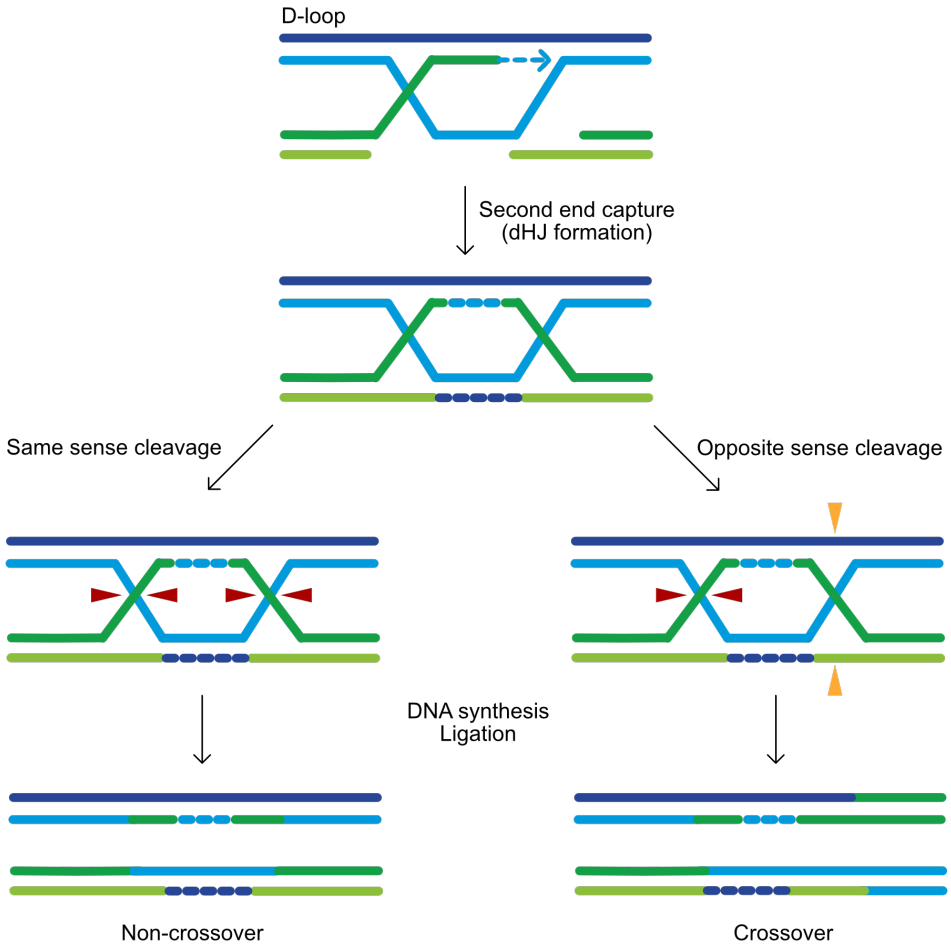


Figure 1.18 **Resolution of double-Holliday junctions (dHJ).** Same sense cleavage by resolvases results in a non-crossover recombination event. Opposite sense cleavage by resolvases results in a crossover recombination event.

Holliday junctions are resolved by the RuvABC complex in *E. coli.* RuvA constrains the junction in a square planar configuration allowing the RuvB helicase to catalyse branch migration. RuvC makes a dual symmetric incision across the Holliday junction intermediate at specific target sequences (Sharples, Ingleston and Lloyd, 1999). Symmetric cleavage results in perfect nicked duplexes that can be directly ligated (West, 1997). RuvA and RuvB homologues are highly conserved throughout bacteria, however RuvC is less conserved. Species lacking RuvC may resolve Holliday junctions using RusA instead. The RecG helicase also performs branch migration, processing of stalled replication forks, and helps direct DNA synthesis during homologous recombination (Singleton, Scaife and Wigley, 2001; Azeroglu *et al.*, 2016); however, RecG is not a resolvase.

Eukaryotic Holliday junction resolution is more complex that the bacterial RuvABC pathway. Multiple Holliday junction resolution pathways are present in eukaryotes and the use of each depends on the species. Many eukaryotic DNA

endonucleases have been proposed to carry out Holliday junction resolvase activity such as Mus81-Mms4 (Eme1), Slx1-Slx4 (BTBD12/Mus312), XPF-ERCC1, Yen1 (GEN1), and Sgs1/MutLγ/Exo1 (Schwartz and Heyer, 2011). The separation and resolution of joined DNA complexes is a complex and multistep process involving sequential nicking of the HR intermediates by endonucleases (Matos *et al.*, 2011).

The archaeal HJ resolving endonuclease, Hjc, has been shown to have analogous resolving properties to the *E. coli* RuvC protein. Similar to RuvC, Hjc cuts Holliday junction symmetrically (Bolt, Lloyd and Sharples, 2001). Some species of archaea contain a second protein that demonstrates Holliday junction resolving activity. For example, the *Sulfolobus solfataricus* Hje protein is able to resolve four-way DNA junctions by introducing a pair of nicks in the stacked-X form of the Holliday junction. Conversely, the *Pyrococcus furiosus* Hjr protein is able to cleave all four arms of a Holliday junction (Kvaratskhelia, Wardleworth and White, 2001). In addition to Hjc, *H. volcanii* contains the Holliday junction nuclease Hef. Neither Hjc or Hef is essential however, when one is deleted the other becomes essential suggesting a redundancy in function. Deletion of *hef* in combination with *radA* is highly deleterious however, the *hjc* and *radA* combined deletion has the same phenotype as the *radA* single deletion. This suggests that Hjc acts exclusively in homologous recombination whereas Hef is able to act in a pathway that bypasses homologous recombination (Lestini, Duan and Allers, 2010).

Asymmetric cleavage

Holliday junctions can also be resolved by asymmetric cleavage of the HR intermediates, this occurs before the double Holliday junction is ligated and results in crossover DNA products (Figure 1.19).

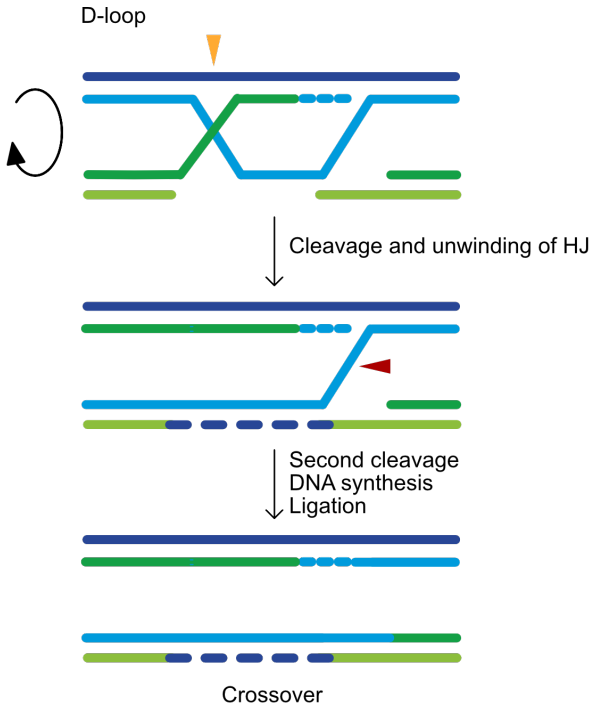


Figure 1.19 **Asymmetric cleavage of D-loops.** Resolvases carry out two sequential cleavage events, the first is indicated by the orange triangle and the second is indicated by the red triangle. The D-loop is unwound between these two cleavage events. Gaps are then filled and ligated generating a cross-over product.

In eukaryotes, the Slx1-Slx4 endonuclease generates a nicked Holliday junction intermediate, which is then targeted by a second endonuclease complex: Mus81-Eme in fission yeast or Mus81-Mms4 in budding yeast (Nishino, Komori, Ishino, *et al.*, 2005; Svendsen *et al.*, 2009; Schwartz and Heyer, 2011). This asymmetric cleavage results in DNA duplex crossover products with gaps that are then filled in by DNA polymerases and ligated.

In archaea, Hef is able to cleave flaps and DNA fork structures. Hef consists of an N-terminal helicase domain and C-terminal nuclease domain linked by a helix-hairpin-helix DNA bind domain and is homologous to FANCM. Hef acts to resolve Holliday junctions by forming homodimers (Komori *et al.*, 2002; Meetei *et al.*, 2005). In *H. volcanii*, Hef is essential in genetic backgrounds where the Holliday junction resolvase Hjc is deleted (Lestini, Duan and Allers, 2010). Hef forms specific localisation foci in the absence of stress, the number of which specifically increase in response to replication arrest, suggesting that Hef stabilises DNA replication by directly interacting with collapsed replication forks (Lestini *et al.*, 2013).

Holliday junction dissolution

An alternative Holliday junction processing method is dissolution. Dissolution of the Holliday junction involves migrating or unwinding the structure using helicases. The resultant DNA hemi-catenanes are then processed by topoisomerases. Unlike Holliday junction resolution, the dissolution pathway generates exclusively non-crossover products and therefore does not rely on endonuclease activity (Figure 1.20) (Schwartz and Heyer, 2011).

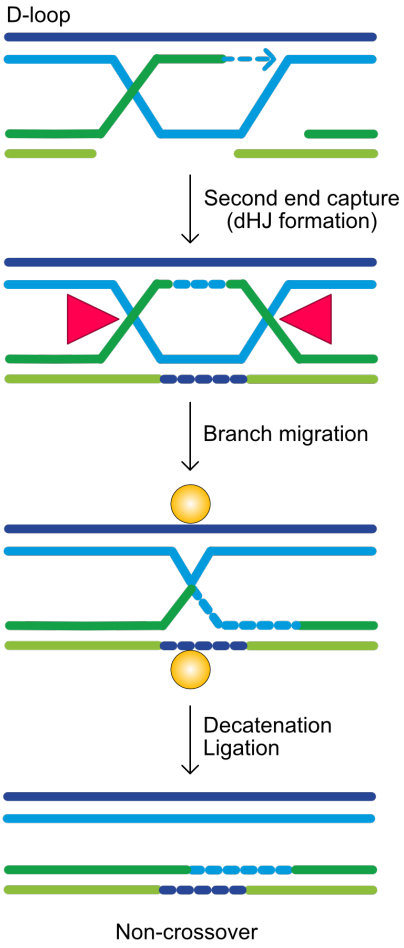


Figure 1.20 **Dissolution of a double Holliday junction.** Two helicases (pink triangles) migrate a double holliday junction forming a hemi-catenane. Decatenation is carried out by topoisomerases generating a non-crossover product.

The DNA hemi-catenanes are formed as a result of HJ migration catalysed by a RecQ family 3'-5' helicases. The RecQ family helicases in *E. coli* is RecQ, in *Saccharomyces cerevisiae* is Sgs1, in *Schizosaccharomyces pombe* is Rqh1, and in humans is BLM (Bizard and Hickson, 2014). Defects in the BLM helicase causes the cancer predisposition syndrome known as "Bloom's syndrome". In humans and *Saccharomyces cerevisiae* the hemi-catenane is then processed by

the TOPOIIIα and Top3 topoisomerases respectively (Fasching *et al.*, 2015). The human Rmi1 and Rmi2 OB-fold proteins stabilise the topoisomerase in an open conformation that favours de-cantenation over relaxation of the DNA substrate (Bachrati and Hickson, 2009; Bocquet *et al.*, 2014).

1.5.4 Regulation of Homologous Recombination

Uncontrolled homologous recombination can lead to aberrant chromosomal rearrangements and therefore tight regulation of HR is paramount. Certain types of damage may be more safely restored via alternative pathways, for example stalled replication forks can be restored using translesion DNA synthesis. Furthermore, the nucleoprotein intermediates generated by HR are toxic in some circumstances, leading to cell cycle arrest and cell death (Krejci *et al.*, 2012).

Negative

Negative regulators of homologous recombination that act to antagonise recombinases, such as Rad51, are known as "anti-recombinases". Anti-recombinases are often DNA helicases or DNA translocases. Examples of anti-recombinases are discussed below.

The *Saccharomyces cerevisiae* Srs2 3'-5' DNA helicase, 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 action prevents untimely or unwanted recombination. The anti-recombinase Srs2 also promotes the SDSA pathway. The mechanism is not yet known but there are two suggested mechanisms. The first is that Srs2 removes Rad51 filaments from D-loops, preventing second end capture by Rad52. The second mechanism involves Srs2 collaborating with nucleases to cleave DNA tails and other intermediate structures (Krejci *et al.*, 2012). Additionally, it is known that Rad51 and Rad55/57 antagonise Srs2 activity (Liu *et al.*, 2011). Whilst no mammalian homologue of Srs2 has been identified, other helicases have demonstrated similar functions. For example RecQ5, BLM, and FANCJ have all shown the ability to disrupt Rad51-ssDNA filaments (Bugreev *et al.*, 2007; Hu *et al.*, 2007; Sommers *et al.*, 2009).

The RecQ helicase Sgs1 in yeast is thought to directly dismantle pre-synaptic filaments, inhibit aberrant invasion events, and resolve recombination intermediates (Mankouri, Craig and Morgan, 2002; Oh *et al.*, 2007). There are several human orthologs of Sgs1 including the cancer-associated helicases BLM, WRN, and RTS. All of these helicases have demonstrated a similar activity to Sgs1 in yeast. The BLM homologue Mus309 in *Drosophila melanogaster* promotes the SDSA pathway by freeing the invading ssDNA from

D-loops and guiding it towards strand annealing (Adams, McVey and Sekelsky, 2003; Bugreev *et al.*, 2007).

HELQ-1, the Hel308 homologue from *Caenorhabditis elegans,* along with RFS-1 have been shown to promote the disassembly of Rad51 filaments from strand invasion intermediates (Ward *et al.*, 2010). The Hel308 helicase and homologues will be discussed in further detail in Section 1.8.

Positive

Recombinase mediators function via several different strategies including promoting and assisting in the assembly of the recombinase nucleoprotein filament on ssDNA, stabilising this structure, promoting recombinase activity, antagonising anti-recombinase activity, or a combination of these.

In mammals the BRCA2, Rad52, Rad54, and Swi5-Sfr2 recombination mediators promote the assembly of Rad51 onto ssDNA and aid in the displacement of RPA (Jensen, Carreira and Kowalczykowski, 2010). In archaea, the recombinase mediator RadB facilitates the loading of RadA onto ssDNA to promote strand invasion and D-loop formation. In *Schizosaccharomyces pombe,* the Shu complex promotes Rad51 function during replication-associated repair and may also function to antagonise Srs2 (Mankouri, Ngo and Hickson, 2007; Krejci *et al.*, 2012). Rad54 and Rdh54/Tid1 are DNA-dependent ATPases that regulate Rad51 by stabilising presynaptic filaments and stimulating Rad51 mediated strand invasion. This in turn promotes D-loop formation and HJ migration (Sung *et al.*, 2003; Mazin *et al.*, 2010).

1.6 Replication Fork Restart

Faithful replication of the genome is essential for accurate transmission of genetic information from generation to generation. Exogenous and endogenous DNA lesions, nicks in the leading strand, and collision with DNA-bound proteins can all hinder the progression of replication by causing the stalling or collapse of replication forks. For example collision with stalled RNA polymerase hinders replication fork progression. Survival of the organism is now dependent on efficient removal of the obstruction and re-establishment of the replication fork. Several pathways are available to re-establish the replication fork including recombination-dependent mechanisms such as break-induced repair (BIR) and Holliday junction-mediated restart, and recombination-independent mechanisms such as fork remodelling and fork resetting mechanisms (Figure 1.21). Recombination-dependent mechanisms of replication fork restart are conservative but complex as they involve the removal and subsequent reloading of replication machinery.

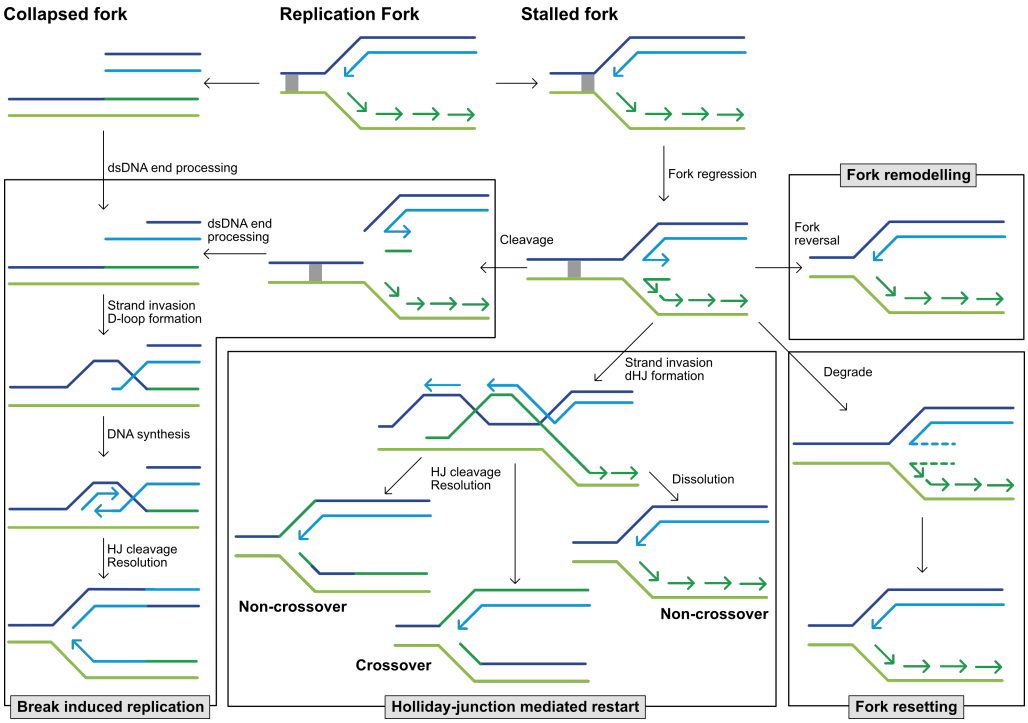


Figure 1.21 **Replication fork restart pathways.** Stalled replication forks can be restarted directly by recombination independent mechanisms such as fork remodelling and fork resetting. Stalled replication forks can also be restarted by recombination-dependent Holliday junction-mediated restart and break induced replication (BIR). Collapsed forks are only repaired via BIR.

1.6.1 Stalled Fork Processing

There are several different pathways to process the stalled fork and reset the active replication fork. Stalled forks can be repaired by either homologous recombination dependent or independent mechanisms. Details of these pathways are discussed below.

Fork Regression

One method of stabilising stalled replication forks is regression of the fork into a Holliday junction. The resultant four-way Holliday junction-like intermediate can be reset into a replication fork by either homologous recombination dependent or independent mechanisms.

The RecG helicase in bacteria facilitates the regression of the replication fork to form a Holliday junction intermediate termed a ‘chicken foot’. Regression of the replication fork allows endonucleases such as RuvC to access and cleave the Holliday junction. RuvAB also demonstrates replication fork regression activity and is able to completely unwind the Holliday junction intermediate (Gupta,

Yeeles and Marians, 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.

In eukaryotes BLM, WRN, and FANCM helicases are all capable of stabilising the stalled replication fork and facilitating the fork regression to form a Holliday junction (Machwe *et al.*, 2007; Gari *et al.*, 2008). Furthermore, BLM may act to repress the formation of abnormal recombination intermediates at stalled replication forks (Ralf, Hickson and Wu, 2006). During the stalling process, long sections of ssDNA are generated. The ssDNA annealing helicase SMARCAL1 assists in stabilising the stalled fork by re-annealing these long sections of ssDNA (Petermann and Helleday, 2010). In *Saccharomyces cerevisiae* the Rad5 DNA helicase demonstrates fork regression activity by unwinding and annealing the nascent and parental strands of replication fork structures (Blastyák *et al.*, 2007).

The Hel308 homologue Hjm from *Sulfolobus tokodaii* displays fork regression activity *in vitro*. Furthermore, Hjm from *Pyrococcus furiosus* demonstrates Holliday junction migration activity *in vitro* suggesting that archaeal Hjm may act analogously to the bacterial RecG, targeting replication forks to form Holliday junctions (Fujikane, Shinagawa and Ishino, 2006; Li *et al.*, 2008). However, the branch migration activity of Hjm is poor compared with the lagging strand or D-loop unwinding activity of Hjm. Hef, the archaeal homologue of FancM, contains an DEAH/D helicase N-terminal domain, which is able to regress the branch point of a replication fork stalled on the leading strand (Horiuchi *et al.*, 2004).

Fork Remodelling

Following regression, the fork can be reversed by helicases to re-establish replication in a recombination-independent manner. Bacterial RecG and RecQ helicases are able to reverse replication fork regression *in vitro,* suggesting that they are capable of resetting replication forks (McGlynn and Lloyd, 2001, 2002; Manosas *et al.*, 2013). The eukaryotic helicases BLM and WRN possess HJ migration activities. In addition to generating regressed replication forks, BLM and WRN may also reverse the formation of Holliday junction-like structures, thus promoting replication fork restart (Constantinou *et al.*, 2000; Karow *et al.*, 2000). The eukaryotic FancM and the archaeal FancM homologue Hef are capable of migrating Holliday junctions and therefore may also act to reverse replication fork regression (Atkinson and McGlynn, 2009; Lestini, Duan and Allers, 2010; Whitby, 2010).

Fork Resetting

Fork resetting involves degradation of the regressed DNA strands to restore the replication fork. In *E. coli,* the RecJ nuclease aided by the RecQ helicase partially degrades nascent DNA strands at the replication fork site (Chow and Courcelle, 2007). In humans, dsDNA at the site of reversed replication forks are degraded in a 5' to 3' polarity by the Dna2 nuclease/helicase and the WRN helicase. In order to prevent excessive nascent strand degradation, RecQ1 inhibits Dna2. It has been suggested that during the repair of replication forks, Dna2 plays a partially redundant role with the exonuclease Exo1 (Karanja *et al.*, 2012; Thangavel *et al.*, 2015).

Holliday junction-mediated resetting

Holliday junction-mediated resetting is one of two recombination-dependent mechanisms that can remodel Holliday junctions into replication forks. During Holliday junction-mediated resetting, the newly regressed DNA strand invades and recombines with the parental duplex DNA. Strand invasion is catalysed by recombinases and forms a double-Holliday junction. Resolution or dissolution of Holliday junctions intermediates, as described in section 1.5.3.2, restores the replication fork.

Break-induced replication

Break-induced replication (BIR) is a homologous recombination-dependent method of restarting replication forks. DNA replication during BIR proceeds via a bubble-like replication fork, resulting in conservative inheritance of the new genetic material. In eukaryotes, the Pif1 helicase promotes DNA synthesis by stimulating the Polδ at D-loops during BIR (Saini *et al.*, 2013; Wilson *et al.*, 2013).

Once a fork has stalled, the replication fork or the regressed replication fork is cleaved by endonucleases generating a one ended DSB. In bacteria, fork regression is carried out by RuvAB or RecG, the RuvC complex then cleaves regressed replication forks (Seigneur *et al.*, 1998; Gupta, Yeeles and Marians, 2014). The archaeal nuclease Hef and HJ-resolvase Hjc are thought to be responsible for generating the one-ended DSB (Komori *et al.*, 2002, 2004; Lestini, Duan and Allers, 2010).

In both instances, the DSB undergoes 5' end resection near the fork junction producing the 3' ssDNA overhang for strand invasion. The eukaryotic endo- and exo-nuclease protein EEPD1, the end resection nuclease Exo1, and the BLM helicase enhance the 5' end resection at stalled replication forks (Wu *et al.*,

2015). Resection of the 5' end prevents the competing NHEJ pathway from repairing the DSB and instead directs the replication fork repair to occur via the homologous recombination-dependent pathway. Avoiding the NHEJ pathway during the repair of replication forks minimises the risk of aberrant fusion of the intermediate dsDNA ends, which is advantageous to the cell.

The homologous recombination enzymes, discussed in detail in Section 1.5, then process the DSB and promote strand exchange. Resolution of the Holliday junction intermediates by the endonuclease Mus81 restores the replication fork in eukaryotes (Mayle *et al.*, 2015). Break-induced replication has a mutation rate a thousand-fold greater than normal DNA replication (Sakofsky, Ayyar and Malkova, 2012).

1.6.2 Replisome Reassembly

Origin-independent reassembly of the *E. coli* replisome is catalysed by the DNA structure-specific factors PriA and PriC. In the presence of large gaps on the leading strand of fork structures, PriC facilitates the loading of the replicative helicase DnaB onto the lagging strand; approximately 20 bp of ssDNA is required for efficient DnaB loading (Heller and Marians, 2005a). The 3'-5' helicases Rep or PriA unwind the nascent lagging strand when initiation of the ultimate Okazaki fragment occurs in close proximity to the replication fork, exposing a sufficient length of ssDNA for PriC-mediated DnaB loading (Heller and Marians, 2005b). Replication can be initiated by a PriA-dependent mechanism if the leading strand is obstructed in close proximity to the fork junction (Heller and Marians, 2005a; Carolina B. Gabbai and Marians, 2010). PriA preferentially acts at fork structures with no gaps on the leading strand, for example D-loop junctions. PriA, PriB, and DnaT act together to load the replicative helicase DnaB onto the displaced strand of the D-loop and reform the replisome. Similar to the PriC-dependent pathway, a leading strand priming event then allows replication to proceed downstream from the original obstruction. Both of these methods of replication restarts result in ssDNA gaps on the leading strand of DNA, which are then repaired by homologous recombination.

Unlike bacteria, eukaryotes and archaea contain multiple origins of replication and therefore, other replication forks on the chromosome can rescue terminally-stalled replication events (Ge, Jackson and Blow, 2007; Newman *et al.*, 2013). In eukaryotes, excess levels of the replicative helicase MCM can license 'dormant' or 'back-up' origins of replication to alleviate replication pressures such as stalled replication forks (Ibarra, Schwob and Méndez, 2008).

1.7 CRISPR-Cas adaptive immunity

The presence of a clustered regularly interspaced short palindromic repeats (CRISPR)-Cas adaptive immune system is conserved throughout the archaea. *H. volcanii* possesses a single type I-B CRISPR-Cas system. CRISPR-Cas immunity proceeds in three stages adaptation, expression, and interference. An invading genetic element releases DNA into the cell which is recognised as foreign and degraded. A degraded fragment of the nucleic acid (known as a protospacer) flanked by a protospacer adjacent motif (PAM), is selected by the Cas1-Cas2 complex and Cas4, and then integrated as a new spacer into the CRISPR locus. This process is known as the "adaptation" stage. During the expression stage, the CRISPR locus is transcribed into a long precursor known as the "pre-crRNA". The endonuclease Cas6 cleaves the pre-crRNA within the repeats generating a pool of CRISPR-derived RNAs (crRNAs), each carrying an individual spacer. The mature crRNAs are incorporated into a multi-protein complex called "Cascade". Cascade comprises Cas5, Cas6, and Cas7, and is important for the stability and maintenance of the crRNAs (Brendel *et al.*, 2014; Maier, Dyall-Smith and Marchfelder, 2015). During CRISPR interference, Cascade detects the PAM sequence and the crRNA then investigates the protospacer. If the cell has previously adapted the protospacer, the crRNA spacer region base pairs with the complementary sequence of the invading DNA. This interaction allows the binding of DNA by the Cascade complex and recruits the Cas3 nuclease for degradation of the invading DNA (Maier *et al.*, 2019) (Figure 1.22).

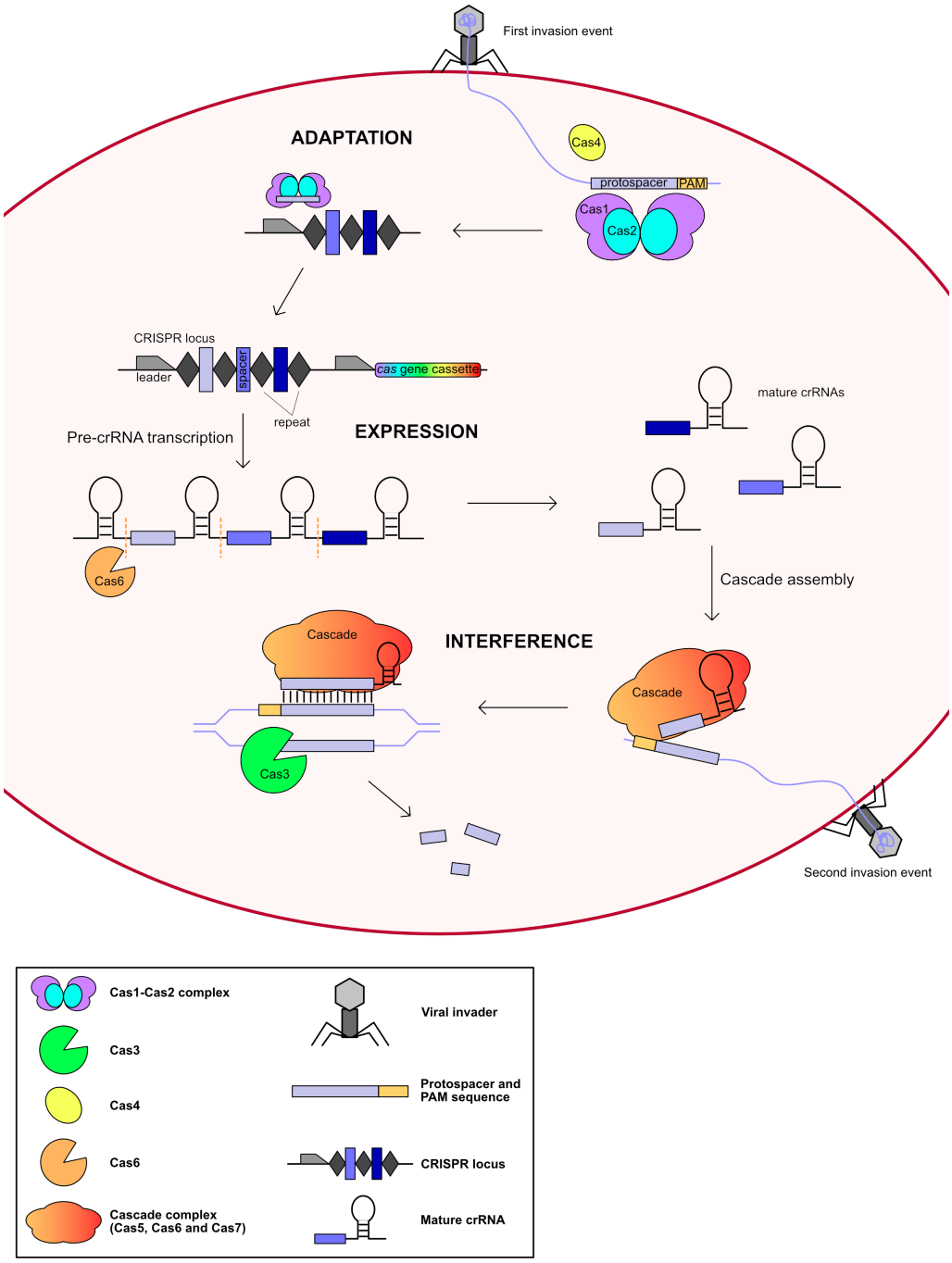


Figure 1.22 **The stages of CRISPR interference in type I systems.** ADAPTATION) Invader 1 releases DNA into the cell which is recognised as foreign and degraded. A protospacer (pale mauve) flanked by PAM sequence (yellow) is selected by the Cas1-Cas2 complex (pink-cyan) and Cas4 (bright yellow), and then integrated as new spacer into the CRISPR locus. EXPRESSION) The CRISPR locus is transcribed into a long precursor known as the pre-crRNA. Cas6 (orange) cleaves the pre-crRNA generating the mature crRNAs. Cascade (orange-red) is assembled incorporating the crRNAs. INTERFERENCE) Cascade detects the PAM sequence and the crRNA then investigates the protospacer. The crRNA spacer region base pairs with the complementary sequence of the invading DNA. Cascade binds DNA and recruits the Cas3 nuclease for degradation of the invading DNA. Figure adapted from (Maier *et al.*, 2019).

A CRISPR interference (CRISPRi) system has been developed for *H. volcanii* utilising these mechanisms. Pre-requisites for CRISPRi include the deletion of genes encoding the nucleases Cas3 and Cas6. By introducing crRNAs with spacers that are complementary to genomic DNA, Cascade is guided to these regions and binds the DNA. Since the *cas3* gene is deleted, the DNA cannot be cleaved and remains intact. Therefore, Cascade acts as a "road-block" by blocking access to the DNA. The *cas6* deletion prevents the maturation of endogenous crRNAs and allows all Cascade complexes in the cell to be loaded with the synthetic crRNA (Maier *et al.*, 2019).

1.8 DNA Helicases

DNA helicases provide a crucial function by generating single-stranded DNA (ssDNA) and RNA (ssRNA) substrates for various stages during nucleic acid metabolism including DNA repair, replication, recombination, transcription, translation, and RNA splicing (Matson, Bean and George, 1994; Chamieh, Ibrahim and Kozah, 2015). This is achieved via the unwinding of complementary duplex strands of nucleic acid, breaking the base-pairing hydrogen bonds, and ultimately leading to the duplex separating. Nucleic acid duplex unwinding is a process dependent on energy derived from the hydrolysis of a nucleoside 5' triphosphate, typically ATP (Hall and Matson, 1999). Helicases load onto and translocate along a nucleic acid single strand in a unidirectional manner, thus can be classified as either 5' to 3' or 3' to 5' helicases (Brosh, 2013).

As a direct result of their essential roles in nucleic acid metabolism it is unsurprising that helicases have been implicated in a multitude of human genetic disorders. The TFIIH transcription factor components XPB and XPD are helicases that play a role in nucleotide excision repair (NER). Mutations within these two helicases have been implicated in the genetic disorders xeroderma pigmentosum, Cockayne syndrome, and trichothiodystrophy (Coin *et al.*, 1999). Cockayne syndrome and trichothiodystrophy are both characterised by premature aging with profound neurological defects. Xeroderma pigmentosum is associated with an extreme predisposition to skin cancer due to an incapability to remove thymidine dimers and 6-4 photoproducts which arise when skin is exposed to high levels of UV light. When the 3' to 5' RecQ family helicases BLM and WRN are mutated, genetic disorders associated with an increased susceptibility to cancers arise (Mohaghegh *et al.*, 2001). Bloom's syndrome is a result of mutations in the helicase BLM. The phenotypes of this disorder include immunodeficiency, facial erythema, short stature, and sub-fertility. Mutations in the WRN helicase however result in Werner's syndrome, which is associated with loss of skin elasticity and subcutaneous fat, and the development of

cataracts. High levels of homologous recombination is observed at a cellular level in Bloom's syndrome patients (German, 1993), whereas Werner's syndrome patients experience illegitimate recombination at a cellular level and a high frequency of chromosomal deletions (Shen and Loeb, 2000). Both disorders result in a high degree of genome instability. Branch migration of Holliday junctions and replication forks is promoted by the FANCM protein, orthologous to the archaeal Hef protein (Meetei *et al.*, 2005). The FANCM protein is composed of an N-terminal helicase domain and a C-terminal nuclease domain. Mutations within the helicase domain has been implicated in Fanconi anaemia, which is typified by developmental defects, bone marrow failure, sensitivity to DNA crosslinking agents, and a predisposition to several cancers (Gari *et al.*, 2008).

1.8.1 Helicase Structure

Six superfamilies of helicases have been defined by sequence similarity between and the organisation of nine signature motifs within the member helicases (Singleton, Dillingham and Wigley, 2007). The nine signature motifs encompass Walker A and Walker B ATPase motifs, which are often annotated as motif I and II respectively in this context (Tuteja and Tuteja, 2004). The full nine conserved motifs are denoted by Q, I, Ia, Ib, II, III, IV, V, and VI (Gorbalenya and Koonin, 1993). These signature motifs often form a 200-700 amino acid cluster known as the 'helicase core'. Within this helicase-core, the Walker A and Walker B boxes are contained within two RecA-like folds (Tuteja and Tuteja, 2004). Walker A and B boxes act to couple the energy generated during ATP hydrolysis to the conformational changes required for the translocation along the nucleic acid strand and ultimately the unwinding of the nucleic acid duplex (Patel and Picha, 2000). It is necessary to note that the presence of these conserved motifs does not ensure an enzyme will have the capability to unwind nucleic acid duplexes (Singleton and Wigley, 2002).

In addition to this motif dependent classification, helicases can also be distinguished based on whether they function monomerically or as a hexameric ring. Superfamily 1 and 2 helicases are typically monomeric with a 3' to 5' directionality. Helicases belonging to superfamilies 3 to 6 however, typically form hexameric rings. The Walker A and B motifs are highly conserved within superfamilies 1 and 2, whilst less conservation is seen of the other seven motifs. Superfamily 3 helicases are composed of only Walker A and B motifs, and a superfamily 3 specific motif III. These helicases are generally found in DNA and RNA viruses (Singleton and Wigley, 2002). Superfamily 4 helicases are bacterial replicative helicases such as DnaB, and possess a 5'- 3' directionality. These helicases also associate with primases in bacteria (Singleton, Dillingham and Wigley, 2007). An example of a superfamily 5 helicase is the bacterial enzyme Rho, which is responsible for the termination of transcription. Rho acts

by unwinding the DNA/RNA hybrid by first binding to and loading onto a specific sequence of nascent RNA (Kaplan and O'Donnell, 2003). MCM complex, the eukaryotic replicative helicase, and RuvB, the helicase responsible for the migration of bacterial Holliday junctions, are both examples of superfamily 6 helicases (Enemark and Joshua-Tor, 2008).

Furthermore, superfamily 3 and 6 helicases are also members of an ATPase superfamily known as the AAA+ family (Neuwald *et al.*, 1999). Many functionally diverse NTPases belong to the vast AAA+ superfamily. The members of this superfamily are characterized by the highly conserved nucleotide-binding and catalytic AAA+ module. The Walker A and Walker B boxes are located within an αβα core domain of this AAA+ module (Snider, Thibault and Houry, 2008).

1.8.2 Mechanism of Helicase Unwinding

Nucleotide strand displacement or unwinding can occur by either active or passive mechanisms. Helicases are defined as passive or active based on whether they primarily stabilise resultant ssDNA or have a direct role in the unwinding of the nucleic acid duplex respectively. Passive unwinding of DNA duplexes occurs as a result of fraying ends; passive helicases will bind the ssDNA stabilising the complex. The ssDNA is then trapped at the subsequent ssDNA/dsDNA junction and unwinding of the duplex commences as the helicase translocates along the fraying strand (Lohman and Bjornson, 1996). There are several models describing active helicase unwinding including the 'inchworm' and the 'active rolling' models.

Inchworm Model

The inchworm model relies on alternative binding of the nucleic acid by the helicase, between two distinguished binding sites, in a coordinated manner. The first site binds ssDNA whilst the second site binds both ssDNA and dsDNA. ATP binding and hydrolysis is required for helicase translocation and nucleic acid unwinding respectively. The resultant conformational changes within the helicase facilitate the 'flipping out' of individual bases within the nucleic acid duplex, this motion is often compared to a 'Mexican wave'. The majority of superfamily 1 and 2 helicases unwind in this manner (Figure 1.23) (Mackintosh and Raney, 2006).

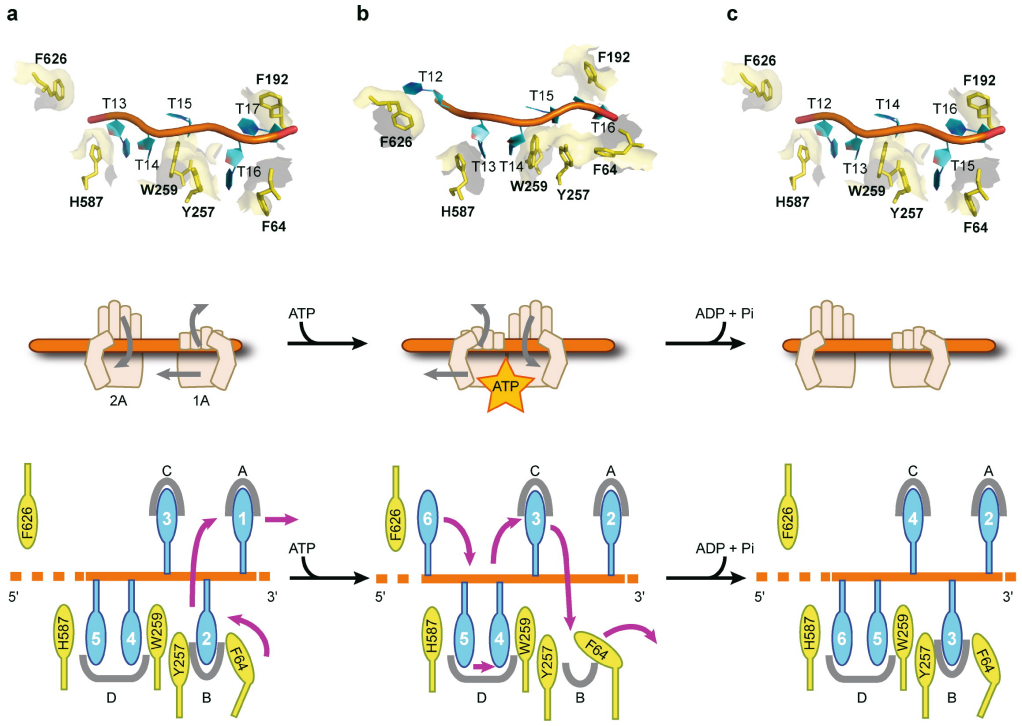


Figure 1.23 **Schematic of the inchworm model of helicase translocation and unwinding.** a) The helicase is bound to the ssDNA tail. b) Upon binding ATP, the helicase undergoes a conformational change, which leads to unwinding of the DNA at the ss/dsDNA junction. c) Following ATP hydrolysis, the helicase returns to its starting conformation. Relative movement of the two RecA-like domains is illustrated by hands (1A and 2A). The open hand represents a loose grip on the DNA and the closed hand represents a tighter grip. During DNA unwinding the conformational changes in the helicase cause the nucleotide bases (blue, numbered 1 to 6 in the 3' to 5' direction) flip between binding pockets of the helicase (lime green residues). Figure adapted from (Singleton, Dillingham and Wigley, 2007).

Co-operative Inchworm Model

The co-operative inchworm model is employed when monomeric helicases encounter a protein block or other such obstacle. In this model, multiple monomers function co-operatively to increase unwinding, thus increasing the likelihood that the obstacle will be overcome. One such example of this is the superfamily T4 phage helicase Dda, though functional as a monomer it was shown that co-operative unwinding was more efficient when a streptavidin block was introduced (Figure 1.24) (Mackintosh and Raney, 2006).

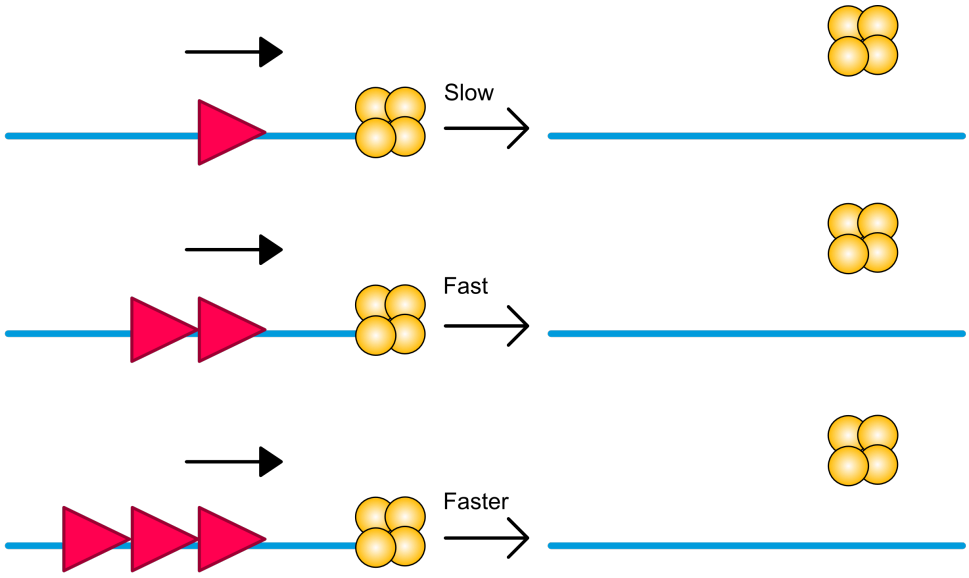


Figure 1.24 **Schematic of the co-operative inchworm model of helicase translocation and unwinding.** Multiple monomeric helicases (pink triangles) assemble along the ssDNA and function together to increase the rate of blockage (yellow circles) clearage. Figure adapted from (Mackintosh and Raney, 2006).

Active Rolling Model

The active rolling model is the proposed mechanism to explain how the *E. coli* Rep helicase unwinds dsDNA. The *E. coli* Rep protein is a stable monomer in the absence of DNA however this model requires the helicase to function in an oligomeric fashion (Korolev *et al.*, 1997). It is proposed that Rep functions as a homodimer where each monomeric subunit has identical nucleic binding sites capable of binding both ssDNA and dsDNA (Soultanas and Wigley, 2001). At any one time, at least one of the subunits is bound to ssDNA ahead of the ss/dsDNA junction. The second subunit is either bound to the same region of ssDNA or it is bound to dsDNA ahead of the replication fork. An ATP dependent isomerisation step enables the two subunits to alternate their affinity for ssDNA and dsDNA. The consequent translocation follows a hand over hand mechanism or rolling of the Rep dimer on the DNA lattice, which is capable of separating multiple base pairs with each Rep subunit binding event (Lohman and Bjornson, 1996) (Figure 1.25).

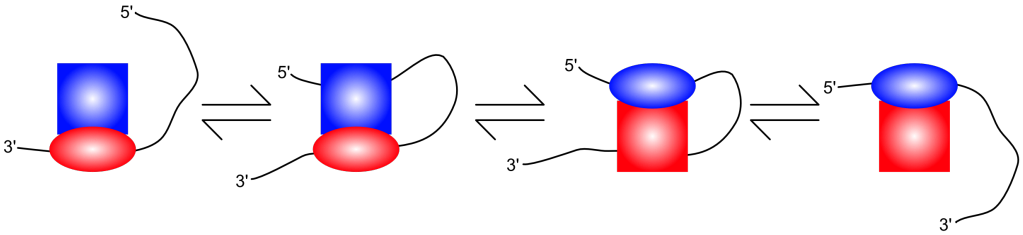


Figure 1.25 **Schematic of the active rolling mechanism of helicase translocation and unwinding.** The dimeric *E. coli* Rep helicase binds to ssDNA with the high affinity subunit (red oval) followed by the low affinity subunit (blue square). One subunit is bound ahead of the ss/dsDNA junction. An ATP isomerisation step switches the affinities of the two Rep subunits. Translocation occurs by rolling of the Rep dimer on the DNA lattice. Figure adapted from (Lohman and Bjornson, 1996).

1.8.3 Hel308

Hel308 is a monomeric 3' to 5' DNA helicase conserved throughout archaea and metazoans however, it is absent from bacteria and fungi. Hel308 is a member of the helicase superfamily 2 (Richards, Johnson, *et al.*, 2008; Woodman and Bolt, 2009). An investigation into *Drosophila melanogaster* mutants, which exhibited hypersensitivity to DNA crosslinking agents, led to the discovery of Hel308 and consequently the human homologue HELQ (Boyd *et al.*, 1981). Dependent on the organism of study, Hel308 is known by other names including HELQ in higher eukaryotes, and Hjm in some archaeal species for example *Pyrococcus furiosus* and *Sulfolobus tokodaii*.

1.8.3.1 Structure of Hel308

Unfortunately, no crystal structure for the *Haloferax volcanii* Hel308 has been obtained to date, there are however resolved crystal structures of three Hel308 homologues. The euryarchaeon *Archaeoglobus fulgidus* Hel308 structure was the first to be resolved both as an apo and a DNA-Hel308 co-crystal (Büttner, Nehring and Hopfner, 2007). In addition to these structures, the structure of Hel308 from the crenarchaeon *Sulfolobus solfataricus*, and the structure of the Hel308 homologue Hjm from the hyperthermophile *Pyrococcus furiosus*, have also been resolved (Richards, Johnson, *et al.*, 2008; Oyama *et al.*, 2009). These crystal structures reveal highly conserved protein folding despite the *Pyrococcus furiosus* Hjm sharing a low amino acid homology of only 30% and 37% with the *Archaeoglobus fulgidus* and *Sulfolobus solfataricus* Hel308 helicases respectively (Oyama *et al.*, 2009).

From these crystal structures, Hel308 was shown to be composed of 5 domains around a central DNA binding core, which is identifiable by a lining of essential DNA binding residues. The 3' ssDNA overhang threads through this DNA

binding core and the five domains drive a β-hairpin between the two strands, subsequently separating the DNA duplex. As with all superfamily 2 helicases Hel308 requires ATP binding and hydrolysis to enable translocation along the DNA; the ATP binding site motifs I, II, V and VI are located within domains 1 and 2 RecA folds. The 'β-hairpin plough' is also located in domain 2. Four α-helices and two parallel β-strands make up a non-canonical winged-helix fold contained within domain 3, thought to be important for the maintenance of the RecA folds structural integrity. This speculation arises from the fact that winged-helix domains are common to nucleic acid binding motifs, and that domains 1 and 3 in Hel308 are packed tightly together (Woodman and Bolt, 2011). Domain 4 is also known as the 'ratchet' domain, it is composed of a seven-helix bundle with the centre α-helix acting as the so-called ratchet (Woodman, Briggs and Bolt, 2007). It is thought that the movement of this ratchet is coupled with ATP hydrolysis via a highly conserved 'FHH' linker which is a part of domain 2 (Gamble-Milner, 2016). Domain 4 together with domains 1 and 3, forms a ring around the ssDNA 3' tail. Finally, domain 5 extrudes from this central pore to bind ssDNA and is able to do so due to its helix-loop-helix structure. Domain 5 is thought to act as an auto-inhibitory domain or 'molecular' brake, in addition to having a role in coupling ATP hydrolysis to the helicase unwinding (Woodman, Briggs and Bolt, 2007; Richards, Johnson, *et al.*, 2008) (Figure 1.26).

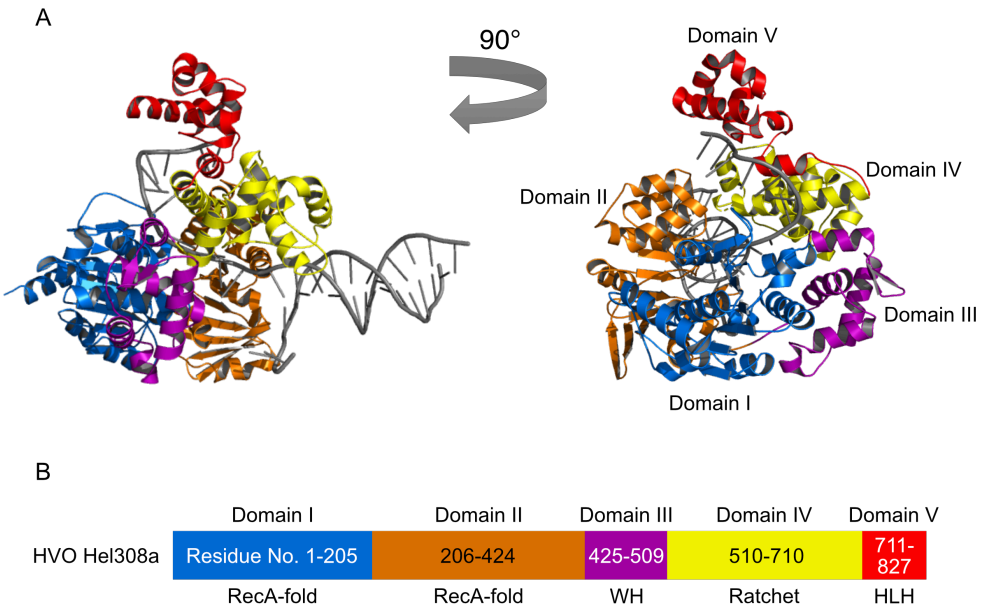


Figure 1.26 ***Archaeoglobus fulgidus* Hel308-DNA co-crystal.** A) Hel308 in complex with a 15 base pair DNA duplex and 10 base single stranded 3' tail. The five domains of Hel308 forms a central pore around ssDNA. B) Linear schematic of the structural domains of *H. volcanii* Hel308. Structure from Protein Data Bank 2P6R (Buttner et al 2007).

Expression information is available at the transcriptomic level for the wildtype strain. The data reveals that Hel308a (HVO_0014) has a transcription level of 430 rkpm found within the medium expression level (Quartile 3) of the dataset.

This is unsurprising as high levels of any helicase can disrupt nucleic acid processing and result in genome instability.

1.8.3.2 Biochemical analysis of Hel308

Both archaeal Hel308 and the human homologue HELQ demonstrate 3' to 5' translocation along ssDNA *in vitro* and have ATPase activity stimulated by ssDNA but not dsDNA (Marini and Wood, 2002; Guy and Bolt, 2005). The Hel308 homologue in *Pyrococcus furiosus* was shown to unwind and migrate Holliday junctions therefore, was named Holliday junction migration (Hjm). Due to this capability it was suggested that Hel308 had a role in the later stages of homologous recombination (Fujikane *et al.*, 2005).

Analysis into a variety of replication fork structures was undertaken to determine specificity of the *Methanothermobacter thermautotrophicus* Hel308, and whether or not the protein has a preferred fork construct for unwinding. The data showed that from a branch point, Hel308 exhibited preference for displacing the lagging (3' overhang) strand structures over leading (5' overhang) strand structures, however the enzyme was unable to unwind parental duplex DNA (Guy and Bolt, 2005). DNA fork substrates require a parental duplex with either a strand nick or a ss-dsDNA junction, for targeted unwinding by Hel308. This is most efficiently achieved when Hel308 engages the fork branch point.

In support of this, the Hel308 homologue Hjm in *Sulfolobus tokodaii* also demonstrated preference for the binding and unwinding of the replication fork nascent strand at a ss-dsDNA junction. However, in contrast with *Methanothermobacter thermautotrophicus* Hel308, *Sulfolobus tokodaii* Hjm has been shown to unwind both leading and lagging strand structures (Fujii *et al.*, 2002; Li *et al.*, 2008). The evolutional divergence between Euryarchaeota and Crenarchaeota may explain the differences between Hjm and Hel308. These differences could elude to divergent cellular functions of these two enzymes within their respective major archaeal subdomains.

Hel308 is also able to dissociate the 'invading' strand in D-loop structures (Guy and Bolt, 2005). In order to unwind replication fork-like structures, Hel308 requires the 3' ssDNA tail to be a minimal length of between 7 and 15 nucleotides (Tafel, Wu and McHugh, 2011). This ability to unwind fork structures with lagging strands suggests that Hel308 may act on replication forks that are damaged and as a consequence DNA replication stalling events on the leading strand template. The *Sulfolobus tokodaii* Hjm has also been proposed to remodel replication forks into 'chicken foot' structures, implicating Hel308 and Hjm homologues in the remodelling of replication forks after stalling events (Li *et al.*, 2008).

In support of this, a GFP recombinant version of the human Hel308 homologue, HELQ, has been shown to localise to stalled replication forks that were generated by treatment with camptothecin. Co-localisation with replication protein A (RPA), Rad51 foci, and FANCD2 was also observed. Rad51 is a RecA family recombinase and FANCD2 is implicated in the promotion of fork repair via homologous recombination (Taniguchi *et al.*, 2002). The data indicate that Hel308 is recruited to stalled replication forks following DNA damage and has a role in the recombination-mediated remodelling of these stalled forks, with the aim to restart replication.

1.8.3.3 Hel308 protein: protein Interactions

Immobilised Hel308 from *Methanothermobacter thermautotrophicus* was shown to interact *in vitro* with purified replication protein A (RPA). Mutational analysis confirmed the interaction is between the C-terminus of Hel308 and RPA (Woodman, Brammer and Bolt, 2011). Immunoprecipitation assays using the mouse HELQ homologue also revealed an interaction with RPA (Adelman *et al.*, 2013). Replication protein A (RPA) is a eukaryotic and archaeal single stranded DNA binding protein, which binds to ssDNA to protect it from degradation and secondary structure formation during DNA replication. Furthermore, RPA has been shown to stimulate the helicase activity of human HELQ directly (Tafel, Wu and McHugh, 2011). The co-localisation, interaction, and direct stimulation of helicase activity between RPA and Hel308 homologues has led to the proposal that RPA may act as a platform to recruit Hel308 to stalled replication forks in archaea.

Multiple assays including gel filtration, affinity pulldown, and yeast two-hybrid analyses in *Sulfolobus tokodaii* have revealed an *in vitro* interaction between the Hel308 homologue Hjm and the Holliday junction resolvase Hjc (Li *et al.*, 2008). Hjc generates two recombinant duplexes upon specific binding of a Holliday junction and symmetrical cleaving of the two opposing strands, this action has been associated with the restart of stalled replication forks (Lestini, Duan and Allers, 2010). The formation of these Hjc-Holliday junction complexes was prevented by the presence of Hjm, suggesting a regulatory role for this Hel308 homologue on the endonuclease activity of Hjc (Hong *et al.*, 2012). The interaction between Hjm and Hjc implicates Hel308 in homologous recombination, however neither the mechanism nor the function of this interaction is known.

Human HELQ has been shown to associate with the RAD51 paralogues RAD51B, RAD51C, RAD51D, and XRCC2, collectively known as the BCDX2 complex. The BCDX2 complex is required for homologous recombination. HELQ was also shown to associate *in vivo* with the DNA damage responsive kinase ATR. ATR is activated in response to persistent ssDNA, a signalling

pathway is subsequently triggered leading to the promotion of homologous recombination (Takata *et al.*, 2013). Furthermore, reciprocal immunoprecipitation assays using mouse HELQ identified both the BCDX2 complex and ATR as interaction partners of HELQ (Adelman *et al.*, 2013). Since BCDX2 promotes replication-coupled homologous recombination, these results suggest that HELQ has a role in the restart of stalled replication forks. The deletion of the Rad51 paralog *rfs-1* in *Caenorhabditis elegans* in combination with *helq-1* leads to a synthetic lethality (Taylor *et al.*, 2015). It is proposed that this double mutation causes a mitotic defect as a consequence of the inability to disassemble RAD51 from the strand invasion intermediates. This was corroborated using purified HELQ-1 and RFS-1, which are both shown to independently bind and promote the disassembly of RAD51 from double stranded, but not single stranded DNA filaments *in vitro* (Ward *et al.*, 2010; McClendon *et al.*, 2016). This indicates that HELQ and RFS-1 have distinct but overlapping roles in homologous recombination.

In *Pyrococcus furiosus*, the proliferating cell nuclear antigen (PCNA) was shown to co-precipitate with the Hel308 homologue Hjm during immunoprecipitation analyses *in vivo*. The co-precipitation was a direct consequence of a C-terminal interaction of Hjm with PCNA. Further analyses into this interaction revealed that PCNA stimulates Hjm helicase activity at DNA fork structures. PCNA anchors DNA polymerase to the template DNA in order to prevent dissociation, and so is known as the ‘sliding clamp’ (Fujikane, Shinagawa and Ishino, 2006).

1.8.3.4 Genetic analysis of Hel308

The deletion of *hel308* is viable in almost every organism in which the gene is found, demonstrating that the protein Hel308 is not essential. An exception to this generalisation is Hel308 in the organism *Sulfolobus islandicus* where it does appear to be essential (Hong *et al.*, 2012; Zhang *et al.*, 2013). In *Haloferax volcanii*, Hel308 deletion mutants are slow growing and show sensitivity to DNA crosslinking agents such as mitomycin C (MMC).

In *Haloferax volcanii*, mutants deleted for the RadA paralogue *radB* exhibit a decrease in the level of homologous recombination to 10% of that in wild type cells. However, a double deletion of *hel308* with *radB* restores, in part, homologous recombination. The data suggest an anti-recombinase role for Hel308 and fits with its role as a guardian of genome stability (Gamble-Milner, 2016).

In *Drosophila melanogaster,* Hel308 deletion mutants show sensitivity to cisplatin, a DNA crosslinking agent and potent inhibitor of DNA replication (Boyd, Sakaguchi and Harris, 1990). The same phenotype is observed in human

cells deleted for HELQ upon MMC treatment (Takata *et al.*, 2013). Overall these data suggest that Hel308 functions to process and repair DNA damaged via crosslinking.

HELQ deficient mice exhibit subfertility, germ cell attrition, a predisposition to tumours, and like human cells, sensitivity to DNA crosslinking agents (Adelman *et al.*, 2013). A similar phenotype, albeit more extreme, was observed in Fanconi anaemia mouse models. Therefore, a double mutant of *Helq* and *Fancd2* was generated in mice to investigate these phenotypic similarities (Parmar, D'Andrea and Niedernhofer, 2009). Double mutant cells showed a greater sensitivity than either of the single mutants following treatment with replication blocking agents such as MMC and camptothecin (Parmar, D'Andrea and Niedernhofer, 2009). There was also a significant increase in spontaneous and MMC-induced chromosomal aberrations, compared with the single *Helq* mutant. These data suggest that HELQ and FANCD2 act in parallel pathways of DNA interstrand cross link repair. Similar findings in human cell FANCD2 RNAi studies, FANCC deleted mice models, and *Caenorhabditis elegans* corroborate this parallel action (Muzzini *et al.*, 2008; Luebben *et al.*, 2013; Takata *et al.*, 2013). HELQ acts in a distinct pathway to the FA complex, but immunoprecipitation assays in mice revealed that HELQ physically interacts with the FANCD2-FANCI heterodimer, suggesting the two pathways may be regulated by a physical interaction (Adelman *et al.*, 2013).

Hel308 has been implicated in genome stability upon the identification of several certain Hel308 single nucleotide polymorphisms (SNPs) being more prevalent in numerous types of head and neck cancers, including esophageal squamous cell carcinoma and gastric adenocarcinoma (Li *et al.*, 2013; Liang *et al.*, 2013; Babron *et al.*, 2014).

Summary

In addition to interacting with replication fork maintenance proteins such as RPA and PCNA, Hel308 interacts with replication fork-like structures *in vitro* and can also resolve D-loop structures. Hel308 has been shown to associate with various proteins involved in homologous recombination such as RAD51, RadA, and the Holliday junction resolvase, Hjc. Furthermore, the Hel308 homologue Hjm has been shown to migrate Holliday junctions, affirming a role in homologous regulation. Though clearly implicated in the restart of stalled replication forks and homologous recombination, the exact function and mechanism of Hel308 remains elusive.

Chapter 2: Materials and Methods

2.1 Materials

All strains and plasmids were constructed by the author unless stated otherwise.

2.1.1 Strains

Table 2.1 ***Haloferax volcanii*** **strains used in this study.**

Parent: H26 (*ΔpyrE2*)			
Strain	**Parent**	**Genotype**	**Notes**
H178	H158	*Δhjc*	Constructed by GN
H358	H338	Δ*hef*	Constructed by ZD
H1391	H1379	Δ*hel308*	Constructed by TA

Parent: H53 (*ΔpyrE2, ΔtrpA*)			
Strain	**Parent**	**Genotype**	**Notes**
H1467	H1427	Δ*hjc,* Δ*hel308::trpA+*	Constructed by TA
H1468	H1427	Δ*hef,* Δ*hel308::trpA+*	Constructed by TA
H2085	H2071	*Δhel308*	Constructed by DW
H3515	H3281	*Δhel308 ΔradA::trpA+*	
H4249	H4248	Δ*pinA::trpA+*	
H4363	H4362	Δ*pinA*	
H4441	H4440	Δ*pinA,* Δ*hef::trpA+*	
H4443	H4442	Δ*pinA,* Δ*hel308::trpA+*	

Parent: H164 (*ΔpyrE2, bgaHa-Bb, leuB-Ag1, ΔtrpA*)			
Strain	**Parent**	**Genotype**	**Notes**
H2117	H2098	Δ*hel308::trpA+*	Constructed by RG-M
H2397	H2356	*hel308-F316A*	Constructed by RG-M
H2398	H2357	*hel308-R743A*	Constructed by RG-M
H2400	H2213	*hel308-D145N*	Constructed by RG-M
H2417	H2369	Δ*hel308::trpA+,* Δ*radB*	Constructed by RG-M
H3826	H3835	*hel308-D145N-F316A*	
H3827	H3833	*hel308-D145N-R743A*	
H3929	H3901	Δ*recJ1::trpA+*	
H3931	H3905	Δ*recJ3::trpA+*	
H3932	H3907	Δ*recJ4::trpA+*	
H4099	H4092	Δ*recJ1*	
H4103	H4095	Δ*recJ3*	
H4105	H4097	Δ*recJ4*	
H4170	H4164	Δ*recJ1,* Δ*recJ3::trpA+*	

H4173	H4167	Δ*recJ3, ΔrecJ4::trpA+*
H4174	H4168	Δ*recJ4, ΔrecJ1::trpA+*
H4181	H1580	Δ*hdrB, Δhel308*
H4190	H4176	Δ*recJ1, ΔrecJ3*
H4192	H4179	Δ*recJ3, ΔrecJ4*
H4193	H4180	Δ*recJ4, ΔrecJ1*
H4194	H4181	Δ*hdrB, Δhel308 {p.tnaM3::7xhis-StrepII-StrepII-()+ pyrE2+ hdrB+}*
H4195	H4181	Δ*hdrB, Δhel308 {p.tnaM3::7xhis-StrepII-StrepII-hel308+ pyrE2+ hdrB+}*
H4196	H4181	Δ*hdrB, Δhel308 {p.tnaM3::7xhis-StrepII-StrepII-hel308-D145N+ pyrE2+ hdrB+}*
H4273	H4268	Δ*recJ1, ΔrecJ3, ΔrecJ4::trpA+*
H4275	H4270	Δ*recJ3, ΔrecJ4,ΔrecJ1::trpA+*
H4376	H4367	Δ*recJ1, ΔrecJ3, ΔrecJ4*
H4386	H4380	Δ*cas3::trpA+*
H4388	H4382	*hel308-F316A, Δcas3::trpA+*
H4389	H4383	*hel308-R743A, Δcas3::trpA+*
H4445	H4444	Δ*hel308::trpA+, Δcas3*
H4447	H4446	*hel308-D145N, Δcas3::trpA+*
H4448	H4181	Δ*hdrB, Δhel308 {p.tnaM3::hel308-K53A+ pyrE2+ hdrB+}*

Parent: H2047 (*ΔpyrE2, Nph-pitA, Δmrr, cdc48d-Ct, ΔtrpA*)			
Strain	**Parent**	**Genotype**	**Notes**
H2131	H2130	*Δhel308::trpA+, pyrE2+*	Constructed by RG-M
H2418	H2372	*his6 tag-hel308-strepII tag*	Constructed by RG-M
H3360	H3045	*hel308-strepII tag*	
H3362	H3047	*his6 tag-hel308*	
H3363	H3048	*strepII tag-hel308-his6 tag*	
H3365	H3050	*his7-strepII-strepII-tag-hel308*	
H3367	H3052	*strepII-strepII tag-hel308*	
H3369	H3054	*hel308-his6 tag*	
H3371	H3056	*strepII tag-hel308 tag*	

DW = David Wallbank, RG-M = Rebecca Gamble-Milner, GN = Greg (Hien-Ping) Ngo, ZD= Zhenhong Duan,

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

Strain	Genotype	Notes
XL1-Blue MRF'	*endA1, gyrA96* (Na1R), *lac* [F' proAB *lacIqZΔM15 Tn10* (TetR)], *Δ(mcrA)183, Δ(mcrCB-hsdSMR-mrr)173, recA1, relA1, supE44, thi-1*	Standard cloning strain enabling blue/white selection in conjunction with pBluescriptII SK+ plasmid derivatives. 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*	For preparation of unmethylated DNA used in *Haloferax volcanii* transformations. *dam-, dcm-*. From (Allers *et al.* 2004) M.G. Marinus via R.G. Lloyd.

2.1.2 Plasmids

All plasmids contain the ampicillin resistance gene, AmpR, to allow for selection during cloning in *E. coli.* All gene deletion constructs contain the *pyrE2* gene to allow for pop-in selection in *H. volcanii.*

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

Name	Use	Notes
pBluescriptII SK+	Standard *E. coli* vector with functional blue/white screening capability	From Stratagene
pTA131	For making deletions in a *ΔpyrE2* background	Constructed by TA (Allers *et al.* 2004)
pTA163	Integrative vector containing *leuB-Ag2* and *pyrE2*. For use in recombination assays.	Constructed by GN
pTA298	*trpA* genomic clone with flanking *Bam*HI sites	Constructed by TA
pTA324	*radA::trpA+* deletion construct.	Constructed by TA
pTA370	*hef* deletion construct	Constructed by ZD
pTA635	Episomal plasmid for *in trans* expression of *radA*	Constructed by TA
pTA1062	Δ*hjc::trpA+*	Constructed by RL
pTA1064	Δ*hef::trpA+*	Constructed by RL
pTA1273	*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-D145N* gene replacement construct	Constructed by TA
pTA1425	N-terminally his6-tagged and a C-terminally StrepII-tagged *hel308* gene replacement construct, under the control of p.tnaA	Constructed by RG-M
pTA1591	*hjc* overexpression construct	Constructed by DW
pTA1592	*hef* overexpression construct	Constructed by DW
pTA1642	*hel308-F316A* gene replacement construct	Constructed by RG-M
pTA1643	*hel308-R743A* gene replacement construct	Constructed by RG-M
pTA1647	*hel308-F316A* gene replacement construct	Constructed by RG-M
pTA1648	*hel308-R743A* gene replacement construct	Constructed by RG-M
pTA1661	Intermediate construct for creating pTA1662	Constructed by RG-M
pTA1662	N-terminally his6-tagged and a C-terminally StrepII-tagged *hel308* gene replacement construct	Constructed by RG-M
pTA1716	*ginS* genomic clone	Constructed by HM
pTA1718	C-terminally StrepII-tagged hel308 gene replacement construct	
pTA1719	N-terminally his6-tagged hel308 gene replacement construct	
pTA1771	Intermediate construct used to create pTA1849	Constructed by TA
pTA1797	Intermediate construct used to create pTA1850	Constructed by TA
pTA1840	*radA* overexpression construct used to generate probe for Southern hybridization	Constructed by DA
pTA1849	Intermediate construct used to create pTA1857	
pTA1850	Intermediate construct used to create pTA1857	
pTA1855	N-terminally StrepII-tagged and a C-terminally His6-tagged *hel308* gene replacement construct	
pTA1856	N-terminally His7-2xStrepII-tagged *hel308* gene replacement construct	
pTA1857	N-terminally 2xStrepII-tagged *hel308* gene replacement construct	

pTA1859	C-terminally His6-tagged *hel308* gene replacement construct
pTA1860	N-terminally StrepII-tagged *hel308* gene replacement construct
pTA1882	*recJ4* genomic clone
pTA1887	*recJ1* genomic clone
pTA1888	*recJ2* genomic clone
pTA1889	*recJ3* genomic clone
pTA1905	*recJ2* genomic clone with *pyrE2* marker
pTA1912	*recJ1* genomic clone with *pyrE2* marker
pTA1913	*recJ3* genomic clone with *pyrE2* marker
pTA1921	*pinA* genomic clone
pTA1923	*pinA* deletion construct
pTA1924	*recJ1* deletion construct
pTA1927	*recJ3* deletion construct
pTA1945	*hel308-K53R-F316A* gene replacement construct
pTA1946	*hel308-K53R-R743A* gene replacement construct
pTA1947	*hel308-D145N-R743A* gene replacement construct
pTA1951	*recJ2* deletion construct
pTA1952	*hel308-D145N-F316A* gene replacement construct
pTA1958	*recJ1::trpA+* deletion construct
pTA1960	*recJ3::trpA+* deletion construct
pTA1970	*pinA::trpA+* deletion construct
pTA1988	*hel308-K53A* gene replacement construct
pTA1989	*hel308-K53A-F316A* gene replacement construct
pTA1990	*hel308-K53A-R743A* gene replacement construct
pTA1991	*recJ4* deletion construct
pTA1996	*recJ2::trpA+* deletion construct
pTA1997	*recJ4::trpA+* deletion construct
pTA2010	*hel308-D145A* gene replacement construct
pTA2080	Episomal plasmid for *in trans* expression of *radA-S101P*
pTA2082	Intermediate construct for creating pTA2090
pTA2083	Intermediate construct for creating pTA2091

pTA2084	Intermediate construct for creating pTA2130	
pTA2085	Intermediate construct for creating pTA2095	
pTA2090	N-terminally his6-tagged and a C-terminally StrepII-tagged recJ1 gene replacement construct	
pTA2091	N-terminally his6-tagged and a C-terminally StrepII-tagged recJ2 gene replacement construct	
pTA2095	N-terminally his6-tagged and a C-terminally StrepII-tagged recJ4 gene replacement construct	
pTA2096	Intermediate construct for creating pTA2118	
pTA2097	Intermediate construct for creating pTA2127	
pTA2112	Intermediate construct for creating pTA2122	
pTA2118	p.tnaM3-His7-2xStrepII expression vector	
pTA2121	p.tnaM3-His7-2xStrepII-Hel308 expression vector	
pTA2122	p.tnaM3-His7-2xStrepII-Hel308-D145N expression vector	
pTA2127	p.syn-His7-2xStrepII expression vector	
pTA2130	Intermediate construct for creating N-terminally his6-tagged and a C-terminally StrepII-tagged recJ3 gene replacement construct	
pTA2184	Intermediate construct for creating pTA2315	
pTA2200	*p.tnaM3-hel308-K53A* expression vector	
pTA2201	p.tnaM3-His7-2xStrepII-Hel308-F316A expression vector	
pTA2315	*ginS* partial deletion construct maintaining ~100bp C-terminus (suspected promoter for *bcp1*)	
pTA2335	Δ*ginS::trpA+* partial deletion construct maintaining ~100bp C-terminus (suspected promoter for *bcp1*)	
p131-Cas3up/do	Δ*cas3*	Anita Marchfelder

pTA131-iCas3KO-trpA	Δ*cas3::trpA+*	Anita Marchfelder
pTA927-cas3FLAGN	*FLAG-cas3* expression construct used to generate radiolabelled probe	Anita Marchfelder

TA = Thorsten Allers, GN = Greg Ngo, ZD = Zhenhong Duan, RL = Roxane Lestini, DW = David Wallbank, RG-M = Rebecca Gamble-Milner, DA = Darya Ausiannikava, HM = Hannah Marriott.

2.1.3 Oligonucleotides

Table 2.4 **Oligonucleotides used in this study.**

Name	Sequence (5'-3')	Description	Used to construct
pBSF2	TTAAGTTGGGTAACGCCAGGG	Amplification of the upstream region to multiple genes: *recJ2, recJ4;* and *hel308a*. Sequencing primer	pTA1923, pTA1951, pTA1991, K53A plasmids
pBSR3	ACCCCAGGCTTTACACTTTATGC	Amplification of the DS region to multiple genes: *recJ2, recJ4*. Sequencing primer	pTA1923, pTA1951, pTA1991
Hel308 NdeI_F	GGGAGATcatATGCGAACTGCGGACCTGACGGGC	Incorporation of a *Nde*I site upstream of h*el308a* start codon: to clone downstream of a tryptophanase promoter	pTA1854
Hel308 SphI_F	GGGAGATGgcATGCGAACTGCGGACCTGACGGGC	Incorporation of a *Sph*I site upstream of h*el308a* start codon: to clone downstream of a tryptophanase promoter	pTA1849, pTA1850
helQ(R)BamHI02	CGAGGAtCCTCATTCGAAATCACCCAGACTGG	Amplification of *hel308a* and introduction of a *Bam*HI site	pTA1849, pTA1850, pTA1854
oriC-1trp-R	CGGCCGCTCTAGAACTAGTGG	Amplification of 2626bp region of pTA1425 containing *StrepII-hel308-6His* CDS	pTA1855
PtnaAFint	GCCTGCCGATTACTTCACATTCGC	Amplification of 2626bp region of pTA1425 containing *StrepII-hel308-6His* CDS	pTA1855
hel308 EcoR	AGGTAGTCGAGCACGCGGTCC	Amplification of internal region of *hel308a*	pTA1952, pTA1965
Hel308 Fint	AGCGCTGGGAGGAGTACGGC	Amplification of internal region of *hel308a*	pTA1952, pTA1965
Hel308 F316A F	CGCGGCGgcCCACCACGCGGGACTCGCCG	Generation of a *hel308-F316A* mutation	pTA1952, pTA1965

Hel308 F316A R	CGTGGTGGgcCGC CGCGCCTTTGGC GACC	Generation of a *hel308-F316A* mutation	pTA1952, pTA1965
Hel308 RTR	GTGTTGACGCCG GCGGCGAGCG	Amplification of internal region of *hel308a*	K53A plasmids
Hel308 K53AF	CGAGCGGgccGAC GCTCATCGCCGA ACTGGCG	Generation of a *hel308-K53A* mutation	K53A plasmids
Hel308 K53AR	CGATGAGCGTCg gcCCGCTCGCCGT CGGG	Generation of a *hel308-K53A* mutation	K53A plasmids
PinA_ F	CGAGTAGTTCGT CGAGCCCTCGG	Amplification of 5861 bp region of chromosome containing *pinA*	pTA1921
PinA_ R	CGTCGAGGAAGT AGAAGTCGCGGT CG	Amplification of 5861 bp region of chromosome containing *pinA*	pTA1921
delpin ABglF	CGCAACaGaTCTC TTACGACACCGA CTGC	Forward inverse PCR primer to delete *pinA*	pTA1923
delpin ANdeR 2	CCGGCACTATCT TCATatGCGTGAG TTTGGGCGGG	Reverse inverse PCR primer to delete *pinA*	pTA1923
RecJ1 NcoI_F	CGTGAGTACCAC ACCcATGGACGG ACCCGTCCCC	Incorporation of an *Nco*I site at the *recJ1* start codon	pTA2082
RecJ1 NheI_ R	CGGTCGACGAgct agCCGCGTTTTCA GCCAACTGC	Incorporation of an *Nhe*I at the *recJ1* stop codon	pTA2082
RecJ2P ciI_F	CCTCGAAGAATA ACTacATGTCCGT GAGCCCCGC	Incorporation of a *Pci*I site at the *recJ2* start codon	pTA2083
RecJ2 NheI_ R	GGGTCGCACGCC GGCgCtagcGCGCA CCTCCC	Incorporation of an *Nhe*I at the *recJ2* stop codon	pTA2083
RecJ3 BspHI _F	CCTAGCGGGAtcA TGAGCGACGAGC ACGCCGGGG	Incorporation of a *Bsp*HI site at the *recJ3* start codon	pTA2084
RecJ3 NheI_ R	CGCCGGCgctaGC CGTCGTCGACAG CTCTTCG	Incorporation of an *Nhe*I at the *recJ3* stop codon	pTA2084
RecJ4 NcoI_F	GCTCAACGCcAT GGATTGGATTAC GCACGAGGAGG	Incorporation of an *Nco*I site at the *recJ4* start codon	pTA2085
RecJ4 NheI_ R	GGATTCGctagcGA ACTGCTCGGCGG CGGCGTCG	Incorporation of an *Nhe*I at the *recJ4* stop codon	pTA2085
RecJ4p robeF	CGAGTACCACTA CTTCACCCGTCG CCCGC	Amplification of internal region of *recJ4*	RecJ4 colony hyb probe

RecJ4probeR1	CCTGATACGACTGGTAGTAGGCTTCC	Amplification of internal region of *recJ4*	RecJ4 colony hyb probe
RecJ2probeF	GCACGACACCGAAGGAAGACACCGACC	Amplification of internal region of *recJ2*	RecJ2 colony hyb probe
RecJ2probeR	GGCGAGGACAGGTCACAGGAGAGCG	Amplification of internal region of *recJ2*	RecJ2 colony hyb probe
RecJ1_F	CGAAGCGGGAGAGGAGGCGAGAGTC	Amplification of 4232 bp region of chromosome containing *recJ1*	pTA1887
RecJ1_R	CCGTCTCTGTTTGTACCTCCG	Amplification of 4232 bp region of chromosome containing *recJ1*	pTA1887
RecJ2_F	GGTAGAACGTGAGGTAGTTGTCGGC	Amplification of 4009 bp region of chromosome containing *recJ2*	pTA1888
RecJ2_R	CGTACCTGCCCTTTCCCTCTTACG	Amplification of 4009 bp region of chromosome containing *recJ2*	pTA1888
RecJ3_F	GTCCTGCGTGAAGTCGAAGAACG	Amplification of 5198 bp region of chromosome containing *recJ3*	pTA1889
RecJ3_R	TTCGTTCTCTTTCTCTCGCAGTCGCTCGC	Amplification of 5198 bp region of chromosome containing *recJ3*	pTA1889
delrecJ4NdeR	CCAATCCATatgGTTGAGCAAACAACCCCTAACTAAAACGATTGTCG	Amplification of the upstream region to *recJ4*	pTA1991
delrecJ4BamF	AGCCGgATCCAGACTCCTCCTCGCGCGG	Amplification of the downstream region to *recJ4*	pTA1991
delrecJ1BamF	GCAGTTGGCTGgAtcCGCGGACTAATCGTCGACCG	Forward inverse PCR primer to delete *recJ1*	pTA1924
delrecJ1NdeR	GGGTCCGTCCATatGTGTGGTACTCACGCCC	Reverse inverse PCR primer to delete *recJ1*	pTA1924
delrecJ3BglF	CGTTCagaTCTCAGACTGCGACCGCTCATTTTTCG	Forward inverse PCR primer to delete *recJ3*	pTA1927
delrecJ3NdeR	CGTCGCTCATAtgCCCGCTAGGTCAGG	Reverse inverse PCR primer to delete *recJ3*	pTA1927
delrecJ2BglF	GCCGTCAGaGAtcTGCGCCGATGAGCCG	Amplification of the downstream region to *recJ2*	pTA1951

delrecJ 2NdeR	GCTCACGGACAT atgGTTATTCTTCG AGG	Amplification of the upstream region to *recJ2*	pTA1951
gintag DSXba R	ACGCtCTaGACGA AATCGGCGCGAG CGTCACCG	Amplification of the downstream region to ginS and introduction of an XbaI site	pTA2315
ginCte rmBa mF	GGTGCGgATCcCC GCCGACGTGGGG ACCATCCTCG	Amplification of C-terminal region of *ginS* (~100 bp) and introduction of a *Bam*HI site	pTA2315
TrpAB glF	GCTCGGTACCCG GaGATCtCGTGGA TAAAACCCC	Amplification of *trpA* incorporating a *Bgl*II site	pTA2335
TrpAB glR	CCAAGCTTGCAT GCCaGATCtGTTA TGTGCGTTCC	Amplification of *trpA* incorporating a *Bgl*II site	pTA2335
NdeLin kBamF	TATGCGACGCGT CGGAATTCCCG	NdeLinkBam linker oligo with *Nde*I and *Bam*HI ends	pTA1924, pTA1991
BamLi nkNde R	GATCCGGGAATT CCGACGCGTCGC A	NdeLinkBam linker oligo with *Nde*I and *Bam*HI ends	pTA1924, pTA1991
BglLin kNdeR	GATCTGGGAATT CCGACGCGTCGC A	NdeLinkBgl linker oligo with *Nde*I and *Bgl*II ends	pTA1927, pTA1951
NdeLin kBglF	TATGCGACGCGT CGGAATTCCCA	NdeLinkBgl linker oligo with *Nde*I and *Bgl*II ends	pTA1927, pTA1951
TRPA R	CGTTGCGACGCG CCCGCTACC	PCR screen for upstream pop-in of pTA1996 at RecJ2 locus	
PrRad AF	CGAGGGCGACGA GAGCTACTGGTA CCCGC	Confirmation of Δ*radA* via PCR	
PrRad AR	AAACGACCGTCC CACCCGTGCCAC CGCC	Confirmation of Δ*radA* via PCR	

Lowercase letters indicate mismatches with the wild-type gene.

2.1.4 Chemicals and Enzymes

Unless otherwise stated, all enzymes were purchased from New England Biolabs (NEB) and all chemicals from Sigma.

2.1.4.1 Media

Haloferax volcanii Media

Sterilisation of all media was achieved via autoclaving (15 minutes, 121˚C). Media are subsequently stored in the dark at room temperature unless otherwise

stated. All plates are stored in sealed bags in the dark at 4˚C to minimise desiccation. Before use plates are dried for a minimum of 30 minutes.

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 pH7.5.

18% salt water: Dilution of 30% SW, autoclaved, then addition of 3 mM $CaCl_2$ once cool.

Trace elements (traditional): 1.82 mM $MnCl_2.4H_2O$, 1.53 mM $ZnSO_4.7H_2O$, 8.3 mM $FeSO_4.7H_2O$, 200 µM $CuSO_4.5H_2O$. Filter sterilised and stored at 4°C.

100 x Trace elements (Modified from Trace elements Duggin *et al* 2015): 13.44 mM EDTA, 4.9 mM $FeCl_3$, 0.37 mM $ZnCl_2$, 0.074 mM $CuCl_2$, 0.077 mM $CoCl_2.6H_2O$, 0.16 mM H_3BO_3, 12.7 mM $MnCl_2.4H_2O$, 0.065 mM $NiSO_4$, 0.062 mM $Na_2MoO_4.2H_2O$, pH to 7.0 with NaOH. Filter sterilised and stored at room temperature.

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

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

10 x Ca: 5% casamino acids, 17.6 mM KOH. Not autoclaved, used immediately.

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

KPO_4 Buffer: 308 mM K_2HPO_4, 192 mM KH_2PO_4, pH to 7.0 with NaOH.

Hv-YPC agar: 1.6% Agar (Bacto), 18% SW, 1 x YPC, 3 mM $CaCl_2$. Microwaved before addition of 10x YPC and $CaCl_2$ to dissolve agar. 10 x YPC added, autoclaved, then $CaCl_2$ added prior to pouring.

Hv-Ca agar: 1.6% Agar (Bacto), 18% SW, 1 x Ca, 0.84% v/v of Hv-Ca Salts, 0.002%v/v of KPO_4 Buffer (pH 7.0). Microwaved before addition of 10x Ca, Hv-Ca Salts, and KPO_4 Buffer to dissolve agar. 10 x Ca added, autoclaved, then Hv-Ca Salts and KPO_4 Buffer added prior to pouring.

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

Hv-Ca+ broth: 18% SW, 1 x Ca, 30 mM Tris.HCl pH 7.0, 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. 18% SW, 1 x Ca, 30 mM Tris.HCl pH 7.0 autoclaved. All other components added once cool.

Haloferax volcanii Media supplements

Table 2.6 **All solutions were filtration sterilised through a 0.2 µm filter.**

Supplement	Abbreviation	Final concentration
Leucine	Leu	50 µg/ml
Uracil	Ura	50 µg/ml
Thymidine	Thy	50 µg/ml (+ 50 µg/ml hypoxanthine for 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

Table 2.7 ***H. volcanii* mutant growth in various media.**

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

Eschericha coli Media

Sterilised by autoclaving and stored at room temperature.

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

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

Eschericha 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	40 μg/ml
Isopropyl β-D-1-thiogalactopyranoside	IPTG	0.2 mM

2.1.4.2 Other Solutions

TE: 10 mM Tris.HCl pH 8.0, 1 mM EDTA
Sodium Acetate: 3 M NaAc pH 5.2, filter sterilised

2.2 Methods

2.2.1 *Escherichia coli* Microbiology

Growth and Storage

Cultures on solid media were grown overnight (16 h) in a static incubator (LEEC) at 37˚C. Small liquid cultures (1-5 ml) were grown for approximately 16 h at 37˚C with 8 rpm rotation. Large liquid cultures (300 ml) were grown for approximately 16 h in an Innova 4330 floor-standing shaking incubator (New Brunswick Scientific) at 37˚C with 110 rpm shaking. Both solid and liquid cultures were stored at 4˚C for a period up to 1 week. For longer durations, a 20% v/v glycerol stock of the liquid cultures were made, then snap frozen on dry ice before storing at -80˚C.

Preparation of Electrocompetent Cells

Electrocompetent cells, *dam*+ and *dam*-, were prepared from the *E. coli* strains XL-1 Blue (resistant to tetracycline) and N2338 were respectively.

A 5 ml culture was grown at 37˚C with 8 rpm rotation for ~ 16 h with the appropriate antibiotic selection. A 10^{-2} dilution of the cells in LB broth, maintaining the appropriate antibiotic selection, was grown at 37˚C with 110rpm shaking until A_{650} of 0.5 – 0.8. Cells were pelleted at 6000 x *g* for 12 minutes at 4˚C and the supernatant discarded. The pellet was resuspended in an equal volume of sterile 1 mM HEPES (pH 7.5) kept on ice. Pelleting and resuspension

of cells was repeated using 0.5 volumes 1 mM HEPES, 0.25 volumes 1 mM HEPES + 10% glycerol, 0.1 volumes 1 mM HEPES + 10% glycerol, and 0.001 volumes 1 mM HEPES + 10% glycerol. Cells were stored in 100 µl aliquots at -80˚C which were first snap frozen on dry ice.

Transformation of *Escherichia coli* via 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.

Approximately 1 µg of DNA, in a maximum volume of 4 µl, 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 SOC was immediately added. Cells were allowed to recover for 1 hour at 37˚C with 8 rpm rotation, before being plated onto LB+2xAmp plates and incubated at 37˚C overnight.

2.2.2 *Haloferax volcanii* Microbiology

Growth and Storage

Cultures on solid media were grown in a static incubator (LEEC) at 45˚C for 5-14 days in plastic bags to minimise desiccation. Small liquid cultures (1-5 ml) were grown for approximately 16 h in a static incubator at 45˚C with 8rpm rotation. Larger liquid cultures (50-333 ml) were grown for approximately 16 h in an Innova 4330 floor-standing shaking incubator (New Brunswick Scientific) at 45˚C with 110 rpm shaking. 2 L cultures were grown for approximately 16 h in a FerMac 360 controlled Bioreactor (Electrolab) at 45˚C with agitation. Agitation was variable dependent on dissolved oxygen. Cultures on solid media were stored at room temperature for a period up to 1 month. Though liquid cultures can be stored at room temperature, pelleting of the cells occurs within a couple of days. For long-term storage, a 20% v/v glycerol, stock of the liquid cultures were made, then snap frozen on dry ice before storing at -80˚C.

Transformation of *Haloferax volcanii* using PEG600

H. volcanii can be easily transformed with the addition of PEG600 (Cline *et al* 1989). Plasmid DNA requires passage through a *dam- E. coli* host prior to transformation due to the presence of the *H. volcanii* restriction endonuclease Mrr which targets methylated DNA (Holmes *et al* 1991). Alternatively, the *dam+* plasmid DNA may be used to directly transform Δ*mrr* strains of *H. volcanii*.

All buffers and solutions were 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: 83 mM EDTA pH 8.0, 15 µl unbuffered spheroplasting solution, ~1-2µg DNA in a final volume of 30 µl per transformation.
60% Polyethylene Glycol 600 (PEG 600): 60% v/v PEG 600 in unbuffered spheroplasting solution.
Spheroplast Dilution Solution: 23% SW, 15% sucrose, 37.5 mM $CaCl_2$.
Regeneration Solution: 18% SW, 1 x YPC, 15% sucrose, 30 mM $CaCl_2$.
Plating Solution: 18% SW, 15% sucrose, 30 mM $CaCl_2$.

A 5 ml culture of *H. volcanii* was grown in Hv-YPC at 45˚C until an A_{650} of 0.6-0.8. Cells were pelleted at 3300 x *g* for 8 minutes in a 14 ml round-bottomed tube (Sarstedt), and the supernatant was removed. The cells were gently resuspended in 1 ml of buffered spheroplasting solution and transferred to a fresh 2 ml round-bottom tube. Cells were pelleted again at 3300 x *g* for 8 minutes and supernatant removed. The cells were resuspended gently in 600 µl buffered spheroplasting solution and aliquoted in 200 µl into fresh 2 ml round-bottom tubes. A 200 µl aliquot was used per transformation. A 20 µl drop of 0.5 M EDTA pH 8.0 was pipetted onto the side of the horizontally held tube, inverted gently, and incubated at room temperature for 10 minutes to allow spheroplast formation. Transforming DNA was added in the same manner as the EDTA and incubated at room temperature for 5 minutes. An equal volume of 60% PEG 600 was pipetted onto the side of the horizontally held tube, mixed by gentle rocking, and incubated at room temperature for 30 minutes. Spheroplasts were diluted by the addition of 1.5 ml spheroplast dilution solution, inverted gently, and incubated at room temperature for 2 minutes. Spheroplasts were pelleted at 3300 x *g* for 8 minutes and the supernatant was removed. The pellet was transferred whole to a sterile 4 ml tube containing 1 ml regeneration solution and was allowed to recover undisturbed for 90 minutes at 45˚C. The pellet was resuspended via tapping of the tube and incubated for a further 3 hours at 45 ˚C with 8 rpm rotation. Cells were transferred to fresh 2 ml round-bottom tube to be pelleted at 3300 x *g* for 8 minutes, the supernatant removed. The cells were resuspended gently in 1 ml plating solution. Appropriate dilutions were made before plating 100 µl on appropriate media. Plates were incubated at 45˚C for 5-14 days.

2.2.3 DNA Extraction from Cells

Plasmid Extraction from *Escherichia coli*

Plasmid extraction from *E. coli* was achieved using Macherey-Nagel Nucleospin (Mini) and Nucleobond AX (Midi) kits. Manufacturer's guidelines were followed. Minipreps and midipreps used 1.5 ml and 300 ml of *E. coli* cell culture respectively. Plasmid DNA from minipreps was eluted in 50 µl elution buffer. Plasmid DNA from midipreps was isopropanol and ethanol precipitated, before resuspending in 200 µl TE. Plasmid DNA samples were stored at -20˚C.

Plasmid Extraction from *Haloferax volcanii*

EDTA Supplemented ST buffer: 1 M NaCl, 20 mM Tris.HCl pH 7.5, 50 mM EDTA.

Macherey-Nagel Nucleospin (Mini) kits were used to isolate circular plasmid DNA from *H. volcanii.* The low plasmid copy number and large amount of cellular debris in *H. volcanii* requires the following amendments to the manufacturer's guidelines:

Starter culture: 30 ml of cell culture was used for minipreps.

Resuspension step: Cell pellets were resuspended in the same total volume as *E. coli* minipreps. However, the cell pellets were initially resuspended in ½ volume of ST buffer supplemented with 50 mM of EDTA, then ½ volume of the standard resuspension buffer (Macherey-Nagel) was added to make up the total volume.

Chloroform extraction: Chloroform extraction steps were required to separate DNA from the cellular debris in order to avoid column blockage. An equal volume of isoamyl alcohol:chloroform mixture (1:24) was added to cell lysates and vortexed to mix. The samples were then centrifuged for 1 minute at 3300 x *g*. The aqueous top layer was transferred to a fresh tube and the chloroform extraction step was repeated until the cell debris at the interface was minimal. 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 5 ml culture of *H. volcanii* was grown in Hv-YPC at 45˚C until an A_{650} of 0.6-0.8. Cells were pelleted at 3300 x *g* for 5 minutes and the supernatant was removed. The cells were resuspended in 200 µl ST buffer followed by addition of an equal volume of lysis solution and inverted to mix. The lysate was then overlaid with 1 ml of 100% EtOH. DNA was spooled at the interface onto a capillary tip until the liquid was homogenous and clear. After washing the DNA twice in 1ml of 100% EtOH, the DNA was air-dried then resuspended in 450 µl of TE. The DNA was isopropanol precipitated before thoroughly resuspending the pellet in 100 µl of TE. Resuspended DNA was stored at 4˚C.

2.2.4 Nucleic Acid Manipulation

PCR Amplification

Amplification of DNA was performed using Q5 Hotstart polymerase (NEB) which is suitable for amplifying templates with a high GC content when used in combination with the Q5 High GC Enhancer. Colony PCRs, to screen large numbers of colonies for a desired plasmid or chromosomal gene, were performed using OneTaq (NEB). All reactions were carried out using a Techne TC-512 thermocycler following the conditions shown below.

Q5 PCR reaction components: 200 µM of each dNTP, 0.5 µM of each primer, 1 ng – 1 µg genomic DNA or 1 pg – 1 ng plasmid DNA template, 1x Q5 Reaction Buffer, 1x Q5 High GC Enhancer, 0.02 U/µl Q5 Hotstart.

OneTaq PCR reaction components: 200 µM of each dNTP, 0.5 µM of each primer, 0.2 µl genomic prepped DNA, 1x OneTaq GC buffer, 0.02 U/µl OneTaq.

Table 2.9 **PCR reaction conditions.**

Step	Q5	OneTaq
Initial Denaturation	98˚C,30 seconds	94˚C, 30
Denaturation	98˚C, 10 seconds	94˚C, 30
Annealing	Tm˚C, 20 seconds	Tm˚C, 20 seconds
Extension	72˚C, 20/30 sec/kb for plasmid and genomic DNA respectively	68˚C, 60 sec/kb
Final Extension	72˚C, 10 minutes	68˚C, 10 minutes

Annealing temperature for primers (Tm˚C) were calculated using the following equation, Equation 2.1 (Howley *et al* 1979).

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

Touchdown PCR

The introduction of novel restriction sites or point mutations requires primers that are not 100% homologous to the template DNA. In these instances, two annealing temperatures were calculated, the first based on the original (mismatched) %homology (Tm_s), the second based on 100% homology (Tm_e). During the first 10 cycles of the reaction the annealing temperature is gradually increased from Tm_s to Tm_e, for the remaining 20 cycles Tm_e remains the annealing temperature.

Colony PCR

In order to screen large numbers of colonies for a desired pop-in orientation, plasmid, or chromosomal gene, colony PCRs were used. *H. volcanii* 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 cells were resuspended in 100 µl of dH_2O by pipetting up and down with the yellow tip used to pick the cells. Boiling at 100˚C lysed the cells which were then cooled on ice. 1 µl of this crude DNA prep was used as template DNA in a PCR reaction with OneTaq polymerase.

Overlap extension PCR

The introduction of point mutations within a gene requires complementary forward and reverse primers containing the desired mutation. PCRs to amplify either side of the desired mutation are conducted using one of these complementary primers and an appropriate external primer. The products from these two PCRs are then mixed in an approximate equal proportion to undergo a third PCR. During the first 10 cycles of the reaction no primers were present and the annealing temperature is calculated based on the overlap of the two products. The external primers only were then added to the mix, which amplified the mutated region for a further 20 cycles (Figure 2.1).

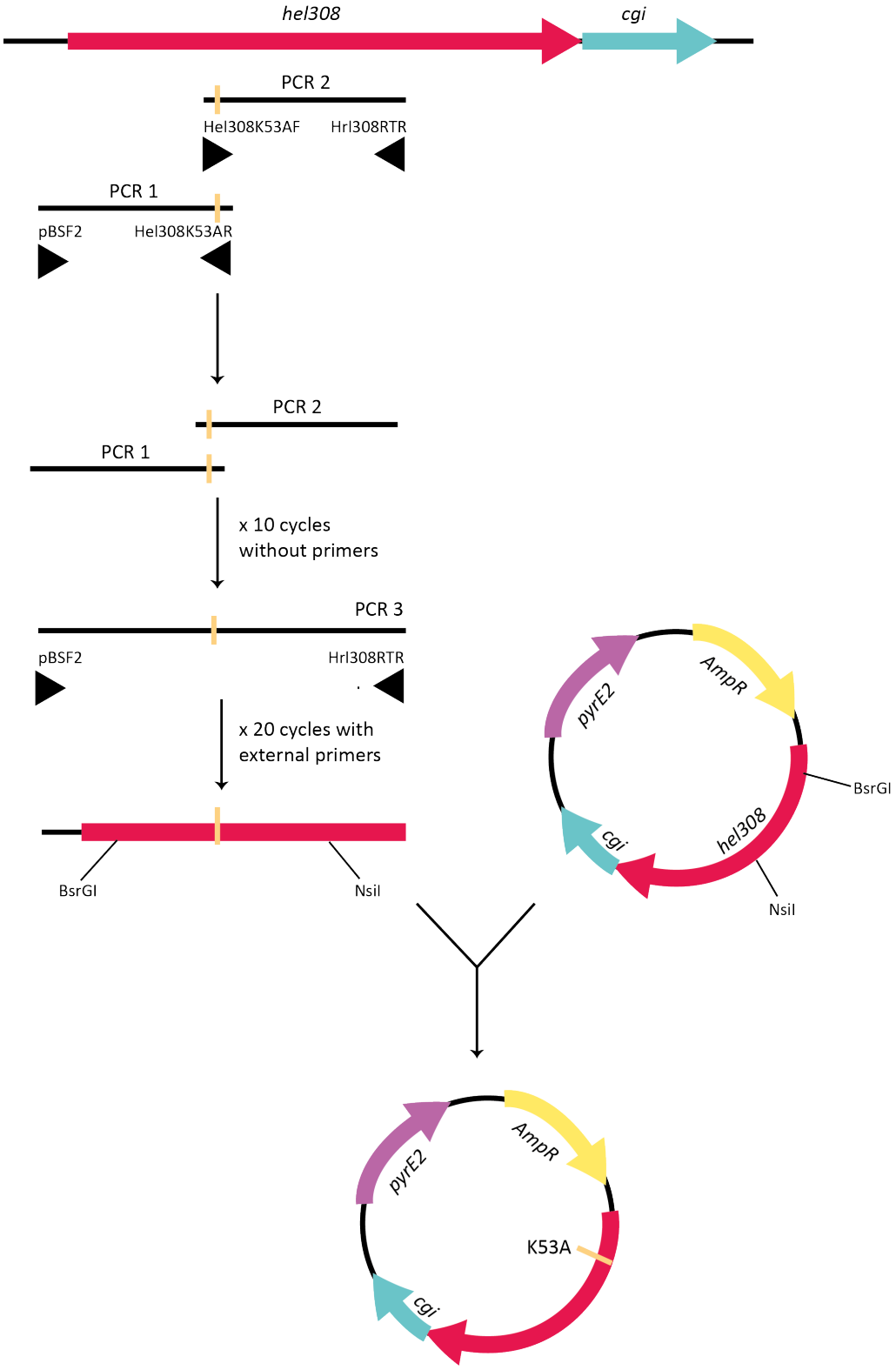


Figure 2.1 **Schematic representation of overlap extension PCR**. PCR 1 and PCR 2 amplify either side of the desired mutation using the reverse and forward complementary primers respectively, and an appropriate external primer. PCR 1 and PCR2 products are mixed in approximate equal proportion and undergo 10 cycles of PCR without primers. External primers are added for the last 20 cycles of PCR. Restriction digest replaces the genomic clone with the mutated fragment.

Restriction Digests

Digests were carried out following the manufacturer's instructions (NEB). For double digests, the NEB buffer was selected such that each enzyme had at least 75% activity. Where this was not possible, a sequential digest was employed cutting with the enzyme requiring the lowest concentration of NaCl first. The subsequent addition of salt allowed cutting with enzymes requiring higher salt concentrations. Genomic DNA was digested for 16 hours, and plasmid DNA for 1 hour.

Blunt-end Filling with Klenow

Blunt ends were produced by the filling in of overhangs, generated via restriction digest, using Klenow enzyme (NEB). Samples were incubated for 30 minutes at 25˚C with 1 units of Klenow per µg of DNA, 1 x NEB Buffer 4, and 1 mM dNTPs, before heat inactivation at 75˚C for 20 minutes.

Dephosphorylation of Vector DNA

The addition of 5 units of Shrimp alkaline phosphatase (rSAP) to a restriction digest was used to prevent self-ligation of vector DNA. Samples were incubated for 30 minutes at 37˚C, before heat inactivation at 65˚C for 10 minutes.

Ligation of DNA

All ligations were carried out using T4 DNA ligase. Reactions contained a molar ratio of ~3:1 insert to vector DNA, 5 units of ligase per µg of DNA, and 1 x T4 ligase buffer. Ligations were incubated at 15˚C for a minimum of 16 hours, DNA was then ethanol precipitated ready for transformation.

Ethanol Precipitation of DNA

DNA was ethanol precipitated by the addition of 1/10 volumes of 3 M sodium acetate pH 5.3 and 2 volumes of 100% EtOH. Samples were incubated at -20˚C for 1 hour before pelleting the DNA at 4˚C, 20000 x *g* for 30 minutes. Pellets were washed in 70% ethanol followed by centrifugation at 4˚C, 20000 x *g* for 10 minutes. The supernatant was removed and the pellets air-dried. The DNA was resuspended in sterile dH_2O or TE.

DNA Sequencing

DNA sequences were obtained using the dideoxy chain termination method (Sanger *et al* 1977) Reactions and analysis were performed by the Deep Seq, University of Nottingham.

Oligonucleotide Synthesis

Oligonucleotides were synthesised by Eurofins MWG, Germany.

Nucleic Acid Quantification

The concentration and purity of plasmid preparations were determined by measuring the absorbance at 260 nm and the 260:280 nm absorbance ratio respectively using a spectrophotometer (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, 20 mM acetic acid, 1 mM EDTA.
5 x Gel Loading Dye: 50 mM Tris.HCl, 100 mM EDTA, 15% w/v Ficoll, 0.25% w/v Bromophenol Blue, 0.25% w/v Xylene Cyanol FF.

TBE buffer and agarose powder (Appleton Scientific) was used for the standard casting and running agarose gels. For higher quality resolution or Southern blotting, TAE buffer and agarose powder (Sigma) was used. DNA samples were resolved with 1 x loading dye, alongside either 1 Kb ladder (NEB or 100 bp ladder (NEB). For visualization of bands, TBE gels were stained with either ethidium bromide or, when the gel was used for DNA extraction, SYBR safe. TAE gels were post-stained with 0.5 µg/ml ethidium bromide for 30 minutes, first with the removal of sample lanes when the gel was used for DNA extraction. TBE gels (10 cm) were run at 100V for 1 hour. TAE gels (25 cm) were run at 50 V or 25 V, when one or two combs were used respectively, with buffer circulation for 16 hours.

Agarose Gel Extraction and Purification of DNA

DNA was visualised by UV light using a UV transilluminator (UVP inc.), the DNA to be excised was first protected from UV exposure using foil. DNA was purified using Macherey-Nagel DNA purification kits. Manufacturer's guidelines were followed.

2.2.5 Genetic Manipulation of *Haloferax volcanii*

Generating a Genomic Clone

Either PCR or restriction digest was used to isolate genomic clones of genes of interest from high quality genomic DNA. PCRs were designed to encompass restriction sites which were at least 500 bp upstream and downstream of the gene of interest. The restriction digest method used similarly located restriction sites.

Generating a Genomic Clone via PCR

After the PCR was conducted, the PCR product was digested overnight and resolved on a TBE gel. DNA was extracted, purified, and ligated into compatible sites of pBluescript II SK+. White colonies were selected and the plasmid confirmed by restriction digest and sequencing.

Generating a Genomic Clone via Restriction Digest

Genomic DNA was digested for 16 hours before being resolved on a 0.75% TAE agarose gel for 16 hours at 50V. The band of expected size (+/- 0.5 Kb) was extracted and DNA was purified. Purified DNA was ligated into compatible sites of pBluescript II SK+ to generate a plasmid library. The library was then transformed into XL1-Blue *E.coli* and transformants were selected on LB +Amp +X-gal. White colonies were patched out and screened by colony hybridisation using a gene specific probe.

Generating a Deletion Construct via PCR

Inverse primers were designed to make a 'clean' deletion of the gene of interest, introducing a *Bam*HI (or compatible) and *Nde*I restriction sites at the site of deletion. Where the inverse PCR was not possible, these primers were used with appropriate external primers to amplify the upstream (US) and downstream (DS) regions either side of the gene of interest. PCR products were digested and appropriate linker was ligated in place of the gene. If the US/DS method was used, the ligated fragment was then digested at appropriate restriction sites and ligated into the original plasmid to replace the gene. The presence of a *Bam*HI (or compatible) site allows for the insertion of the *trpA* marker in place of the gene. The *Nde*I site allows for the insertion of a tagged version of the gene in place of the native gene, and doubles up as the gene's new start codon.

Generating a Gene Replacement Construct

Gene replacement constructs were made by inserting the gene of interest (usually containing a point mutation) along with the gene's flanking regions into pTA131. The method for doing this varied depending on the mutation. For details on individual gene replacement plasmids constructed for this study see relevant result chapters.

Gene Deletion/Replacement

Strains deleted for *pyrE2* (Δ*pyrE2*) are transformed with a plasmid containing a deletion construct for the desired gene, typically a *trpA* marker is included in place of the gene, and the *pyrE2* marker. The strains were plated on media lacking uracil, typically Hv-Ca (+ necessary additives), to select for the integrated *pyrE2*-marked plasmid (pop-in). A single pop-in colony was picked and grown, non-selectively, overnight in a 5 ml Hv-YPC culture to an A_{650} of 1.0. The culture was diluted 1/500 into fresh Hv-YPC and the growth and dilution was repeated. The removal of the uracil selection by growth in Hv-YPC allows the loss of the integrated plasmid and the native gene. Successful pop-outs are selected for on Hv-Ca + 5-FOA media (+ necessary additives). Both wild-type and deletion mutants may be recovered depending upon the location of the recombination event. The inclusion of the *trpA* marker in place of the gene enables direct selection for the deletion mutants (Figure 2.2). Colonies were then restreaked onto selective media and tested for the desired genotype.

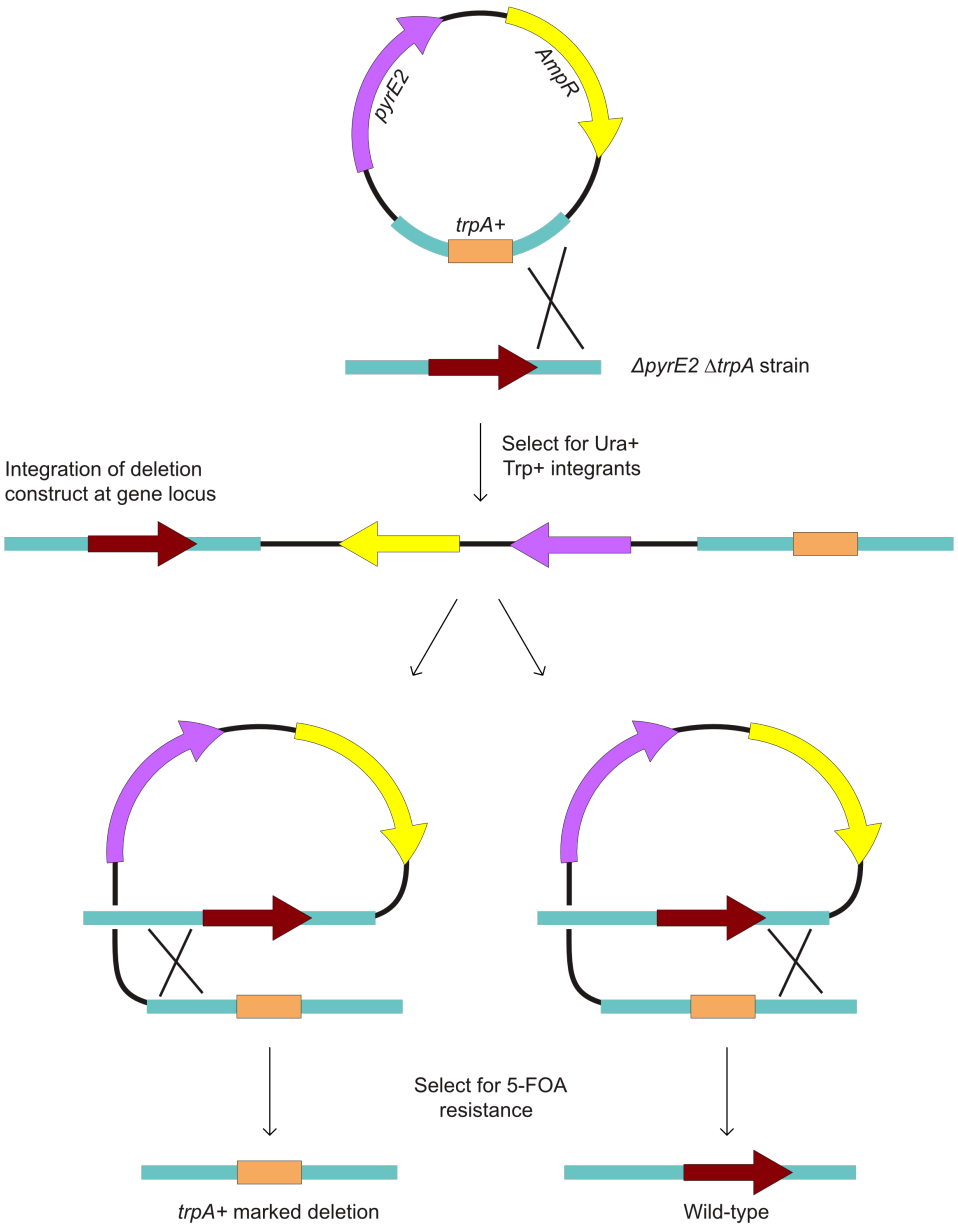


Figure 2.2: **Gene deletion by pop-in/pop-out.** Δ*pyrE2* strains are transformed with a *pyrE2*+ deletion construct. Pop-ins are selected for by their ability to grow on media lacking uracil. Relieving the uracil selection allows for pop-out to occur. Recombination between regions of homology can either be upstream or downstream of the gene. Pop-out is selected for by plating on 5-FOA. The resultant gene allele will either be a deletion or wild-type. Replcing the gene with the *trpA*+ marker in Δ*trpA* strains enables direct selection for deletion mutants.

Deletion of *radA*

RadA is the recombinase that facilitates homologous recombination, which is essential for the pop-in and pop-out steps of gene deletion and gene replacement. Therefore, for a *radA* deletion to occur the strain must be complemented with an episomal copy of the *radA* gene to facilitate the pop-out step (Delmas *et al.*, 2009). After Δ*pyrE2* Δ*trpA* strains have been transformed with the Δ*radA::trpA*+ deletion construction (pTA325) and plated on selective media, integrants were transformed with an episome (pTA637) to allow *in trans* complementation of *radA*. This episome also expresses a Mevinolin resistance gene *mevR*, in addition to the *pyrE2* gene. The drug mevinolin inhibits HMG-CoA reductase, which is required for the production of isoprenoids. Isoprenoids replace fatty acids as the side chains to glycerol in archaeal cell membranes, thus are essential to archaeal cells. The presence of *mevR* on the episome ensures that the episome is retained by the cell in the presence of mevinolin. The pop-out was generated by growing a colony in YPC+Mev to an A_{650} of 1.0, then diluted 10^{-2} in fresh YPC+Mev. When this culture reached an A_{650} of 1.0, it was diluted 10^{-2} in fresh YPC to allow the loss of the episome from the strain. Cultures were then plated onto Hv-Ca +5-FOA to select for pop-outs that have also lost the *pyrE2*-marked episome. Colonies were restreaked onto selective media and tested for the desired genotype (Figure 2.3).

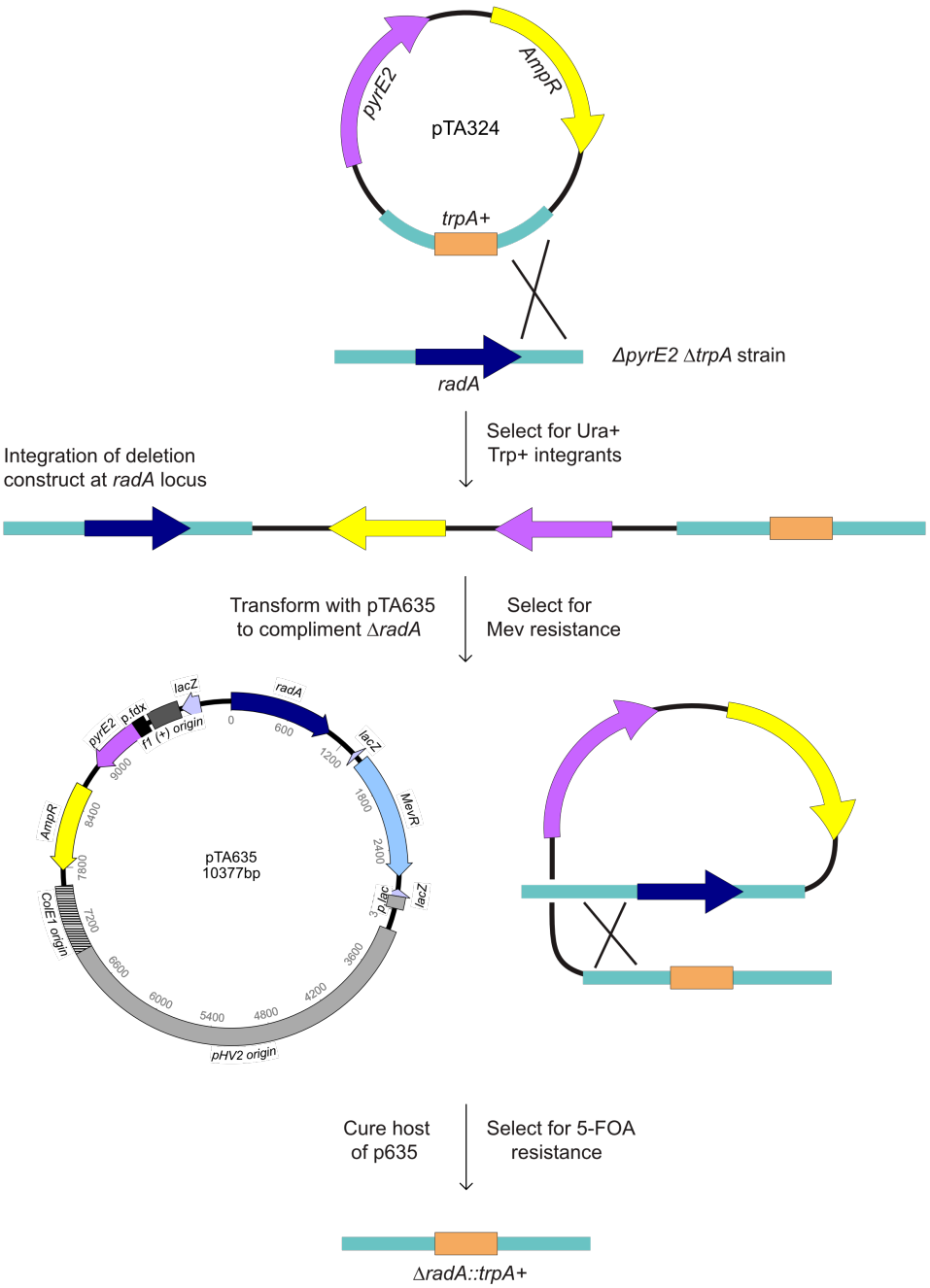


Figure 2.3: ***radA* gene deletion.** A Δ*pyrE2,* Δ*trp* strain is transformed with pTA324, a deletion construct containing a *radA* deletion (5' *radA* still present) and *pyrE2* marker (pop-in). The strain is then transformed with pTA635 (or pTA2080), episomal *radA*+ (*radA-S101P*+) complementing plasmid containing *pyrE2* and *MevR* markers. Pop-out is able to occur due to the presence of *radA*+ on pTA635 (*radA-S101P*+ on pTA2080). Strains are selected for both the pop-out of TA324 and the loss of pTA635 by plating out on 5-FOA (Delmas et al 2009).

2.2.6 Genotype Screening

The presence or absence of a specific gene in a pop-out strain is screened in a variety of different ways. If a selectable marker was used in the generation of the strain (e.g. *trpA*+), the colonies may be plated on selective media. This alone is not sufficient evidence of the presence/absence of a specific gene due to *H. volcanii* being polyploid. Strains plated on selective media could have a mixture of both wild-type and mutant alleles on different chromosome copies, this is known as a merodiploid cell. Therefore, selecting homozygous strains requires additional confirmation methods such as colony hybridisation and Southern blotting.

Colony Lift

20 x 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, alongside controls, were patched onto Hv-YPC plates using sterile wooden toothpicks and incubated at 45˚C for ~3 days. The patched colonies were lifted from the plate by placing a circle of GE Healthcare Amersham Hybond – N^+ positively charged membrane onto the surface for 1 minute. Lysis of the cells was achieved by transferring the membrane colony side up to Whatman paper soaked in 10% SDS and incubating for ≥5 minutes. Similarly, the membrane is transferred to Whatman paper soaked in denaturing solution for ≥5 minutes to denaturing proteins and DNA. Neutralisation was achieved by transferring the membrane to Whatman paper soaked in neutralising solution for ≥5 minutes twice. The membrane was briefly washed ≤30 seconds in 2 x SSPE and was air dried. DNA was crosslinked to the membrane with 120 mJ/cm^2 UV.

Southern Blot by Vacuum Transfer

20 x 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. The digested DNA was resolved on a 200 ml 0.75% TAE gel at 50 V for 16 hours with buffer circulation. The gel was post-stained with 0.5 μg/ml ethidium bromide for 30 minutes. The DNA embedded in the gel was acid-nicked in 0.25 M HCl for 15 minutes, washed in dH_2O for 10 minutes, and denatured in denaturing solution for 45 minutes. A 15 x 25 cm GE healthcare Amersham Hybond – XL 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. The membrane was briefly washed ≤30 seconds in 2 x SSPE and was air dried before DNA was crosslinked to the membrane with 120 mJ/cm^2 UV.

Hybridisation

100 x Denhardt's Solution: 2% Ficoll 400, 2 % polyvinyl pyrrolidone (PVP) 360, 2% BSA (bovine serum albumin, Fraction V).
20 x SSPE: 3 M NaCl, 230 mM NaH_2PO_4, 32 mM EDTA, pH 7.4.
Prehybridisation Solution: 6 x SSPE, 1% SDS, 5 x Denhardt's solution, 200 μg/ml salmon sperm DNA (Roche, boiled for 5 minutes prior to addition).
Hybridisation Solution: 6 x SSPE, 1% SDS, 5% dextran sulphate
Low Stringency Wash Solution: 2 x SSPE, 0.5% SDS.
High Stringency Wash Solution: 0.2 x SSPE, 0.5% SDS.

Membranes were incubated in prehybridisation solution at 65˚C with rotation for ≥3 hours. 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 100˚C for 5 minutes, incubated with the radioisotope and HiPrime random priming mix (Roche) at 37˚C for 15 minutes, and then purified on a BioRad P-30 column. The purified DNA probe was then added to 10 mg/ml salmon sperm DNA to a total volume of 500 μl. The DNA mix was denatured at 100˚C for 5 minutes and quenched on ice. For Southern blots, 1 μl of 1μg/ml 1 kb ladder was also included in the radiolabelling reaction. The prehybridisation solution was replaced with 30 ml hybridisation solution before the addition of the DNA probe. Membranes were incubated in hybridisation solution at 65˚C with rotation overnight. The membranes were then washed twice in low stringency wash solution, the first for 10 minutes and the second for 30 minutes. A further two washes of 30 minutes, this time in high stringency wash solution, was carried out. Membranes were air dried before being wrapped in Saran wrap and exposed to a phosphorimager screen (Fujifilm BAS Cassette 2325) ≥24 hours. The screen was then scanned using a GE Healthcare Typhoon.

2.2.7 Phenotyping of *Haloferax volcanii*

Tryptophan Gradient plates

To generate a gradient of tryptophan across a plate, with the desired tryptophan concentration at one side fading to no tryptophan on the other side, plates were first poured with 17 ml Hv-Ca +Ura +Trp (of the desired concentration) on a 7° slant to form a tapered wedge (Figure 2.4). Once set, the plate was placed flat

and the wedge was covered with 43 ml Hv-Ca + Ura agar, lacking tryptophan (Bryson and Szybalski, 1952; Hawkins *et al.*, 2013).

5 ml cultures of *H. volcanii* strains were grown with 8 rpm rotation in Hv-YPC at 45˚C until an A_{650} of 0.6-0.8. These were then diluted into fresh Hv-YPC and incubated at 45˚C until an A_{650} of 1.0. Serial dilutions of the cultures in 18% SW to 10^{-4} were prepared. Autoclaved paintbrushes (The Range) were first wetted in 18% SW. Using a fresh paintbrush for each strain, the paintbrush was dipped into the diluted culture and then painted in one direction across the gradient plate. Using the same paintbrush dipped again in the diluted culture, a second line was painted over the first in the opposite direction. Once dry, the plates were incubated at 45˚C for 4-5days.

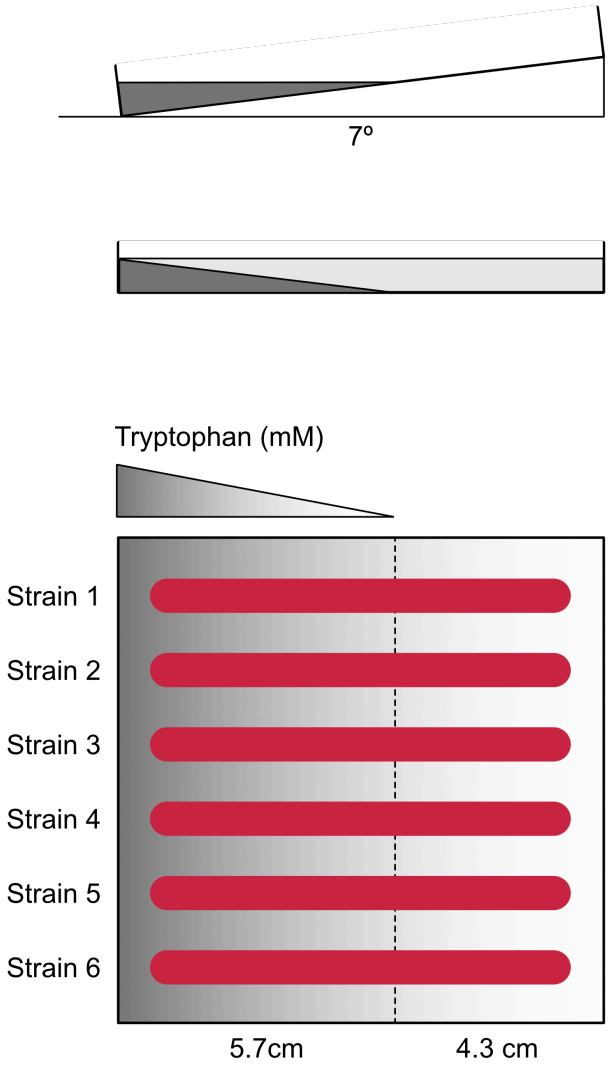


Figure 2.4: **Tryptophan gradient plates.** Plates were first poured with 17 ml Hv-Ca +Ura +Trp (of the desired concentration) on a 7° slant to form a tapered wedge. Once set, the plate was placed flat and the wedge was covered with 43 ml Hv-Ca + Ura agar, lacking tryptophan. Strains are painted across the tryptophan gradient.

Standard Growth Assay

Standard growth rate in liquid media was determined using an Epoch 2 Microplate Spectrophotometer (BioTek). Cultures were prepared in 5 ml Hv-YPC broth and grown to mid-late exponential phase, which corresponds to an A_{650} of 0.2-0.8. These cultures were diluted and again grown to mid exponential phase. Serial dilutions of the cultures were made before loading 150 µl in duplicate, alongside appropriate blanks, to the wells of a 96 well microtiter plate (Corning). The plate was sealed around the edges with microporous tape (Boots) and incubated at 45˚C with double orbital shaking at 1000 rpm for 72 hours in the Epoch. Readings at A_{600} were taken every 15 minutes and converted to a 1 cm pathlength by dividing the raw A_{600} value by 0.14. The generation time was calculated by plotting the growth on a log2 scale and using the following equation, 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 A_{650}
B = start A_{650}

Generation times varied between experiments therefore comparisons were only made between sets of strains within the same experiment i.e. strains that have been incubated on the same 96 well microtiter plate with the A_{600} being measure simultaneously during a single run. Since generation times vary between experiments, the generation times stated are not absolute however, the relationships between the strains are consistent. Therefore, growth curves generated by this method are used to illustrate the relationship of generation times between a set of strains.

Ultraviolet (UV) Irradiation Sensitivity

5 ml of Hv-YPC was inoculated with 1 colony and grown overnight at 45˚C with 8 rpm rotation. The culture was diluted into 5 ml of fresh Hv-YPC and grown to an A_{650} of 0.4. A range of serial dilutions (10^0 – 10^{-6}) of the cells in 18% salt water were made and duplicate 20 µl samples were spotted onto Hv-YPC agar and allowed to air-dry. Plates were exposed to UV light at 254 nm, 1 J/m^2/sec for varying amounts of time and shielded from visible light to prevent photo-reactivation DNA repair. Plates were incubated at 45˚C for 4-7 days. During this period, colonies were counted and survival fractions were calculated relative to a non-irradiated control.

Mitomycin C (MMC) Sensitivity

5 ml of Hv-YPC was inoculated with 1 colony and grown overnight at 45˚C with 8 rpm rotation. The culture was diluted into 5 ml of fresh Hv-YPC and grown to an A_{650} of ~0.4. A range of serial dilutions (10^{0} – 10^{-6}) of the cells in 18% salt water were made and duplicate 20 µl samples were spotted onto Hv-YPC agar containing 0 – 0.02 µg/ml MMC. Plates were allowed to air-dry before incubation at 45˚C for 4-7 days. During this period, colonies were counted and survival fractions were calculated relative to an untreated control. MMC plates were made fresh and used within 2 days.

Recombination Assay

The recombination frequency between the chromosome and a plasmid can be determined in strains derived from a H164 genetic background. These strains have the *leuB-Ag1* frame-shift mutant allele and are therefore unable to grow on media lacking leucine. A recombination event between this allele and a second frame-shift mutant *leuB* allele, *leuB-Aa2* provided on pTA163, generates a wild-type *leuB* allele allowing the strain to now grow on media lacking leucine (Lestini *et al* 2010) (Figure 2.5).

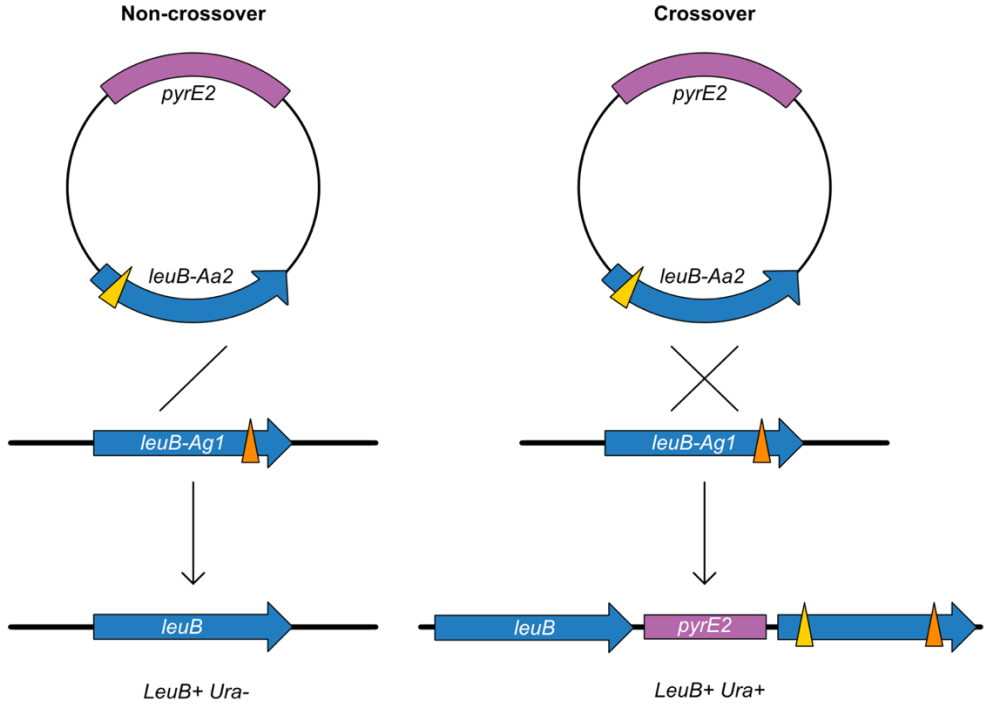


Figure 2.5: **Recombination assay.** Δ*pyrE2* strains with a chromosomal *leuB-Ag1* allele (*leu*-) are transformed with pTA163, containing *pyrE2* and *leuB-Aa2* allele (*leu*-). A recombination event between the plasmid and chromosomal mutant *leuB* alleles generates a wildtype *leu*+ allele and therefore strains are able to grow on media lacking leucine. The proportion of crossover (CO) and non-crossover (NCO) events are determined by the retention or loss of the *pyrE2* marker found on pTA163: CO are *pyrE2*+ *(ura+),* and NCO are Δ*pyrE2 (ura-)*.

Strains were transformed with 1 µg of pTA163 (*pyrE2+, leuB-Aa2*), transformants were allowed to recover undisturbed for 90 minutes at 45˚C. The pellet was resuspended via tapping of the tube and incubated for a further 3 hours at 45 ˚C with 8 rpm rotation. Transformants were plated on Hv-Min +Trp +Ura at $10^0 – 10^{-5}$ dilutions to select for cells that had undergone a recombination event between the two mutant *leuB* alleles, generating a wild-type *leu+* allele. Reversion of this allele is extremely rare and has no effect on the results of this assay (Haldenby 2007). Transformants were additionally plated on non-selective Hv-YPC at $10^{-4} – 10^{-6}$ dilutions, to determine the total viable count.

Crossover (CO) recombination events (integration of the *pyrE2* marked plasmid) and non-crossover (NCO) recombination events (conversion to wild-type *leuB* without integration of the *pyrE2* gene) can be distinguished by patching in duplicate on Hv-Min +Trp plates with and without +Ura. Patches on Hv-Min +Trp +Ura act as a control, ensuring all of the colonies patched were *leu+,* whilst only colonies which have undergone a CO event will grow on the Hv-Min +Trp plates lacking Ura. The fraction of CO events (*leuB+ pyrE2+*) is derived by comparison to the total recombination frequency, the remaining recombination events were NCO (*leuB+ pyrE2-*).

Transformation efficiency of each strain was also calculated. Strains were transformed with 1 µg of the episomal plasmid pTA354 (*pyrE2+*), transformants were allowed to recover undisturbed for 90 minutes at 45˚C. The pellet was resuspended via tapping of the tube and incubated for a further 3 hours at 45 ˚C with 8 rpm rotation. Transformants were plated on Hv-Ca +Trp at $10^0 – 10^{-3}$ dilutions to select for cells that had taken up the plasmid. Transformants were additionally plated on non-selective Hv-YPC at $10^{-4} – 10^{-6}$ dilutions, to determine the total viable count. The recombination efficiency for each strain was normalised to the transformation efficiency.

Flow Cytometry

DNA content and cell size of *H. volcanii* cells were determined via flow cytometry (Delmas, Duggin and Allers, 2013). 5 ml Hv-YPC broth was inoculated and grown at 45˚C with 8 rpm rotation in two successive dilutions until an A_{650} of 0.4 was reached. Acridine orange solution was added to the cells at a final concentration of 1 µg/ml. The sample was run on an Apogee A40 flow cytometer with settings of: Sample flow 1.5 µl /minute, volume 150 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 10,000 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 Protein Purification

A 5 ml starter culture was grown overnight in Hv-YPC supplemented with 1 x Trace elements (Modified) at 45˚C to an A_{650} of ~1.0, then diluted 10^{-2} into fresh Hv-YPC supplemented with 1 x Trace elements and grown for a further 8 hours to an A_{650} of ~1.0. The culture was diluted into 50 ml Hv-YPC supplemented with 1 x Trace elements to allow growth to A_{650} of ~0.5 over a period of 24 hours. This culture was used to inoculate 2 L Hv-YPC supplemented with 1 x Trace elements and grown for approximately 16 h in a FerMac 360 controlled Bioreactor (Electrolab) at 45˚C with agitation to an A_{650} of ~0.4. At this point the culture may be treated with 0.5 µg/ml MMC and incubated for 1 hour. If native protein interactions were to be analysed, then the culture was not treated with MMC. The cells were pelleted at 3300 x *g* for 8 minutes at 4˚C. Pellets were resuspended in 7 ml of either buffer A or B dependant on whether the protein was to be purified via Nickel or Streptactin affinity Chromatography respectively. Cells were lysed via sonication (3-6 x 30 seconds at ≤8 µm amplitude) on ice. Cell lysates were then pelleted at 20000 x *g* for 15 minutes at 4˚C, and the supernatant was filtered sequentially through 0.8 µm, 0.45 µm, and 0.2 µm filters to remove DNA.

Nickel affinity Chromatography

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

First IMAC Sepharose 6 Fast Flow beads (GE Healthcare) were charged with Ni^{2+}. Beads (0.5 ml per column) were washed twice with 10 column volumes (CV) of dH_2O, followed by equilibration of the column with 2 CV 0.2M $NiSO_4$ for 30 minutes at 4˚C with rotation. 6 subsequent washes of the beads were performed, twice with 10 CV H_2O, once with 10 CV Buffer A + 500 mM imidazole, and finally 3 washes with 10 CV Buffer A + 20 mM imidazole. Beads were resuspended in 1 CV Buffer A + 20 mM imidazole. The filtered supernatant, from the cell lysates, was incubated with the Ni^{+} charged beads for 1 hour at 4˚C with rotation. Working at 4˚C, the slurry was applied to a Poly-Prep column (Bio-Rad), the flow-through collected and reloaded onto the column. The column was washed with 10 CV Buffer A and then with 20 CV Buffer A, before eluting in 4 CV Buffer A +100 mM imidazole, 4 CV Buffer A +200 mM imidazole, and 4 CV Buffer A +500 mM imidazole. If protein sample required purification via *Strep*-tactin affinity chromatography, the elutions were first combined and concentrated using a Vivaspin 20 ultrafiltration spin column (GE healthcare). This method also acts to remove imidazole from the sample. A column with an appropriate molecular weight cut off (MWCO) for the protein was selected and equilibrated with dH_2O followed by Buffer B or desired storage

buffer. Samples were added to the column for centrifugation at 8000 x *g* in a swing bucket at 4˚C until the desired final volume was reached. For application onto a *Strep*-tactin affinity column, 1-2 ml of sample is desired.

Strep-tactin affinity Chromatography

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

Working at 4˚C, *Strep*-tactin Sepharose (0.5 ml per column) was applied to a Poly-Prep column (Bio-Rad) and equilibrated twice with 2 CV Buffer B. Protein sample was applied to the column, the flow-through collected and reloaded onto the column. The column was washed 5 times with 1 CV Buffer B. Bound protein was eluted in 0.8 CV, 1.4 CV, and 0.8 CV of elution buffer.

Protein Storage

Elutions were stored at 4˚C for the short-term. Long-term storage of proteins required the addition of 10% glycerol before snap freezing on dry ice and storage at -80˚C.

TCA Protein Precipitation

Resuspension buffer: 4% SDS, 0.2 M Tris pH 7.4, 0.15 M NaOH.
4x protein sample buffer (4x Binding Buffer B): 80 mM HEPES pH 7.5, 8 M NaCl, 4 mM PMSF
4x Laemmli buffer: 50 mM Tris pH 6.8, 100 mM DTT, 2% SDS, 0.1% bromophenol blue, 10% glycerol.

$1/10^{th}$ volume of 0.15% deoxycholate was added to samples, vortexed to mix, and incubated at room temperature for 10 minutes. $1/10^{th}$ of the original volume of 72% trichloroacetic acid (TCA) solution was then added and incubated for 5 minutes at room temperature. Samples were pelleted at 16,000 x *g* for 8 minutes and the supernatant removed. The pellet was resuspended in a 3:1 solution of resuspension buffer: 4 x protein sample buffer. The protein was then heat denatured by boiling at 94˚C for 5 minutes in 1 x Laemmli buffer and analysed via SDS-PAGE.

SDS-Polyacrylamide Gel Electrophoresis

10% SDS-PAGE gel (resolving): 10% 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% SDS-PAGE gel (stacking): 3% acrylamide/bisacrylamide Protogel, 0.25 M Tris pH 6.8, 0.2% SDS, 0.125% AMPS, 0.125% TEMED.
10x SDS-PAGE running buffer: 0.25M Tris, 1.92 M glycine, 1% SDS.
4x Laemmli buffer: 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). Isopropanol was layered onto the freshly poured 10% resolving gel to leave a level surface. The isopropanol was washed off once the gel had set, the 3% stacking gel applied and the appropriate come inserted. Laemmli buffer was added to protein samples at a 1x concentration, protein was then denatured by boiling at 94˚C for 5 minutes before resolving alongside Blue Protein Standard, Broad Range ladder (NEB) at 200V, 36 mA for 90 minutes in 1x SDS-PAGE running buffer. If the gel was to be used for Western blot, samples were run alongside Precision Plus Protein Kaleidoscope Protein Standard (BioRad) instead. For visualisation of proteins, gels were either incubated with Page-Blue Protein Staining Solution (Thermo Scientific) or silver stained. Alternatively, gels were analysed via Western blot (see below for details).

Protein Identification by Western Blot

Transfer buffer: 12 mM Tris base, 96 mM Glycine, 20% v/v Methanol.
10 x TBS: 198 mM Tris base, 936 mM NaCl2. pH adjusted to 7.6 with 1 M HCl.
TBST: 1 x TBS, 0.1% Tween20.

Freshly run, unstained SDS-PAGE gels were removed from cassettes and the stacking gel and foot removed. Sponges, Whatman paper and, Amersham Protan 0.45 μm NC membrane (GE Heathcare) were soaked in transfer buffer for 5 minutes before the Western blot apparatus was assembled, ensuring there were no air bubbles. Transfer was run at 25 V for 2 hours in NovexTM XCell IITM blotting apparatus.

When transfer was complete, the membrane was blocked with 5% w/v milk for 1 hour at room temperature, with agitation. The membrane was then incubated with an HRP conjugated anti-His_6 monoclonal antibody (120 ng/mL in 5% w/v milk, Protein Tech) overnight at 4˚C with agitation. The membranes were then washed 4 times of 5 minutes with TBST. The membrane was then placed in

Saran wrap and incubated with Amersham ECL prime reagent (GE Healthcare). Excess was removed after 1 minute. Chemiluminescence was detected using a Fusion FX system (Vilber).

Mass Spectrometry

Mass spectrometry was performed by the Cambridge Centre for Proteomics, University of Cambridge. Proteins in gel bands were reduced (DTT) and alkylated (iodoacetamide) and subjected to enzymatic digestion with Trypsin. The resulting peptides were loaded onto an autosampler for automated electrospray ionization-tandem mass spectrometry (MS/MS) analysis. All LC-MS/MS experiments were performed using a Dionex Ultimate 3000 RSLC nanoUPLC system and a QExactive Orbitrap mass spectrometer (Thermo Fisher Scientific Inc, Waltham, MA, USA). Peptides were separated on a PepMap C_{18} reversed-phase, 2µm particle size, 100A pore size, 75 µm-inner-diameter, 50-cm column at a flow rate of 300 nL/min. Protein Discoverer software version 2.1 (Thermo Scientific) was used to generate a peak list file of uninterpreted fragment mass data, which was used to search against the UniProt archaea database (ref: 20190528, 3149228 sequences; 899013487 residues) and common contaminant sequences (123 sequences; 40594 residues) using the MASCOT search engine. Only protein identifications with probability-based MOWSE scores above a threshold of $P < 0.05$ were accepted.

Chapter 3: Separation of function in Hel308

3.1 Background

Structure of Hel308a

No crystal structure for the *Haloferax volcanii* Hel308 has been obtained to date; there are resolved crystal structures of three archaeal Hel308 homologues. The Hel308 structure from the euryarchaeon *Archaeoglobus fulgidus* was the first to be resolved both as an apo and a DNA-Hel308 co-crystal (Büttner, Nehring and Hopfner, 2007). In addition, the structure of Hel308 from the crenarchaeon *Sulfolobus solfataricus* and the structure of the Hel308 homologue Hjm from the hyperthermophile *Pyrococcus furiosus* have been resolved (Richards *et al.*, 2008; Oyama *et al.*, 2009). These crystal structures reveal highly conserved protein folding, despite the *Pyrococcus furiosus* Hjm sharing a low amino acid sequence identity of only 30% and 37% with the *Archaeoglobus fulgidus* and *Sulfolobus solfataricus* Hel308 helicases, respectively (Oyama *et al.*, 2009).

From these crystal structures, Hel308 has been shown to comprise 5 domains around a central DNA binding core, which is identifiable by a lining of essential DNA binding residues. The 3' ssDNA overhang threads through this DNA binding core and the five domains drive a β-hairpin between the two strands, thereby separating the DNA duplex. In common with all superfamily 2 helicases, Hel308 requires ATP binding and hydrolysis to enable translocation along the DNA. The ATP binding site motifs I, II, V and VI are located within domains 1 and 2 RecA folds, and the 'β-hairpin plough' is also located in domain 2. Four α-helices and two parallel β-strands make up a non-canonical winged-helix fold contained within domain 3, thought to be important for the structural integrity of the RecA fold. This proposal arises from the fact that winged-helix domains are common to nucleic acid binding motifs, and that domains 1 and 3 in Hel308 are packed tightly together (Woodman and Bolt, 2011a). Domain 4 is also known as the 'ratchet' domain, it is composed of a seven-helix bundle with the centre α-helix acting as the so-called ratchet (Woodman, Briggs and Bolt, 2007). It is thought that the movement of this ratchet is coupled with ATP hydrolysis via a highly conserved 'FHH' linker, which is a part of domain 2 (Gamble-Milner, 2016). Domain 4, together with domains 1 and 3, forms a ring around the ssDNA 3' tail. Finally, domain 5 extrudes from this central pore to bind ssDNA and is able to do so via its helix-loop-helix structure. Domain 5 is thought to act as an auto-inhibitory domain or 'molecular brake', in addition to having a role in coupling ATP hydrolysis to the helicase unwinding (Woodman, Briggs and Bolt, 2007; Richards *et al.*, 2008) (Figure 1.26).

Hel308b

In addition to the canonical Hel308, a second *hel308* gene known as *hel308b* (HVO_0971) is present in *Haloferax volcanii*. A gene duplication event is the probable cause of the presence of this second *hel308* gene. In addition to *Haloferax volcanii*, a *hel308b* like gene has also been identified in four closely related haloarchaeal species. This second helicase is shorter in length than the canonical Hel308, 639 amino acids long compared with the 827 amino acid long canonical Hel308, consequently Hel308b lacks domain 5 which is key in the regulation of ATP hydrolysis (Figure 3.1). Mutants deleted for *hel308b* show no sensitivity to DNA-crosslinking agents unlike the *Δhel308* mutant (Gamble-Milner, 2016). However, similar to the *Δhel308* mutant, the *Δhel308b* mutant does exhibit a recombination defect compared to the wild type. This suggests that Hel308b may have a regulatory role in homologous recombination, Hel308b may even be involved in the promotion of homologous recombination. The lack of domain 5 implies that the exact role of Hel308b is independent of helicase function (Gamble-Milner, 2016). Mutants deleted for both *hel308a* and *hel308b* demonstrate a sensitivity to DNA-crosslinking agents and a recombination defect compared to the wild type similar to that of the *hel308a* single delete mutants (Gamble-Milner, 2016). Hel308b (HVO_0971) is expressed in the wildtype strain at a transcription level of 112 rkpm. Although the expression of Hel308b is less than that of Hel308a, which has a transcription level of 430 rkpm, Hel308b is still found within the medium expression level (Quartile 3) of the dataset.

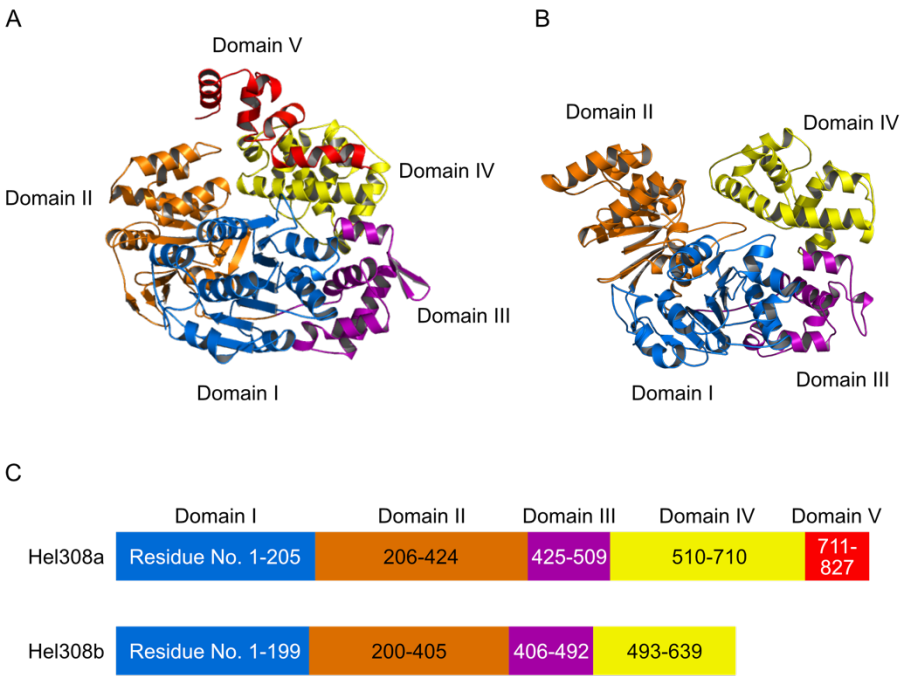


Figure 3.1: ***Haloferax volcanii* Hel308a and Hel308b predicted structures.** A) Predicted structure of Hel308a (HVO_0014). B) Predicted structure of Hel308b (HVO_0971). C) Linear schematic of the structural domains of *H. volcanii* Hel308a and Hel308b. Hel308a and Hel308b structures predicted from the respective amino acid sequences using Phyre2 (Kelley et al 2015).

Genetic Analysis of *hel308* Point Mutants

ATPase-null mutants

The binding and hydrolysis of ATP is carried out by the Walker A and Walker B motifs, respectfully. Walker A motifs have a looped structure, as revealed by crystal structures, with a consensus sequence of GXXXXGKT/S (where X can be any residue). The looped structure wraps around a nucleotide enabling the highly conserved lysine and threonine to bind to the γ-phosphate (Ramakrishnan, Dani and Ramasarma, 2002). The consensus sequence of the Walker B motif is DEAD/H. The first aspartic acid residue (underlined DEAD/H) coordinates Mg^{2+} to enable the binding of ATP. The glutamate residue (underlined DEAD/H) is essential for ATP hydrolysis (Caruthers and McKay, 2002) (Figure 3.2).

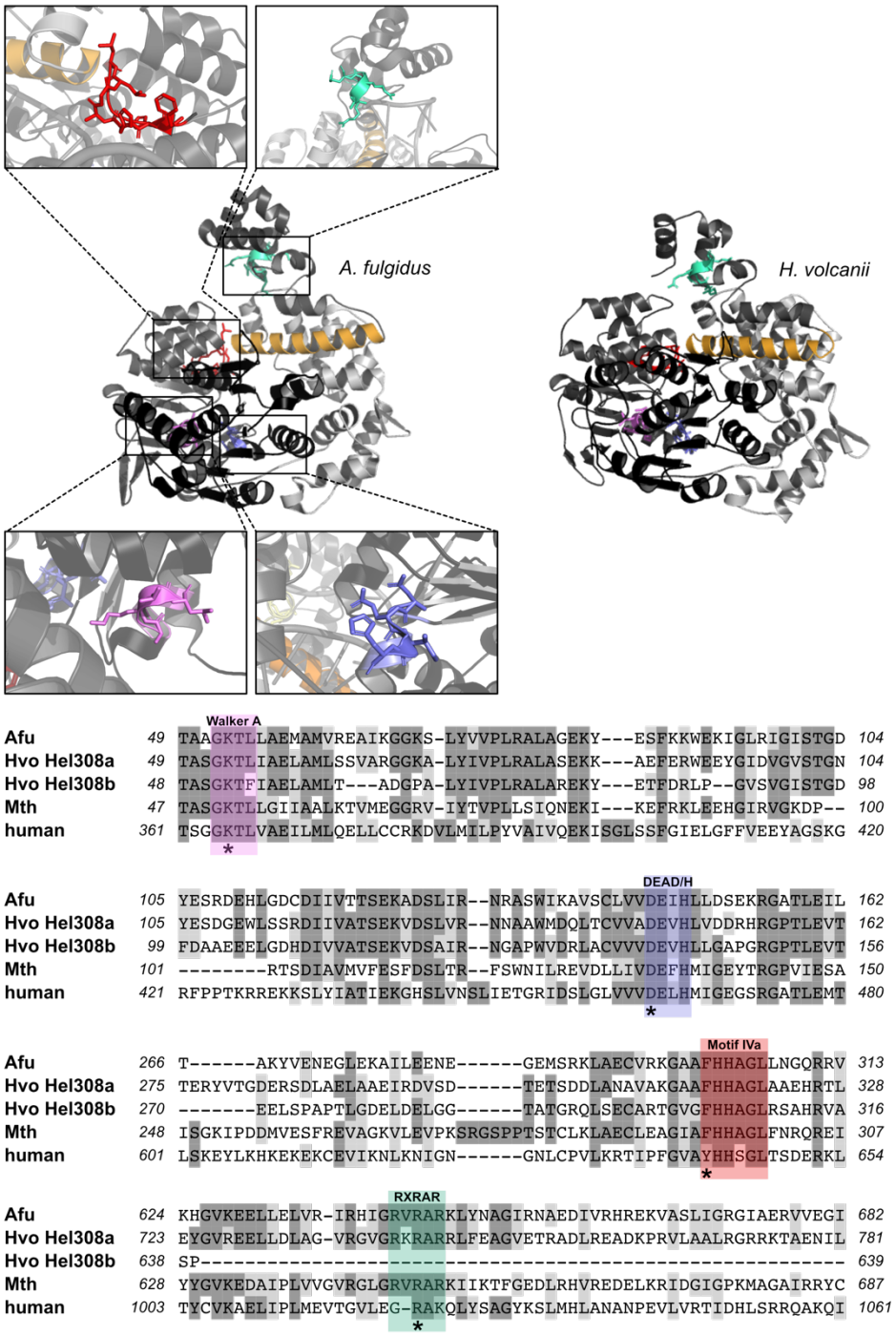


Figure 3.2: **Hel308 protein sequence alignment with the Hel308-DNA co-crystal structure *Archaeoglobus fulgidus* (Afu) and the predicted structures of Hel308a and Hel308b from *Haloferax volcanii* (Hvo) highlighting conserved residues examined.** Conserved motifs examined were the Walker A motif in Domain 1 (pink), the DEAD/H motif in Domain 1 (blue), Motif IVa in Domain 2 (red), and the triple arginine motif, RXRAR, in Domain 5 (green). * indicates the mutated residues in *Haloferax volcanii* (Hvo). Hel308-DNA co-crystal structure taken from Protein Data Bank 2P6R, DNA not shown (Buttner et al 2007). *Haloferax volcanii* Hel308a and Hel308b structures were predicted from the respective amino acid sequence using Phyre2 (Kelley et al 2015).

In Hel308-D145N (lab strain H2400), the negatively charged aspartic acid (D) was mutated to the polar uncharged asparagine (N). The coordination of Mg^{2+} requires negative charge of the aspartic acid side chain, the substitution with the uncharged glutamate side chain interferes with this organisation resulting in a loss of affinity for ATP. H2400 was generated by Rebecca Gamble-Milner (Gamble-Milner, 2016) and was confirmed via Southern blot in this study. A construct to introduce *hel308-K53R* mutation was made prior to this study (see 3.3.2 Materials: Plasmids for details). Both lysine (K) and arginine (R) residues are positively charged however, the bulky side chain of arginine is thought to sterically hinder the binding of ATP (Hishida *et al.*, 2002). For completion, this study generated a construct to introduce the *hel308-K53A* mutation. The introduction of point mutations in place of key residues within these motifs would uncouple ATP hydrolysis and helicase translocation along ssDNA. Thus, ATPase-null mutants will also be helicase-null (Figure 3.3).

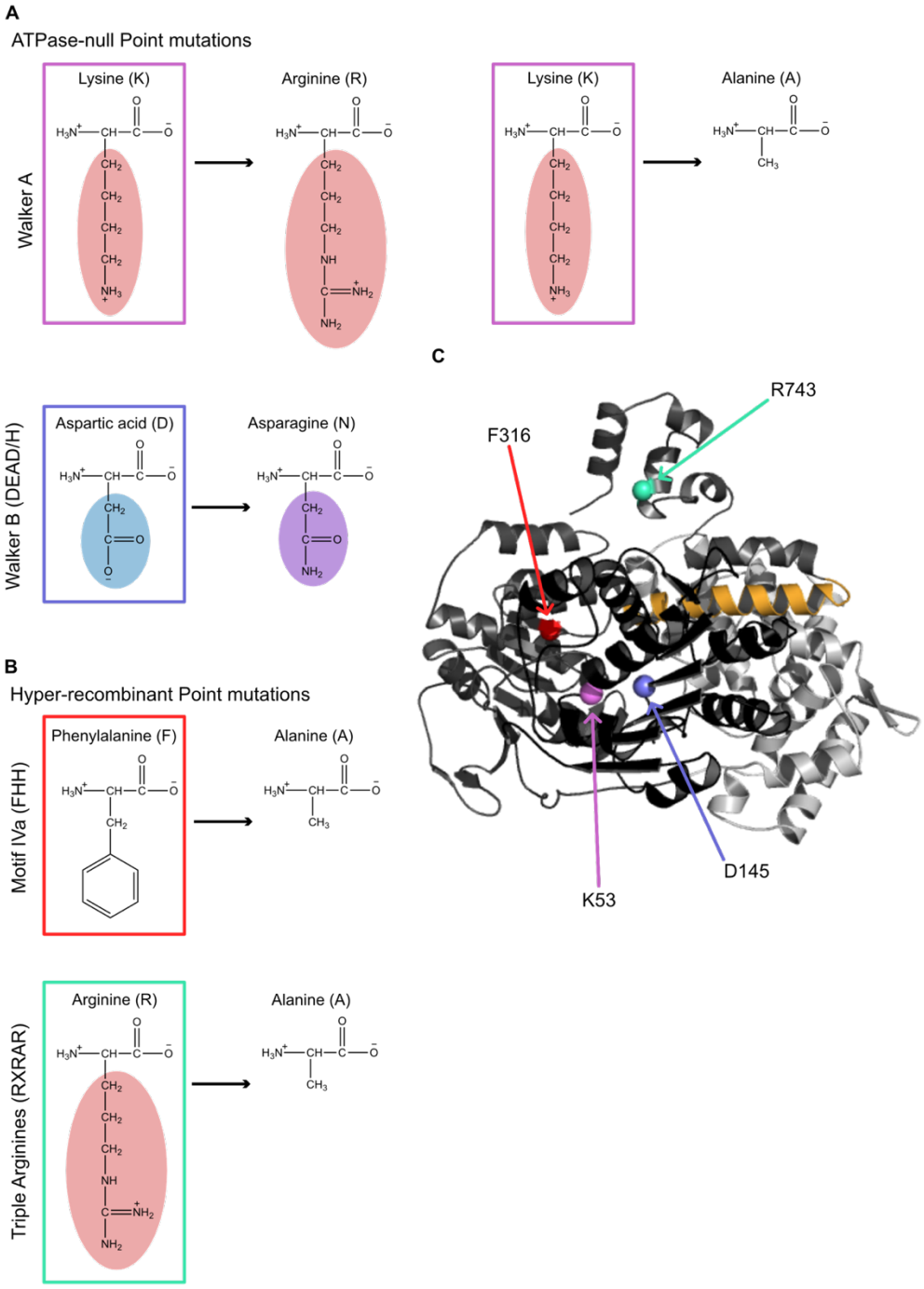


Figure 3.3: **Location of K53, D145, F316, and R743 in the predicted structure of Hel308a.** R groups shaded if positively charged (red), negatively charged (blue), and polar (purple). Hydrophobic R groups remain unshaded. A) The ATPase-null amino acid substitutions examined. Lysine (K53) was substituted for both arginine and alanine. Aspartic acid (D145) was substituted for asparagine. B) The hyper-recombinant amino acid substitutions examined. Phenylalanine (F316) was substituted for alanine. The central arginine residue (R743) was substituted for alanine. C) K53 is located in the RecA-fold of domain I, indicated with a pink sphere. D145 is located in the RecA-fold of domain I, indicated with a blue sphere. F316 is located in domain II, indicated with a red sphere. R743 is located in the HLH of domain V, indicated with a green sphere. *Haloferax volcanii* Hel308 structure predicted from the amino acid sequence using Phyre2 (Kelley et al 2015).

Hyper-recombinant mutants

F316A

Motif IVa (HHAGL) and the highly conserved 'FHH' linker are located within domain 2 and are responsible for coupling ATP hydrolysis to the so-called 'ratchet' in domain 4 (Büttner, Nehring and Hopfner, 2007). During Hel308a translocation, the 3' tail of unwound DNA is threaded through the helicase central pore to aid the separation of the DNA duplex. The *A. fulgidus* Hel308a-DNA co-crystal reveals that the linker residues F300 and H301, corresponding to the F316 and H317 residues in *H. volcanii*, line the surface of the helicase central pore in order to bind the passing 3' tail of unwound DNA (Figure 3.2). Previous genetic analysis in *H. volcanii* and biochemical analysis in *Methanothermobacter thermautotrophicus* has mutated the bulky hydrophobic phenylalanine (F) residue to the small hydrophobic alanine (A) residue. The difference in the side chain size is thought to interfere with inter-domain and DNA interactions. Genetic analysis of the F316A mutation in *H. volcanii* revealed a hyper-recombinant phenotype whilst biochemical analysis of the equivalent mutation in *M. thermautotrophicus* (F295A) revealed decreased ATPase activity and, counterintuitively, hyper-helicase activity (Gamble-Milner, 2016; Simmons, 2018)(Figure 3.3).

R743A

Archaeal Hel308 helicases contain a highly conserved triplet of arginine residues (RXRAR, where X can be any residue) located within domain 5. Domain 5 is thought to be an auto-inhibitory domain or 'molecular brake', responsible for coupling ATP hydrolysis to helicase unwinding. The *A. fulgidus* Hel308a-DNA co-crystal reveals that the central arginine, R644 (underlined RX<u>R</u>AR), binds the emerging 3' tail of unwound DNA via an interaction between the helix-loop-helix motif structure and the ssDNA phosphate-sugar backbone (Büttner, Nehring and Hopfner, 2007). By contrast, the flanking arginine residues, R642 and R646, form bridging contacts between domain 5 and the helicase ratchet domain 4 (Woodman, Briggs and Bolt, 2007). The recognition of replication forks by Hel308 might be dependent on the interaction between the RXRAR motif and ssDNA (Fujikane *et al.*, 2005) (Figure 3.2).

Substituting any of the arginine residues in the RXRAR motif for an alanine residue would have drastic effects on DNA binding, ATPase activity, and helicase activity (Woodman, Briggs and Bolt, 2007; Richards *et al.*, 2008; Woodman, Brammer and Bolt, 2011; Gamble-Milner, 2016). This amino acid substitution results in a loss of positive charge and therefore the loss in ability to interact with the negatively-charged phosphate-sugar backbone. Biochemical analysis in *M. thermautotrophicus* has revealed that mutating the first arginine

to an alanine (R647A, underlined RXRAR) increased ATPase activity, suggesting that this residue functions as part of a 'molecular brake', preventing ATP hydrolysis prior to the helicase successfully binding DNA. Furthermore, mutating both the first and final arginine residues (R647A and R651A, underlined RXRARR) abolishes helicase activity *in vitro*. Lastly, mutation of the central arginine (R649A, underlined RXRAR) resulted in DNA-binding defects and consequently, reduced DNA-stimulated ATP hydrolysis (Woodman, Briggs and Bolt, 2007; Woodman, Brammer and Bolt, 2011). However, in *Sulfolobus solfataricus* the mutation of the central arginine (R662A, underlined RXRAR) has little to no affect on ssDNA binding. Furthermore, the deletion of the entirety of domain 5 resulted in a binding affinity only slightly weaker than that of the wild-type helicase. Despite their decreased DNA binding affinity, both mutant Hel308 helicases experienced an increase in helicase activity (Richards *et al.*, 2008). Genetic analysis in *H. volcanii* showed that mutating the central arginine (R743A, underlined RXRAR) gives rise to a hyper-recombinant phenotype, not unlike the F316A mutation (Gamble-Milner, 2016) (Figure 3.3).

Hel308 as an anti-recombinase

Homologous recombination involves the exchange of nucleotide sequences between near-identical DNA duplexes for which RadA, the archaeal recombinase, is essential. RadA forms a nucleoprotein filament on ssDNA with the assistance of RadB. The RadA nucleoprotein filament then binds dsDNA molecules in order to search for a region of homology (Wardell *et al.*, 2017). Once found, RadA catalyses strand invasion and D-loop formation (Kil, Baitin and Lanzov, 2000) (Figure 3.4).

In *Haloferax volcanii*, mutants deleted for the RadA paralogue *radB* exhibit a decrease in the level of homologous recombination to 5% of that seen in wild-type cells. However, a double deletion of *hel308* with *radB* restores, in part, the level of homologous recombination. These data suggest an anti-recombinase role for Hel308 (Gamble-Milner, 2016).

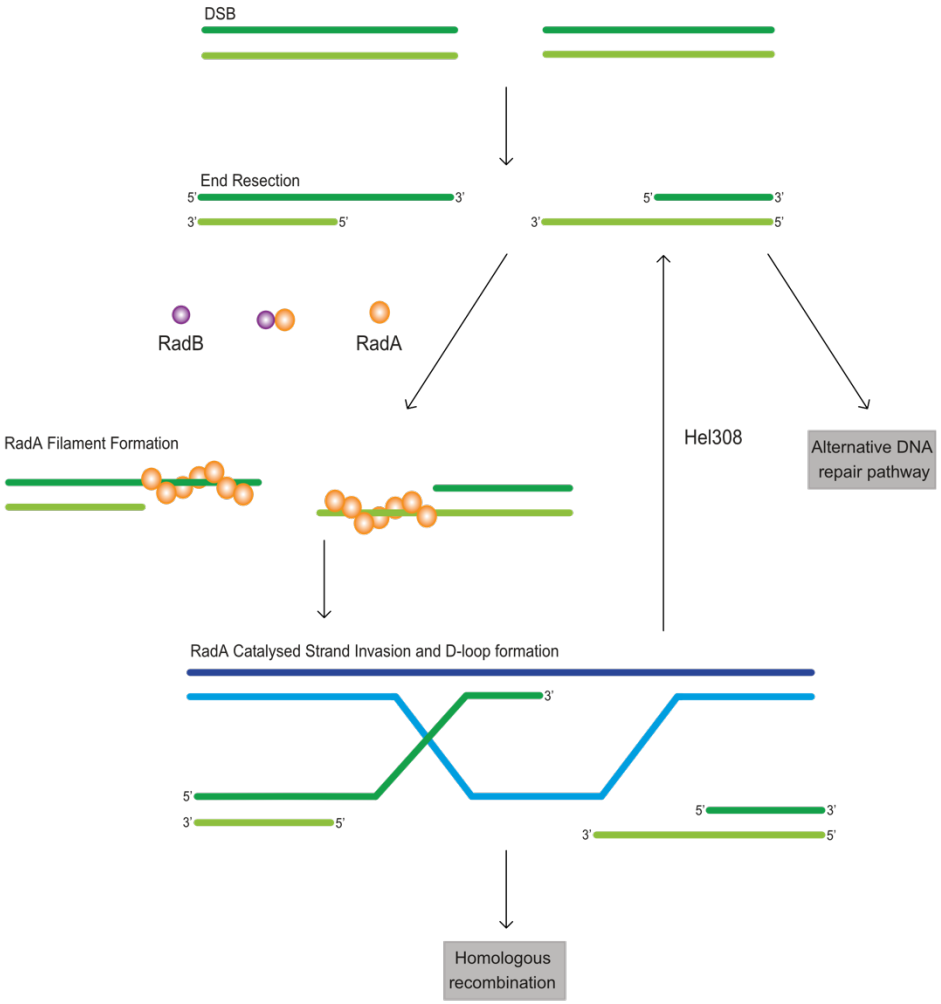


Figure 3.4: **Hel308 acts as an anti-recombinase.** DNA damage, such as double strand breaks, can be repaired by homologous recombination or by alternative DNA repair pathways. RadB promotes the homologous recombination pathway as well as assisting RadA nucleofilament formation on ssDNA. Hel308 could act as an anti-recombinase to disassemble D-loops and promote alternative DNA repair pathways.

Recombination, and consequently RadA, are essential for both the pop-in and pop-out steps of the gene deletion process (Woods and Dyall-Smith, 1997). Therefore, in order to delete *radA* from the genome, an episomal copy of the *radA* gene (pTA635, *radA*+ *mevR*+) must be present during the pop-out step in order to complement the chromosomal *radA* deletion. The *radA*+ complementing plasmid, pTA635, is later cured (Delmas *et al.*, 2009). Strains deleted for *radA* have a severe growth defect (Woods and Dyall-Smith, 1997) (Figure 3.5).

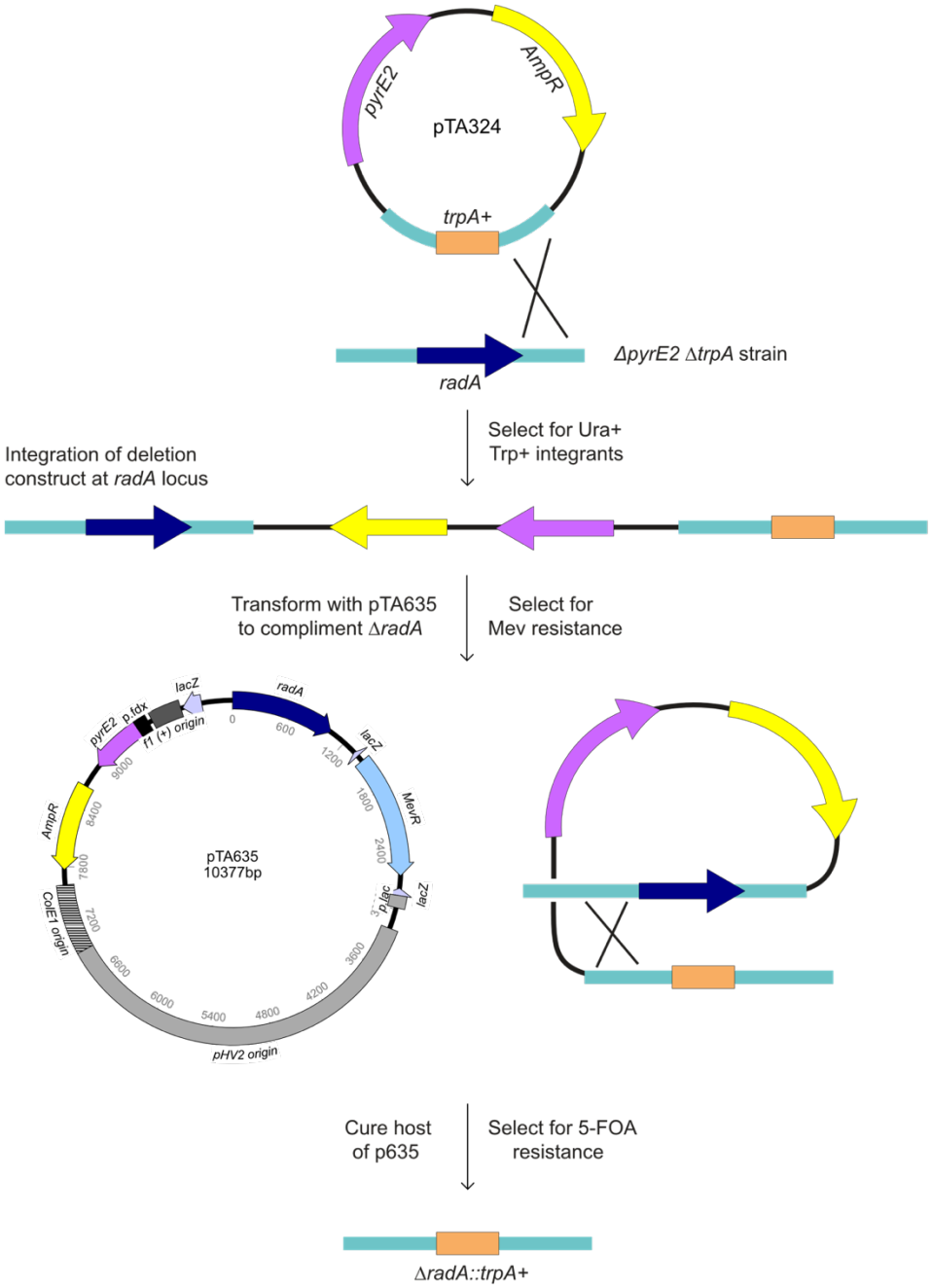


Figure 3.5: ***radA* gene deletion.** A Δ*pyrE2, Δtrp* strain is transformed with pTA324, a deletion construct containing a *radA* deletion (5' *radA* still present) and *pyrE2* marker (pop-in). The strain is then transformed with pTA635 (or pTA2080), episomal *radA+* (*radA-S101P+*) complementing plasmid containing *pyrE2* and *MevR* markers. Pop-out is able to occur due to the presence of *radA+* on pTA635 (*radA-S101P+* on pTA2080). Strains are selected for both the pop-out of TA324 and the loss of pTA635 by plating out on 5-FOA (Delmas et al 2009).

3.2 Aims and Objectives

It is of interest to further explore the role Hel308a plays in recombination. The objectives of this chapter were to:

- Generate strains deleted for *hel308, radA,* and *radB* in combination.
- Analyse the growth and determine generation time of strains deleted for *hel308, radA,* and *radB* in combination.
- Analyse the phenotypes of *hel308, radA,* and *radB* deletion strains after treatment with DNA damaging agents such as UV, which causes double-strand breaks, and MMC, an interstrand crosslinking agent.
- Generate ATPase-null and hyper-recombinant mutant *hel308* strains, individually and in combination.
- Analyse the growth and determine generation time of ATPase-null and hyper-recombinant mutant *hel308* strains.
- Analyse the phenotypes of ATPase-null and hyper-recombinant mutant *hel308* strains after treatment with DNA damaging agents such as UV and MMC.
- Analyse the recombination frequency of ATPase-null and hyper-recombinant mutant *hel308* strains, including the crossover: non-crossover product ratio.

3.3 Materials

3.3.1 Strains

Table 3.1 ***Haloferax volcanii*** **strains used in this chapter.**

Parent: H53 (*ΔpyrE2, ΔtrpA*)			
Strain	**Parent**	**Genotype**	**Notes**
H2085	H2071	*Δhel308*	Constructed by DW
H3515	H3281	*Δhel308 ΔradA::trpA+*	

Parent: H164 (*ΔpyrE2, bgaHa-Bb, leuB-Ag1, ΔtrpA*)			
Strain	**Parent**	**Genotype**	**Notes**
H2117	H2098	Δ*hel308::trpA+*	Constructed by RG-M
H2397	H2356	*hel308-F316A*	Constructed by RG-M
H2398	H2357	*hel308-R743A*	Constructed by RG-M
H2400	H2213	*hel308-D145N*	Constructed by RG-M
H2417	H2369	Δ*hel308::trpA+*, Δ*radB*	Constructed by RG-M
H3826	H3835	*hel308-D145N-F316A*	
H3827	H3833	*hel308-D145N-R743A*	
H4181	H1580	Δ*hdrB,* Δ*hel308*	

H4194	H4181	Δ*hdrB,* Δ*hel308 {p.tnaM3::7xhis-StrepII-StrepII-()+ pyrE2+ hdrB+}*
H4195	H4181	Δ*hdrB,* Δ*hel308 {p.tnaM3::7xhis-StrepII-StrepII-hel308+ pyrE2+ hdrB+}*
H4196	H4181	Δ*hdrB,* Δ*hel308 {p.tnaM3::7xhis-StrepII-StrepII-hel308-D145N+ pyrE2+ hdrB+}*
H4448	H4181	Δ*hdrB,* Δ*hel308 {p.tnaM3::hel308-K53A+ pyrE2+ hdrB+}*

DW = David Wallbank, RG-M = Rebecca Gamble-Milner

3.3.2 Plasmids

Table 3.2 Plasmids used in this chapter.

Name	Use	Notes
pBluescriptII SK+	Standard *E. coli* vector with functional blue/white screening capability	From Stratagene
pTA131	For making deletions in a *ΔpyrE2* background	Constructed by TA (Allers *et al.* 2004)
pTA163	Integrative vector containing *leuB-Ag2* and *pyrE2*. For use in recombination assays.	Constructed by GN
pTA324	*radA::trpA+* deletion construct.	Constructed by TA
pTA635	Episomal plasmid for *in trans* expression of *radA*	Constructed by TA
pTA1273	*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-D145N* gene replacement construct	Constructed by TA
pTA1642	*hel308-F316A* gene replacement construct	Constructed by RG-M
pTA1643	*hel308-R743A* gene replacement construct	Constructed by RG-M
pTA1840	*radA* overexpression construct used to generate probe for Southern hybridization	Constructed by DA
pTA1856	N-terminally His7-2xStrepII-tagged *hel308* gene replacement construct	
pTA1945	*hel308-K53R-F316A* gene replacement construct	

pTA1946	*hel308-K53R-R743A* gene replacement construct
pTA1947	*hel308-D145N-R743A* gene replacement construct
pTA1952	*hel308-D145N-F316A* gene replacement construct
pTA1988	*hel308-K53A* gene replacement construct
pTA1989	*hel308-K53A-F316A* gene replacement construct
pTA1990	*hel308-K53A-R743A* gene replacement construct
pTA2080	Episomal plasmid for *in trans* expression of *radA-S101P*
pTA2200	*p.tnaM3-hel308-K53A* expression vector

TA = Thorsten Allers, GN = Greg Ngo, RG-M = Rebecca Gamble-Milner, DA = Darya Ausiannikava.

pTA163 (Integrative vector containing *leuB-Ag2* and *pyrE2*)

The integrative vector, pTA163, contains *leuB-Ag2* and *pyrE2* for use in recombination assays and was constructed prior to this study by Greg Ngo in 2002 (Lestini, Duan and Allers, 2010) (Figure 3.6).

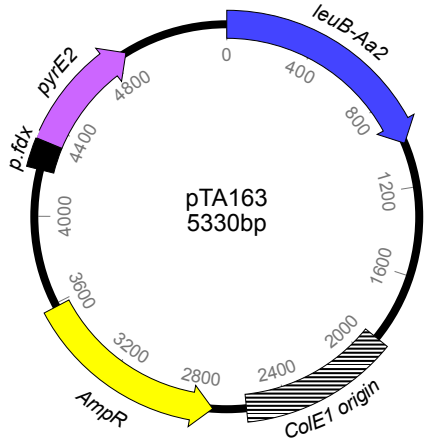


Figure 3.6: **pTA163.** pTA163 integrative vector containing *leuB-Ag2* and *pyrE2* for use in recombination assays. Constructed by Greg Ngo in 2002 (Lestini, Duan and Allers, 2010)

pTA298 *trpA* (HVO_0789)

The *trpA* genomic clone with flanking *Bam*HI sites, pTA298, was constructed prior to this study by Thorsten Allers in 2004 (Figure 3.7).

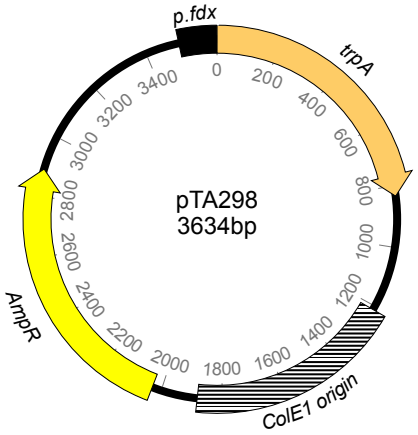


Figure 3.7: **pTA298.** The *trpA* genomic clone with flanking *Bam*HI sites. Constructed by Thorsten Allers in 2004

pTA324 (*radA::trpA*+ deletion construct)

The *radA::trpA*+ gene deletion construct, pTA324, was constructed prior to this study by Thorsten Allers in 2004 (Figure 3.8).

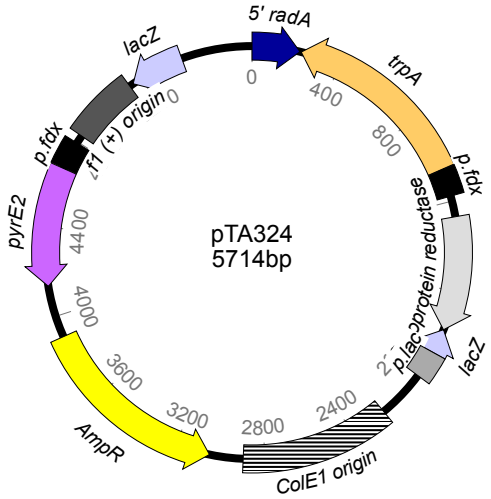


Figure 3.8: **pTA324.** The *radA::trpA*+ gene deletion construct, 5' *radA* still present. Constructed by Thorsten Allers in 2004

pTA635 (*radA in trans* expression construct)

The *radA in trans* expression construct, pTA635, was constructed prior to this study by Thorsten Allers in 2004 (Figure 3.9).

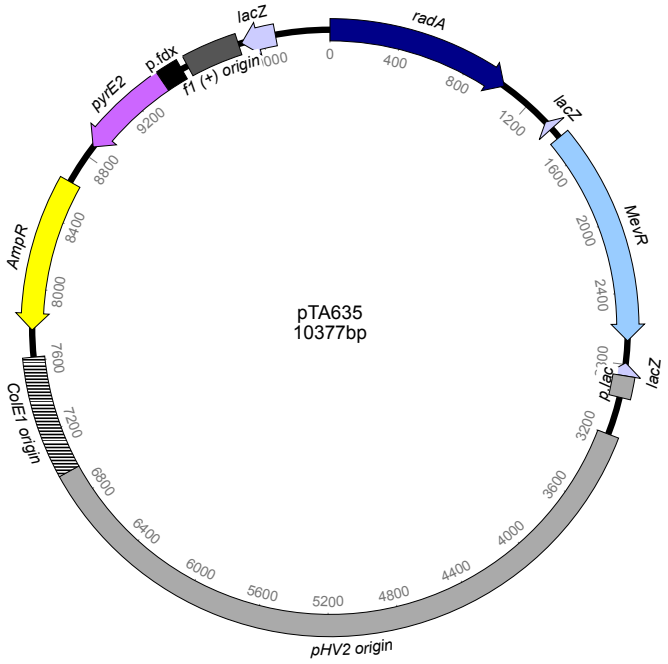


Figure 3.9: **pTA635.** *radA in trans*-expression construct. Constructed by Thorsten Allers in 2004

pTA1273 (*hel308::trpA*+ deletion construct)

The *hel308::trpA*+ gene deletion construct, pTA1273, was constructed prior to this study by Thorsten Allers in 2010 (Figure 3.10).

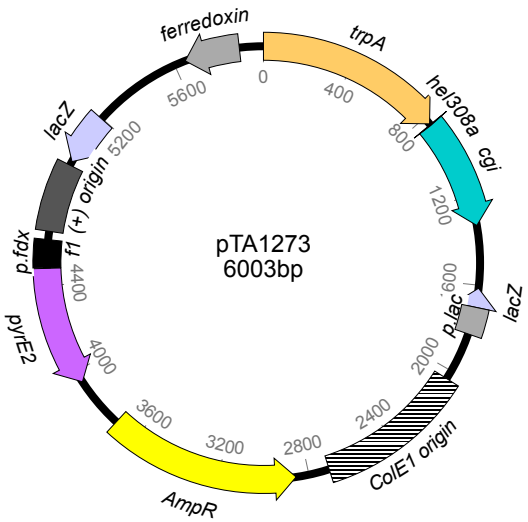


Figure 3.10: **pTA1273.** *hel308::tprA*+ gene deletion construct. Constructed by Thorsten Allers in 2010.

pTA1316 *hel308* (HVO_0014)

The *hel308* genomic clone, pTA1316, was constructed prior to this study by Thorsten Allers in 2011 (Gamble-Milner, 2016) (Figure 3.11).

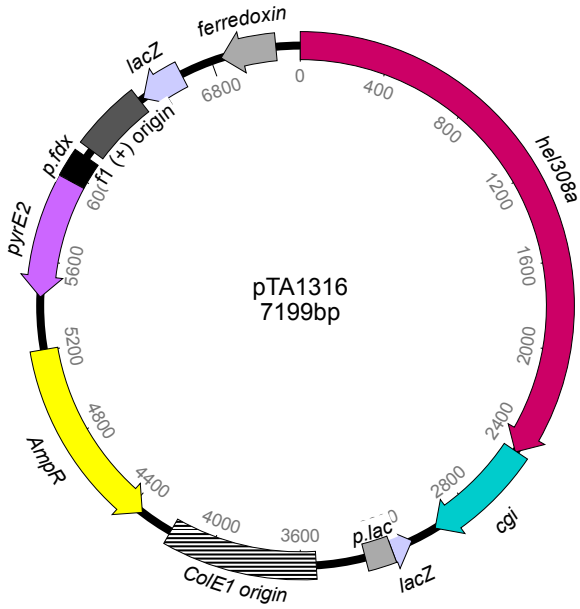


Figure 3.11: **pTA1316.** *hel308, cgi* genomic clone. Constructed by Thorsten Allers in 2011 (Gamble-Milner, 2016)

pTA1334 (*hel308-K53R* gene replacement construct)

The *hel308-K53R* gene replacement construct, pTA1334, was constructed prior to this study by Thorsten Allers in 2011 (Gamble-Milner, 2016) (Figure 3.12).

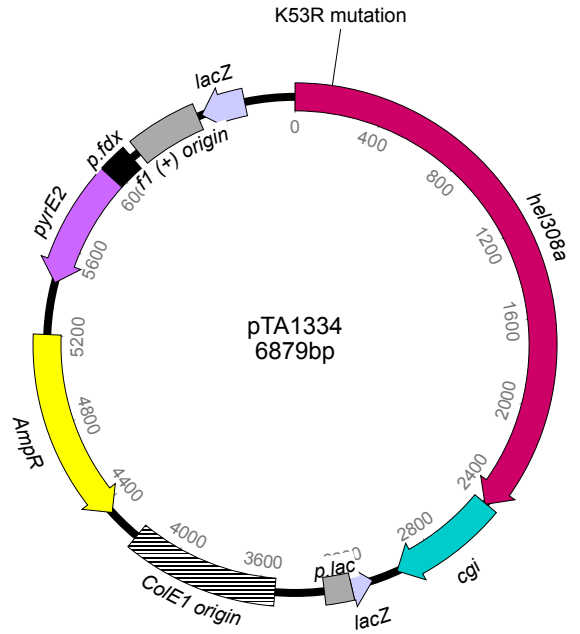


Figure 3.12: **pTA1334.** *hel308-K53R* gene replacement construct. Constructed by Rebecca Gamble-Milner in 2011 (Gamble-Milner, 2016)

pTA1335 (*hel308-D145N* gene replacement construct)

The *hel308-D145N* gene replacement construct, pTA1335, was constructed prior to this study by Thorsten Allers in 2011 (Gamble-Milner, 2016) (Figure 3.13).

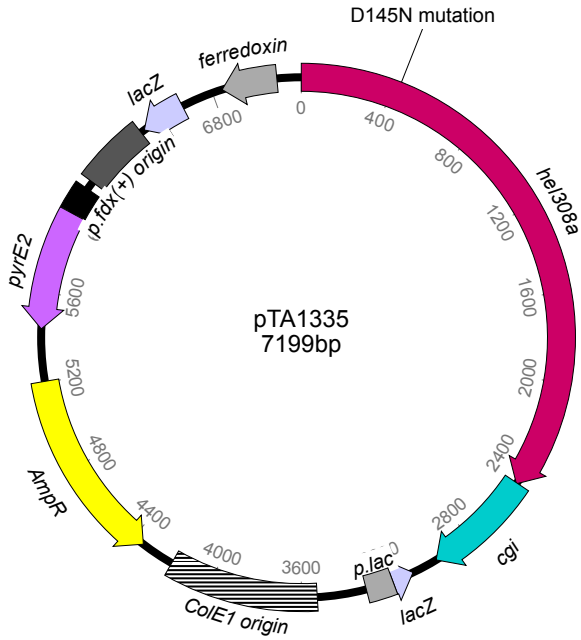


Figure 3.13: **pTA1335.** *hel308-D145N* gene replacement construct. Constructed by Rebecca Gamble-Milner in 2011 (Gamble-Milner, 2016)

pTA1485 (*radA-S101P* episomal expression construct)

The *radA-S101P* episomal expression construct, pTA1485, was constructed prior to this study by Antoine Doyen in 2014 (Figure 3.14)

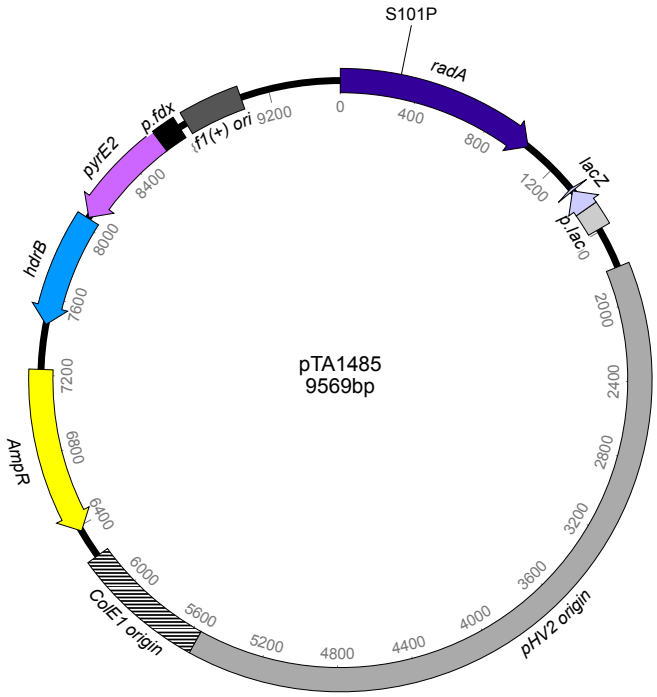


Figure 3.14: **pTA1485.** *radA-S101P* episomal expression construct. Constructed by Antoine Doyen in 2014.

pTA1530 (*radB* deletion construct)

The *radB* gene deletion construct, pTA1530, was constructed prior to this study by David Wallbank in 2014 (Figure 3.15).

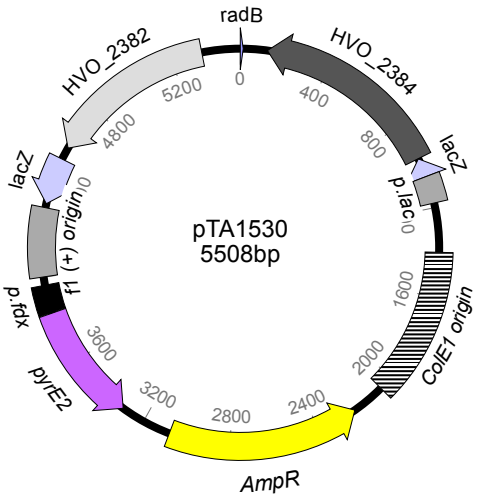


Figure 3.15: **pTA1530.** *radB* gene deletion construct. Constructed by David Wallbank in 2014.

pTA1642 (*hel308-F316A* gene replacement construct)

The *hel308-F316A* gene replacement construct, pTA1642, was constructed prior to this study by Rebecca Gamble-Milner in 2015 (Gamble-Milner, 2016) (Figure 3.16).

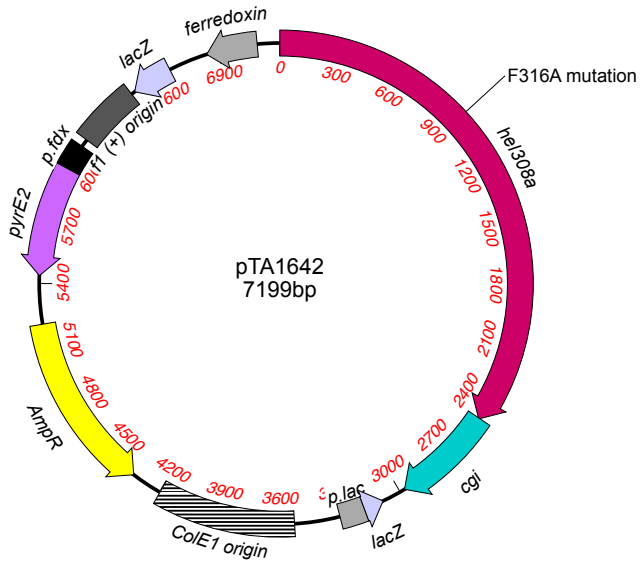


Figure 3.16: **pTA1642.** *hel308-F316A* gene replacement construct. Constructed by Rebecca Gamble-Milner in 2011 (Gamble-Milner, 2016)

pTA1643 (*hel308-R743A* gene replacement construct)

The *hel308-R743A* gene replacement construct, pTA1643, was constructed prior to this study by Rebecca Gamble-Milner in 2015 (Gamble-Milner, 2016) (Figure 3.17).

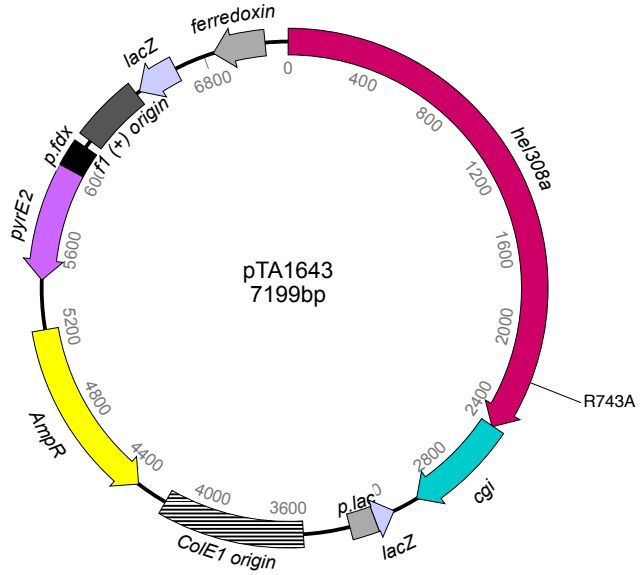


Figure 3.17: **pTA1643.** *hel308-R743A* gene replacement construct. Constructed by Rebecca Gamble-Milner in 2011 (Gamble-Milner, 2016)

pTA1840 (*radA* overexpression construct)

The *radA* overexpression construct, pTA1840, was constructed prior to this study by Darya Ausiannikava in 2016 (Figure 3.18)

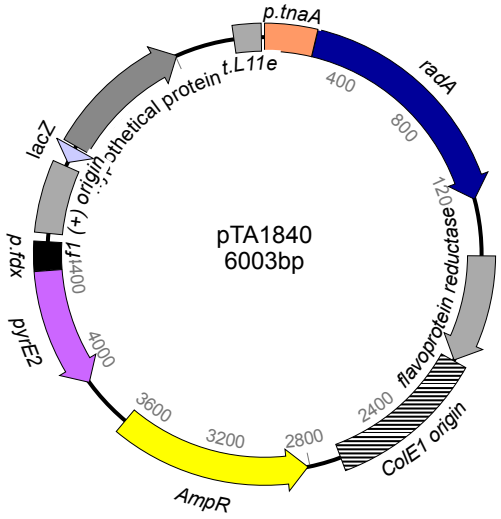


Figure 3.18: **pTA1840.** *radA* overexpression construct. Constructed by Darya Ausiannikava in 2016.

3.3.3 Oligonucleotides

Table 3.3 **Oligonucleotides used in this chapter.**

Name	Sequence (5'-3')	Description	Used to construct
pBSF2	TTAAGTTGGGTAACGCCAGGG	Amplification of the upstream region to multiple genes: *recJ2, recJ4;* and *hel308a*. Sequencing primer	pTA1923, pTA1951, pTA1991, K53A plasmids
Hel308 NdeI_F	GGGAGATcatATGCGAACTGCGGACCTGACGGGC	Incorporation of a *Nde*I site upstream of h*el308a* start codon: to clone downstream of a tryptophanase promoter	pTA1854
Hel308 SphI_F	GGGAGATGgcATGCGAACTGCGGACCTGACGGGC	Incorporation of a *Sph*I site upstream of h*el308a* start codon: to clone downstream of a tryptophanase promoter	pTA1849, pTA1850
helQ(R)BamHI02	CGAGGAtCCTCATTCGAAATCACCCAGACTGG	Amplification of *hel308a* and introduction of a *Bam*HI site	pTA1849, pTA1850, pTA1854
oriC-1trp-R	CGGCCGCTCTAGAACTAGTGG	Amplification of 2626bp region of pTA1425 containing *StrepII-hel308-6His* CDS	pTA1855

PtnaA Fint	GCCTGCCGATTA CTTCACATTCGC	Amplification of 2626bp region of pTA1425 containing *StrepII-hel308-6His* CDS	pTA1855
hel308 EcoR	AGGTAGTCGAGC ACGCGGTCC	Amplification of internal region of *hel308a*	pTA1952, pTA1965
Hel308 Fint	AGCGCTGGGAGG AGTACGGC	Amplification of internal region of *hel308a*	pTA1952, pTA1965
Hel308 F316A F	CGCGGCGgcCCA CCACGCGGGACT CGCCG	Generation of a *hel308-F316A* mutation	pTA1952, pTA1965
Hel308 F316A R	CGTGGTGGgcCGC CGCGCCTTTGGC GACC	Generation of a *hel308-F316A* mutation	pTA1952, pTA1965
Hel308 RTR	GTGTTGACGCCG GCGGCGAGCG	Amplification of internal region of *hel308a*	K53A plasmids
Hel308 K53AF	CGAGCGGgccGAC GCTCATCGCCGA ACTGGCG	Generation of a *hel308-K53A* mutation	K53A plasmids
Hel308 K53AR	CGATGAGCGTCg gcCCGCTCGCCGT CGGG	Generation of a *hel308-K53A* mutation	K53A plasmids
PrRad AF	CGAGGGCGACGA GAGCTACTGGTA CCCGC	Confirmation of Δ*radA* via PCR	
PrRad AR	AAACGACCGTCC CACCCGTGCCAC CGCC	Confirmation of Δ*radA* via PCR	

3.4 Results

3.4.1 Hel308 as an anti-recombinase

To further explore the role of Hel308 in homologous recombination, *hel308* was deleted in combination with the homologous recombination gene, *radA*, and the homologous recombination mediator gene, *radB.* The genes *hel308*, *radA,* and *radB* had all been individually deleted prior to the commencement of this study. Mutant strains deleted for both *hel308* and *radB* were also generated prior to this study.

3.4.1.1 Development of an improved *radA in trans* expression construct for deletion of *radA* in a strain deleted for *radB*

Deletion of *radB* reduces recombination to 5% of that in wild-type cells and the introduction of the S101P mutation in *radA* suppresses this phenotype (Wardell *et al.*, 2017) (Table 3.4). The use of a *radA-S101P+* complementing plasmid should assist the pop-out step in a *radB* deleted background by increasing the level of recombination.

radA *radB*	+ +	+ Δ	*S101P* Δ	*S101P* +
Recombination Frequency vs WT	1	0.018x	0.11x	1.36x
Generation time (h)	2	5.5	3.3	2

Table 3.4 **Recombination frequency and growth rate for radA-S101P strains.** *radA-S101P* alleviates, in part, both the recombination and growth defects of Δ*radB* strains. Figure adapted from Wardell et al., 2017

pTA2080 (*radA-S101P in trans* expression construct)

The *radA-S101P* episomal expression construct, pTA1485, was digested with the restriction enzymes *Kpn*I and *Bam*HI to produce a 1669bp *radA-S101P* fragment, which was inserted into the *Kpn*I and *Bam*HI sites of p635 to generate the *radA-S101P in trans* expression construct with the *MevR* and *pyrE2* markers, pTA2080 (Figure 3.19).

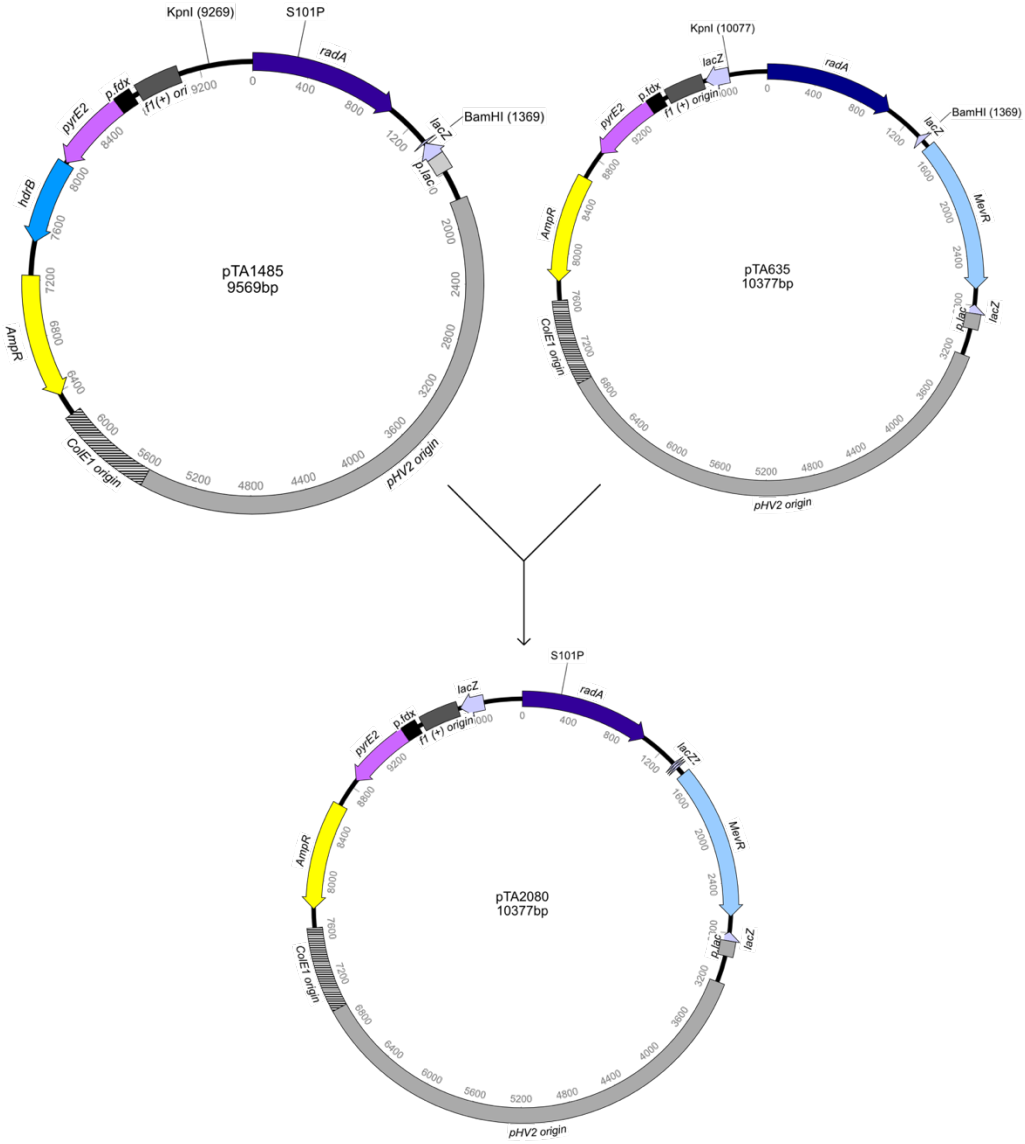


Figure 3.19: **Construction of *radA-S101P in trans*-expression construct, pTA2080.** pTA1485 was digested with *Kpn*I and *Bam*HI. The *radA-S101P* fragment was ligated into the *Kpn*I and *Bam*HI sites of p635 to generate the *radA-S101P in trans*-expression construct, pTA2080, with the *MevR* and *pyrE2* markers.

3.4.1.2 Generation of *hel308, radA,* and *radB* deleted strains

Δ*hel308* Δ*radA::trpA+*

The *radA::trpA+* deletion construct, pTA324, was introduced into the H2085 (Δ*pyrE2,* Δ*trpA,* Δ*hel308*) to generate the pop-in strain H3229. The episomal *radA+* complementing plasmid, pTA635, was then introduced into H3229, to generate the strain H3281. An amendment to the pop-out step was required. Strains were grown over two consecutive overnights in YPC +Mev to select for the *radA+* complementing plasmid, pTA635. Pop-out was allowed to occur due to the presence of *radA+* on pTA635 (and the presence of uracil in YPC, as usual). Strains were then grown for a further overnight in YPC only and plated on 5-FOA in order to select for both the pop-out of pTA324 and the loss of pTA635. Few colonies were obtained, all were restreaked and screened via Southern blot, two colonies of which looked correct. PCR was used as a second method to further verify the Δ*hel308* Δ*radA::trpA+* strain, H3515. (Figure 3.20).

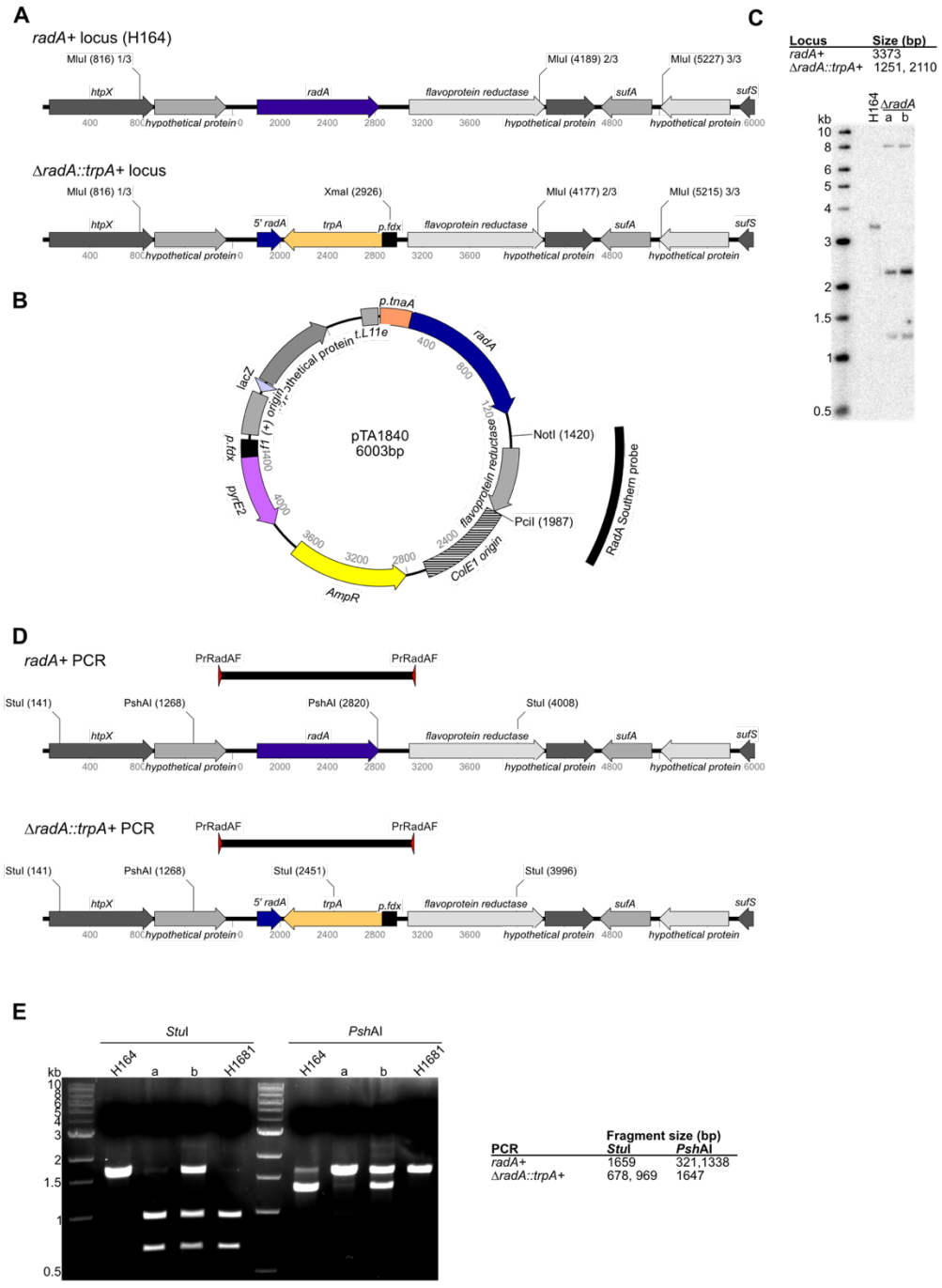


Figure 3.20: **Deletion of *radA* from H2085 (Δ*pyrE2,* Δ*trpA,* Δ*hel308*).** A) Allele for *radA*+ and Δ*radA::trpA*+. The *Mlu*I and *Xma*I sites used to digest genomic DNA are shown. B) A 567bp Δ*radA* probe for southern hybridisation was made by digesting pTA1840 with *Pci*I and *Not*I. C) Predicted fragment size and Southern blot, showing Δ*radA::trpA*+ candidates. D) *radA*+ and Δ*radA::trpA*+ PCR products from PrRadAF to PrRadAR. The *Psh*AI and *Stu*I sites used to digest the PCR product DNA are shown. E) Predicted fragment size (bp) of digested PCR product DNA and agarose gel, showing candidate 'a' is Δ*radA::trpA*+ like H1681 (Δ*radA::trpA*+, confirmed via Southern blot by Laura Mitchell in 2015), and candidate 'b' is merodiploid with fragment sizes corresponding to both *radA*+ and Δ*radA::trpA*+ genotypes. Candidate 'a' was assigned H3515.

Δ*hel308* Δ*radB* Δ*radA::trpA+*

In order to generate the triple deletion mutant (Δ*hel308* Δ*radB* Δ*radA::trpA+*), the *trpA* gene was first deleted from H1844 (Δ*pyrE2,* Δ*hel308,* Δ*radB*) using the Δ*trpA* construct, pTA98, to generate the strain H3696, which was confirmed via *trp*⁻ phenotype and colony hybridisation. The *radA::trpA+* deletion construct, pTA324, was then introduced into the H3696 (Δ*pyrE2,* Δ*trpA,* Δ*hel308,* Δ*radB*) to generate the pop-in strain H3840. The episomal *radA-S101P+* complementing plasmid, pTA2080, was then introduced into H3840, to generate the strain H4364. Further amendment to the pop-out step was required. Strains were grown over two consecutive overnights in YPC +Mev to select for the *radA+* complementing plasmid, pTA2080. Pop-out was allowed to occur due to the presence of *radA-S101P+* on pTA2080. Strains were then grown for a further two consecutive overnights in YPC only, followed by six continuous overnights in Ca +ura media, which lacks tryptophan, in an attempt to force all genome copies to adopt the Δ*radA::trpA+* genotype before plating on 5-FOA. Few colonies were obtained, all were restreaked and screened via Southern blot. All candidate strains were merodiploid (Figure 3.21).

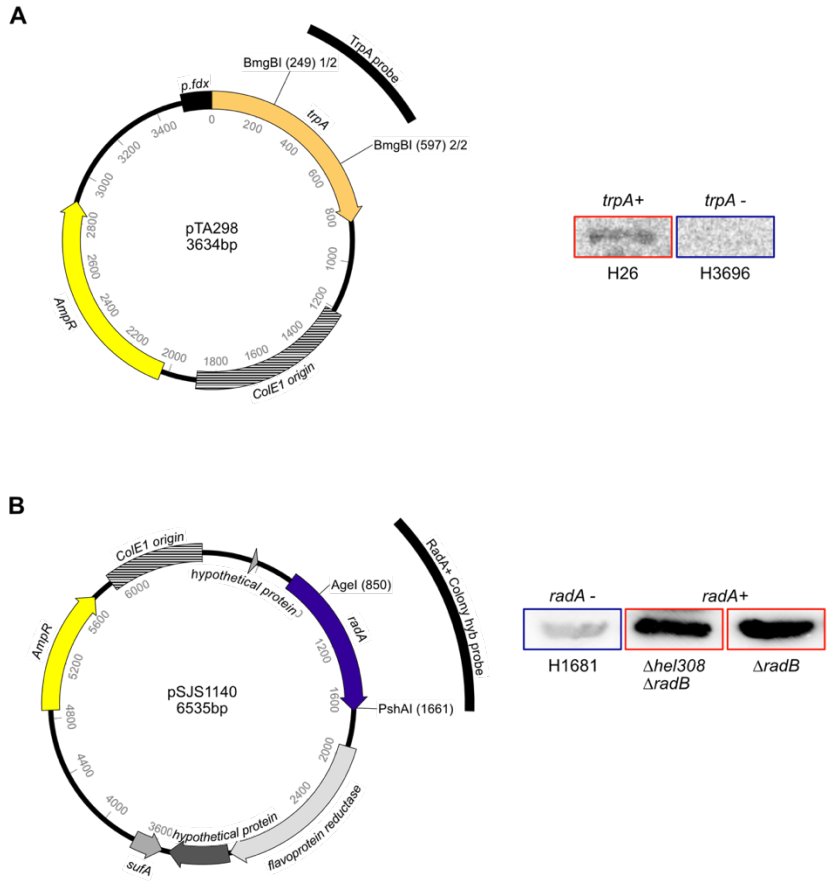


Figure 3.21: **Attempt to generate the *hel308, radB, radA* triple deletion mutant (Δ*pyrE2,* Δ*trpA,* Δ*hel308,* Δ*radB,* Δ*radA::trpA*+) and the *radB, radA* double deletion mutant (Δ*pyrE2,* Δ*trpA,* Δ*radB,* Δ*radA::trpA*+).** A) The 348bp Δ*trpA* probe for colony hybridisation was made by digesting pTA298 with *Bmg*BI, and the corresponding hyb. B) The 811bp Δ*radA* probe for colony hybridisation was made by digesting pSJS1140 with *Age*I and *Psh*AI, and the corresponding hyb, showing candidates for both the double and triple delete mutants are *radA*+.

Δ*radB* Δ*radA::trpA*+

In order to generate a strain deleted for *radB* and *radA,* the *trpA*+ marker was first removed from H3298 using the unmarked Δ*radB* construct, pTA1530, to generate the strain H4535, which was confirmed via *trp*⁻ phenotype. The *radA::trpA*+ deletion construct, pTA324, was then introduced into the H4535 (Δ*pyrE2,* Δ*trpA,* Δ*radB*) to generate the pop-in strain H4708. The episomal *radA-S101P*+ complementing plasmid, pTA2080, was then introduced into H4708, to generate the strain H4712. Strains were grown over two consecutive overnights in YPC +Mev to select for the *radA*+ complementing plasmid, pTA2080. Pop-out was allowed to occur due to the presence of *radA-S101P*+ on pTA2080. Strains were then grown for a further two consecutive overnights in YPC only, followed by six continuous overnights in Ca +ura in an attempt to force all genome copies to adopt the Δ*radA::trpA*+ genotype before plating on 5-FOA. Colonies were patched and screened via colony hybridisation using a

radA+ radiolabelled probe to check for *radA* deletions. Out of 157 colonies screened, no *radA* deletions in a H4535 (Δ*pyrE2,* Δ*trpA,* Δ*radB*) background were found (Figure 3.21, B).

3.4.1.3 Characterisation of *hel308* and *radA* deleted strains

Growth Rate

Standard growth rates in liquid media were determined for the Δ*radA* and Δ*hel308* strains. Strains were grown over two consecutive overnights to form a vigorously growing culture. The generation time in exponential phase of cell growth was calculated using an Epoch2 Microplate Spectrophotometer (BioTek) by measuring the A_{600} of the culture at 15 minute intervals over a 72-hour time course. A growth curve was plotted and the generation time during exponential phase for each strain was calculated (Figure 3.22).

In this experiment the wild-type strain H26 had a generation time of 2.75h. The strain H298 (Δ*radA*) had a severe fitness defect with a generation time of 7h. The strain H1391 (Δ*hel308*) also had a significant fitness defect with a generation time of 5.25h. Interestingly, the strain deleted for both *hel308* and *radA* (H3515) was healthier than either single deletion, with a generation time of 3.75h.

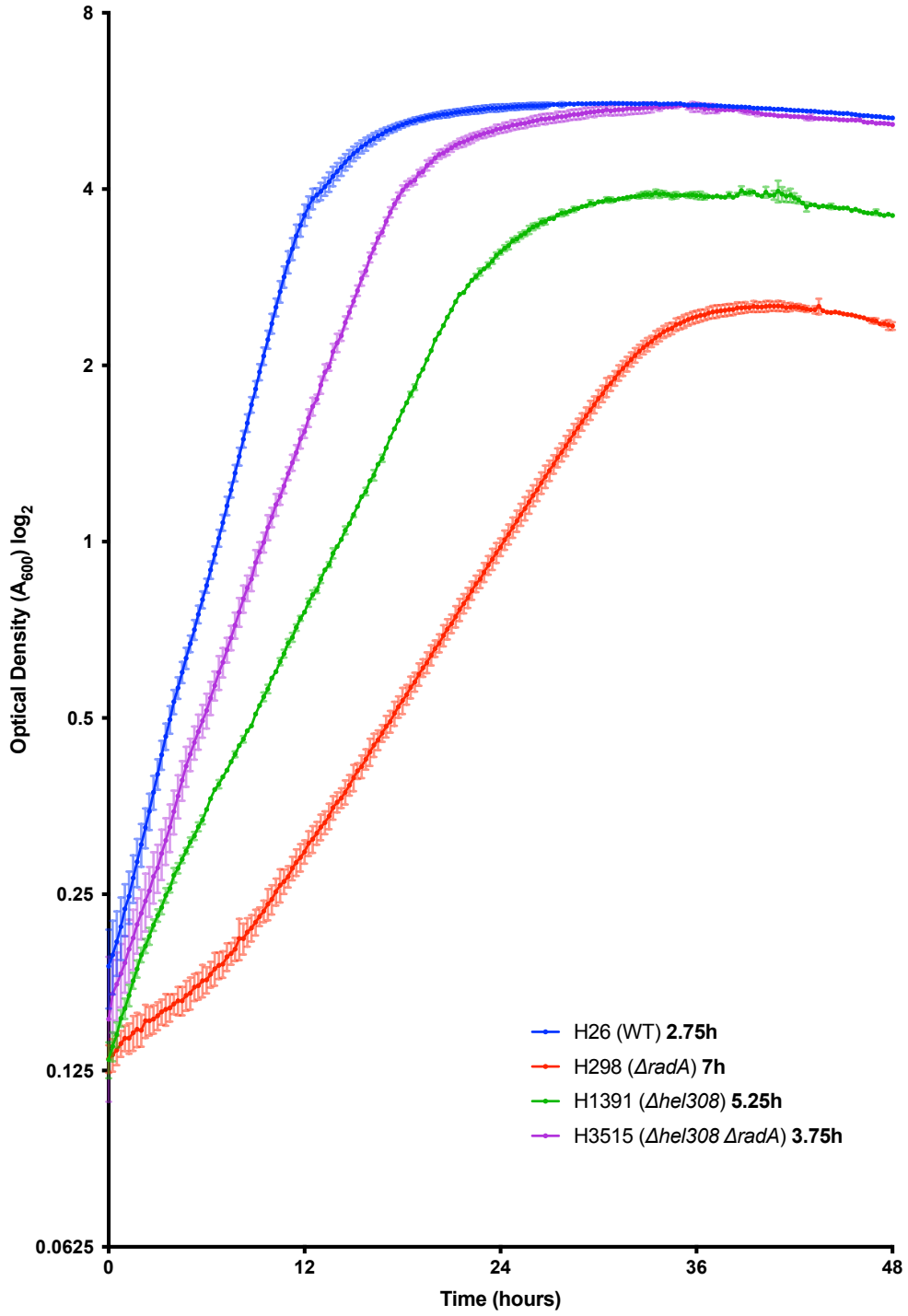


Figure 3.22: **Exponential growth rate of Δ*radA* and Δ*hel308* strains.** H26 (*radA+, hel308+*), H298 (Δ*radA*), H1391 (Δ*hel308*), and H3515 (Δ*hel308,* Δ*radA*) were grown in liquid culture for two consecutive overnights. The strains were then diluted and grown for 72 hours. Growth was measured by A_{600}. Generation time is indicated next to each strain. All strains were incubated on the same 96-well plate and measured simultaneously using an Epoch2 Microplate Spectrophotometer (BioTek). Data points are a mean of two technical repeats within two biological repeats. Standard deviation is shown.

Survival Following Treatment with DNA-damaging Agents

Strains deleted for *radA* are sensitive to MMC treatment and UV irradiation. Strains deleted for *hel308* are also sensitive to MMC treatment but to a lesser extent. Therefore, it was of interest to determine the response following UV irradiation or treatment with MMC for the *radA, hel308* double delete strain, H3515 (Figure 3.23).

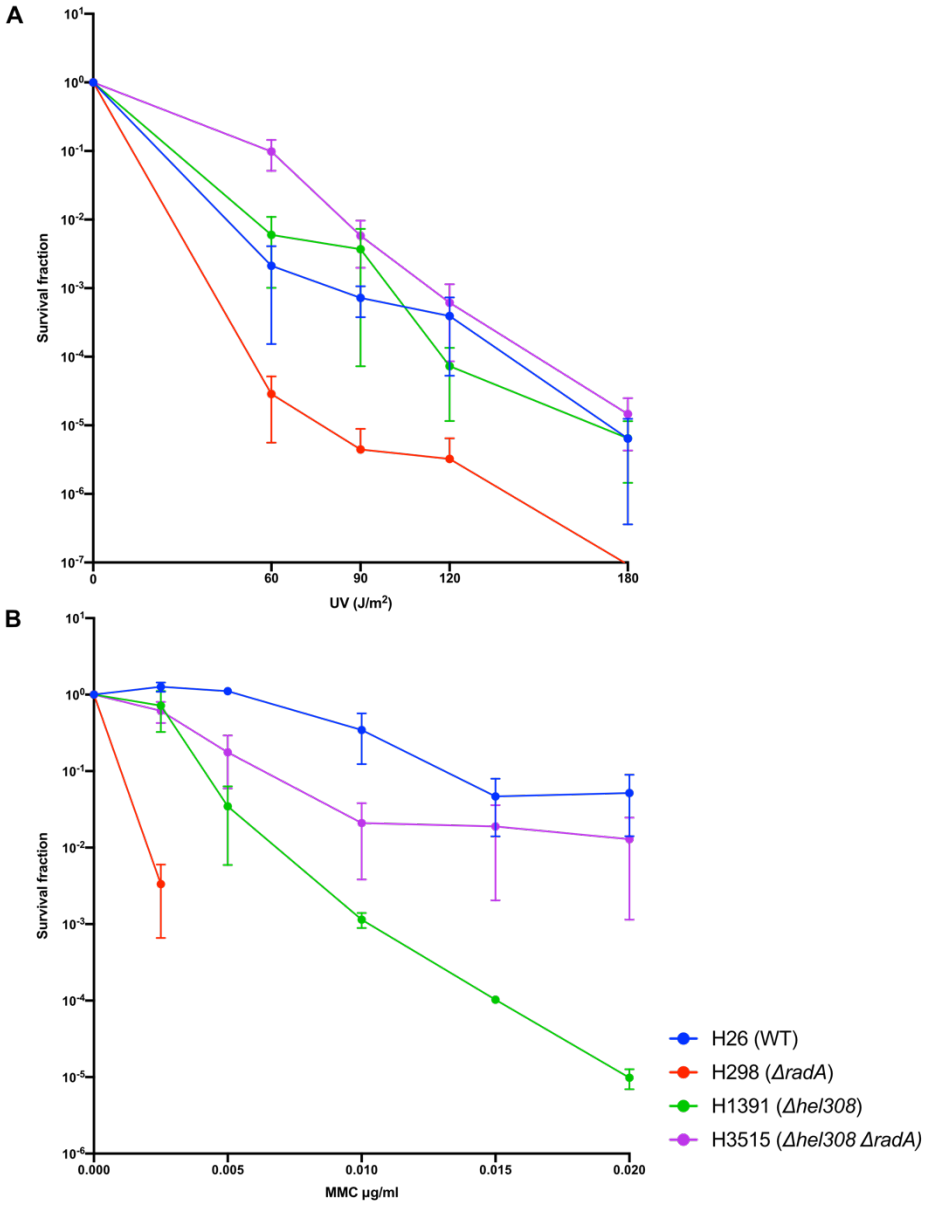


Figure 3.23: **Survival fraction of Δ*radA* and Δ*hel308* strains following treatment with DNA damaging agents.** H298 (Δ*radA*) has a significant decrease in survival fraction compared to wildtype H26 (*hel308+, radA+*) after UV irradiation (A) and treatment with MMC (B). H1391 (Δ*hel308*) has a decrease in survival fraction compared to wildtype H26 (*hel308+, radA+*) treatment with MMC only. No observed difference in survival fraction for H3515 (Δ*hel308,* Δ*radA*) compared to H26 after UV irradiation or treatment with MMC. Survival fraction is calculated relative to un-treated control. Each data point is generated as the mean of at least 3 independent trials. Standard deviation is shown.

There was no observed difference in survival fraction for H3515 (Δ*hel308,* Δ*radA*) compared to H26 (WT) after UV irradiation. MMC treatment did cause more cell death in the H3515 strain compared to the wildtype however, the difference was not significant. These data agree that the *radA, hel308* double delete strain, H3515, is healthier than either single deletion strain, H298 (Δ*radA*) or H1391 (Δ*hel308*).

3.4.2 Separation of function Mutations in Hel308

3.4.2.1 Generation of ATPase-null and hyper-recombinant mutant *hel308* strains

Walker A

K53R

The *hel308-K53R* gene replacement construct, pTA1334, was previously introduced into H2117 (Δ*pyrE2, bgaHa-Bb, leuB-Ag1,* Δ*trpA,* Δ*hel308::trpA+*) to generate the *hel308-K53R* strain, H2399, by Rebecca Gamble-Milner (Gamble-Milner, 2016). However, due to time constraints, the strain was not fully verified during that study. Southern blot and colony PCR strategies were used to verify H2399 but the data revealed that H2399 was incorrect (Figure 3.24).

Attempts to generate the *hel308-K53R* recombination strain were made using the original pop-in strain, H2212 (generated by Rebecca Gamble-Milner) and by repeating the transformation of H2117 (Δ*pyrE2, bgaHa-Bb, leuB-Ag1,* Δ*trpA,* Δ*hel308::trpA+*) with pTA1334. However, neither strategy produced a *hel308-K53R* strain. Replacing the *trpA+* marked *hel308* deletion allows the relatively quick screening of pop-out colonies via *trpA* phenotype. A *hel308-K53R* colony should be *trp-*, but all colonies screened (200) reverted to a *trpA+* phenotype, suggesting that the *hel308-K53R* mutation is more detrimental to the cell than the deletion of the entire *hel308* gene.

Gene replacement constructs combining the K53R mutation and the hyper-recombinant point mutations, F316A and R743A, were also generated.

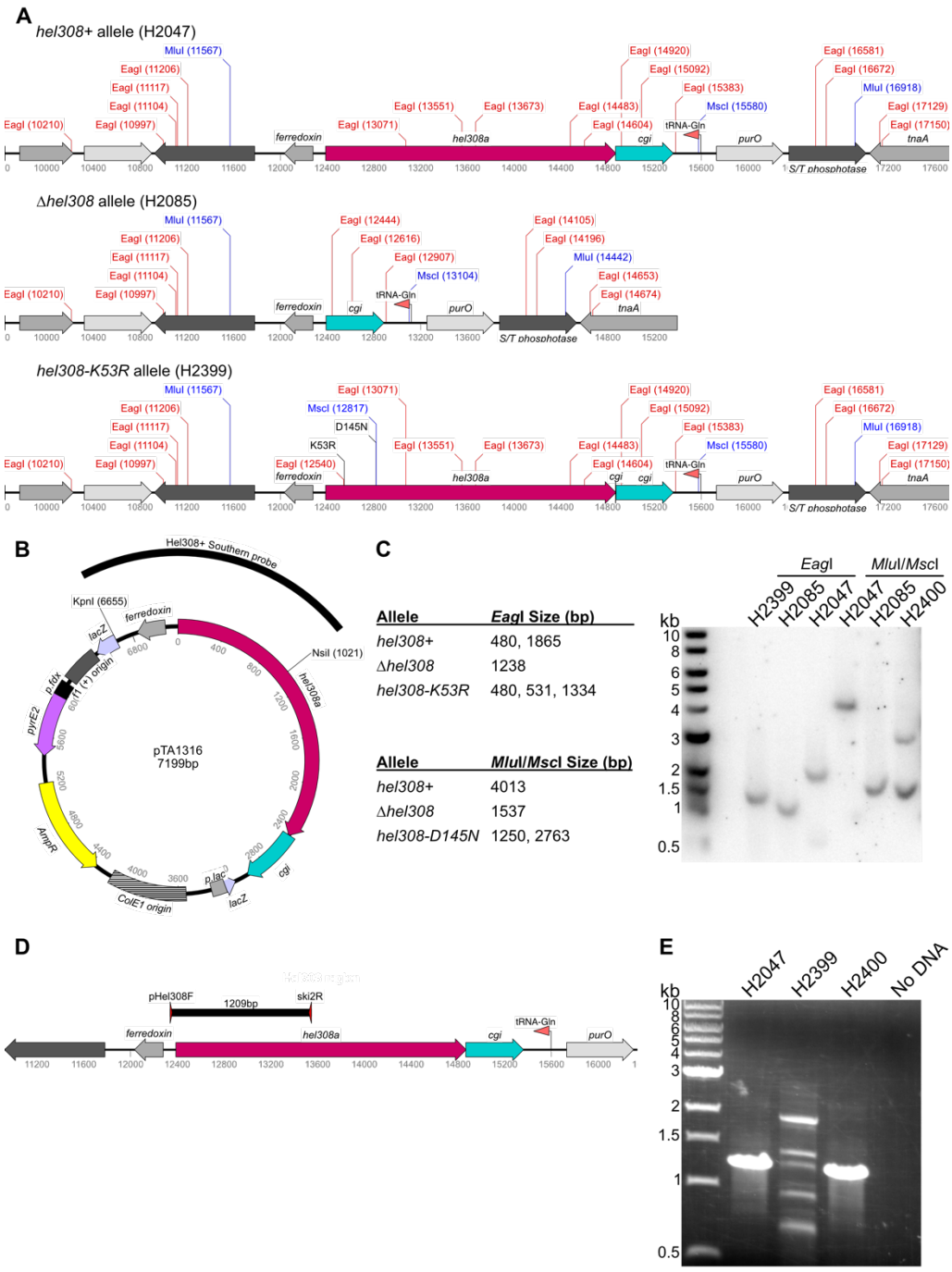


Figure 3.24 **Gene replacement of Δ*hel308::trpA*+ with *hel308-K53R* and *hel308-D145N* in H2117 (Δ*pyrE2, bgaHa-Bb, leuB-Ag1,* Δ*trpA,* Δ*hel308::trpA*+)**. A) Gene allele for *hel308+,* Δ*hel308,* and *hel308-K53R*. The *Eag*I sites used to digest genomic DNA for H2399 verification are shown in red. The *Mlu*I and *Msc*I used to digest genomic DNA for H2400 verification are shown in blue. B) A 1565bp *hel308+* probe for southern hybridisation was made by digesting pTA1316 with *Kpn*I and *Nsi*I. C) Predicted fragment size and Southern blot. D) 1208bp *hel308+* PCR product from pHel308F to ski2R. E) PCR agarose gel. Southern blot and PCR data show that H2400 is *hel308-D145N* however, H2399 was incorrect.

pTA1945 (*hel308-K53R-F316A* gene replacement construct)

The *hel308-F316A* gene replacement construct, pTA1642, was digested with the restriction enzymes *Bsp*EI and *Eco*RI, and ligated into the *Bsp*EI and *Eco*RI sites of pTA1334 to generate the *hel308-K53R-F316A* gene replacement construct, pTA1945. Construct confirmed by DNA sequencing (Figure 3.25).

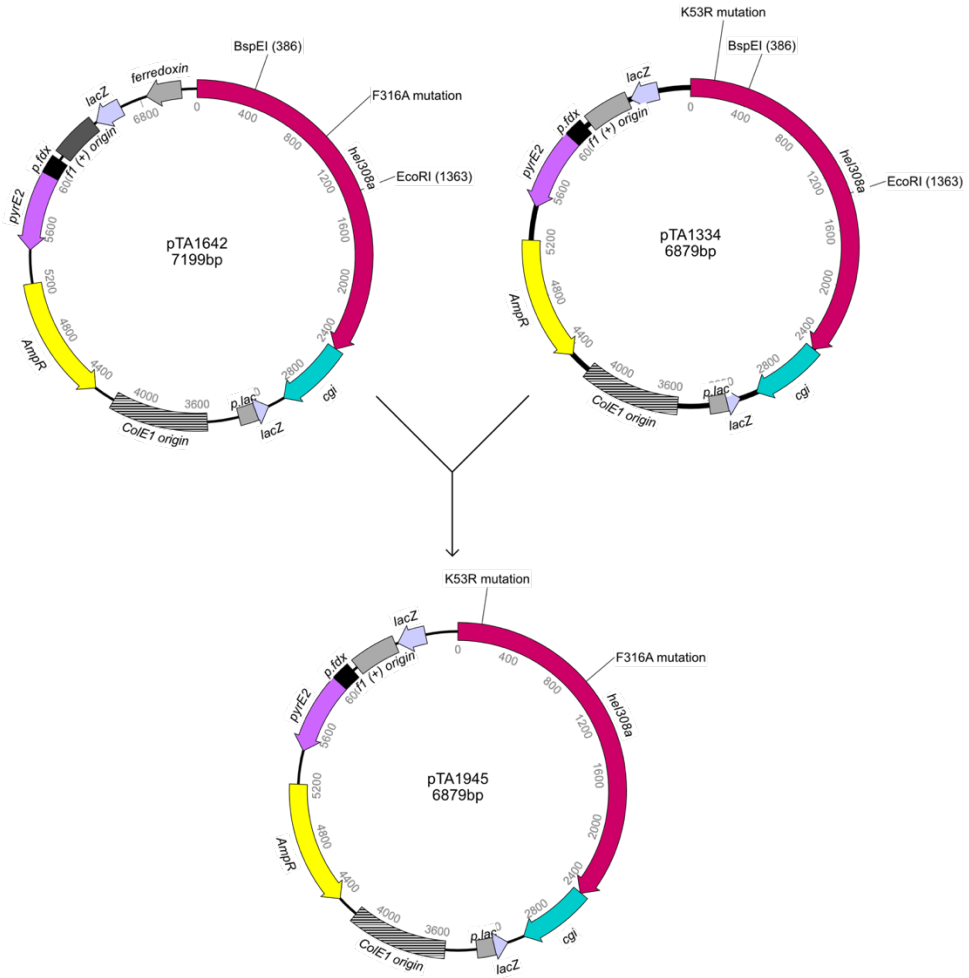


Figure 3.25: **Construction of *hel308-K53R-F316A* gene replacement construct, pTA1945.** pTA1642 was digested with *Bsp*EI and *Eco*RI, and ligated into the *Bsp*EI and *Eco*RI sites of pTA1334 to generate the *hel308-K53R-F316A* gene replacement construct, pTA1945.

pTA1946 (*hel308-K53R-R743A* gene replacement construct)

The *hel308-R7423A* gene replacement construct, pTA1643, was digested with the restriction enzymes *Blp*I and *Eco*RI, and ligated into the *Blp*I and *Eco*RI sites of pTA1334 to generate the *hel308-K53R-R743A* gene replacement construct, pTA1946. Construct confirmed by DNA sequencing (Figure 3.26).

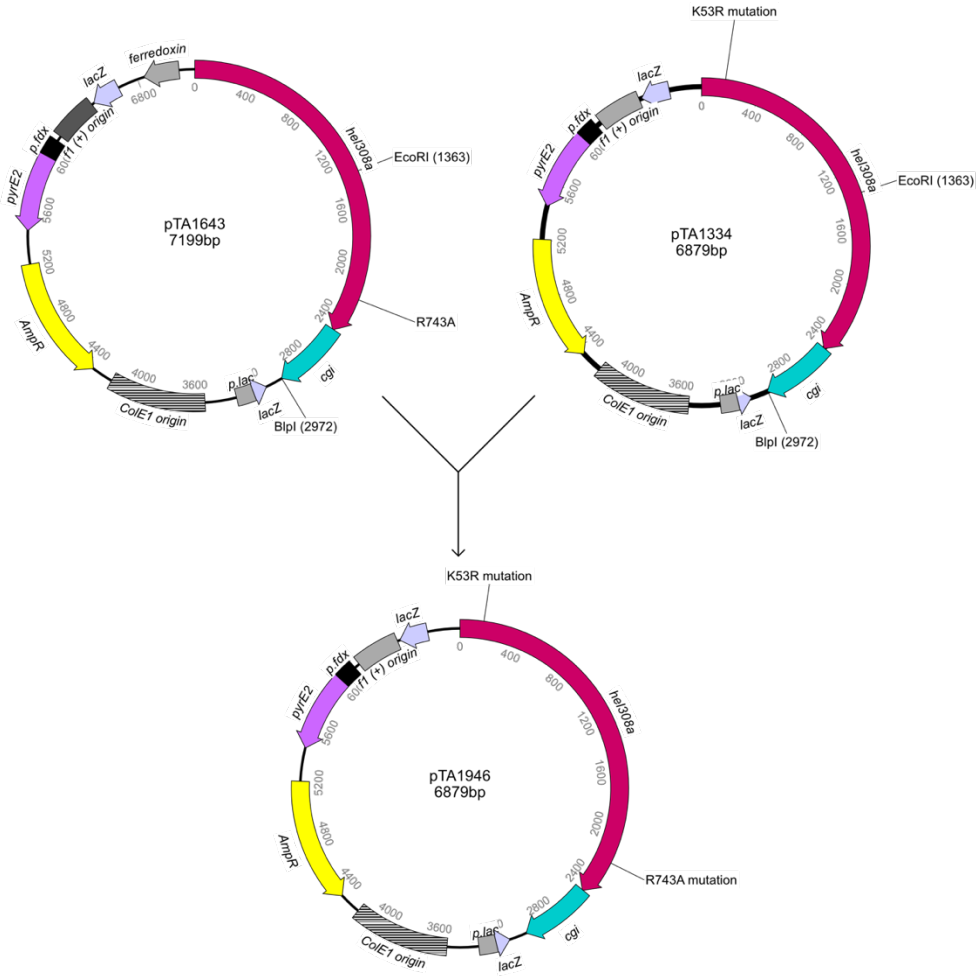


Figure 3.26: **Construction of *hel308-K53R-R743A* gene replacement construct, pTA1946.** pTA1643 was digested with *Blp*I and *Eco*RI, and ligated into the *Blp*I and *Eco*RI sites of pTA1334 to generate the *hel308-K53R-R743A* gene replacement construct, pTA1946.

pTA1945 and pTA1946 were introduced into H2117 (Δ*pyrE2, bgaHa-Bb, leuB-Ag1,* Δ*trpA,* Δ*hel308::trpA+*). However, like the single K53R mutation, all colonies screened (200 each) reverted to a *trp+* phenotype suggesting that the Hel308-K53R mutation is more detrimental to the cell than the deletion of the entire gene.

K53A

Both lysine (K) and arginine (R) residues are positively charged but the bulky side chain of arginine is thought to sterically hinder the binding of ATP (Hishida *et al.*, 2002). It is possible that this amino acid substitution prevents correct folding of Hel308 causing the protein to aggregate, explaining why the Δ*hel308::trpA+* genotype is preferential. In order to examine this theory, a construct to introduce the *hel308-K53A* mutation was generated. Gene

replacement constructs combining the K53A mutation and the hyper-recombinant point mutations, F316A and R743A, were also produced.

pTA1988 (*hel308-K53A* gene replacement construct)

The *hel308* genomic clone, pTA1316, was used as template DNA for overlap extension PCR with the primers pBSF2, Hel308K53AR, Hel308K53AF and Hel308RTR, to generate a *hel308* fragment containing the K53A point mutation. The fragment was digested with the restriction enzymes *Kpn*I and *Bsp*EI, and ligated into the *Kpn*I and *Bsp*EI sites of pTA1316 to generate the *hel308-K53A* gene replacement construct, pTA1988. Construct confirmed by DNA sequencing (Figure 3.27).

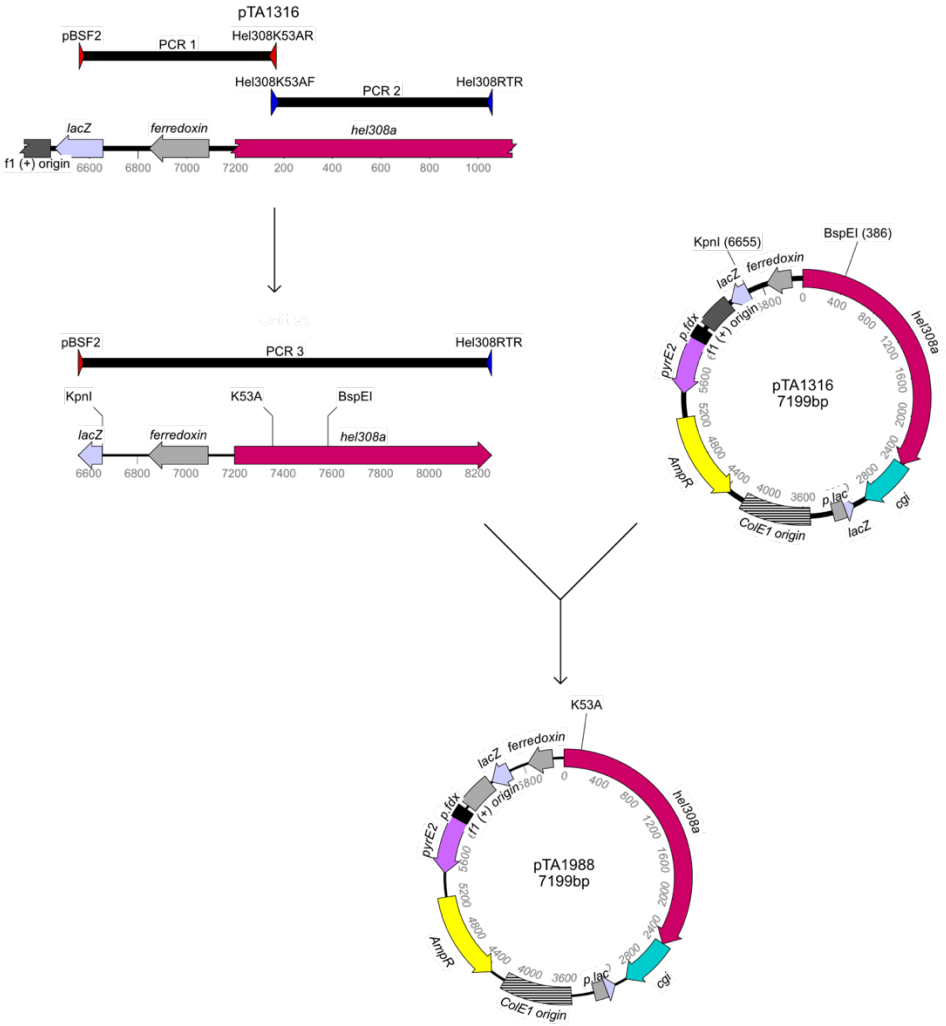


Figure 3.27: **Construction of *hel308-K53A* gene replacement construct, pTA1988.** pTA1316 was used as template DNA to PCR amplify from pBSF2 to Hel308K53AR to yield the PCR 1 product, and from Hel308K53AF to Hel308RTR to yield the PCR 2 product. PCR 1 and PCR 2 products were mixed in approximate equal proportion and underwent 10 cycles of PCR without primers. External primers, pBSF2 and Hel308RTR, were added for the last 20 cycles of PCR to generate a *hel308* fragment containing the K53A point mutation. The fragment was digested with the restriction enzymes *Kpn*I and *Bsp*EI, and ligated into the *Kpn*I and *Bsp*EI sites of pTA1316 to generate *hel308-K53A* gene replacement construct, pTA1988.

pTA1989 (*hel308-K53A-F316A* gene replacement construct)

The *hel308* genomic clone, pTA1316, was used as template DNA for overlap extension PCR with the primers pBSF2, Hel308K53AR, Hel308K53AF and Hel308RTR, to generate a *hel308* fragment containing the K53A point mutation. The fragment was digested with the restriction enzymes *Kpn*I and *Bsp*EI, and ligated into the *Kpn*I and *Bsp*EI sites of pTA1642 to generate *hel308-K53A-F316A* gene replacement construct, pTA1989. Construct confirmed by DNA sequencing (Figure 3.28).

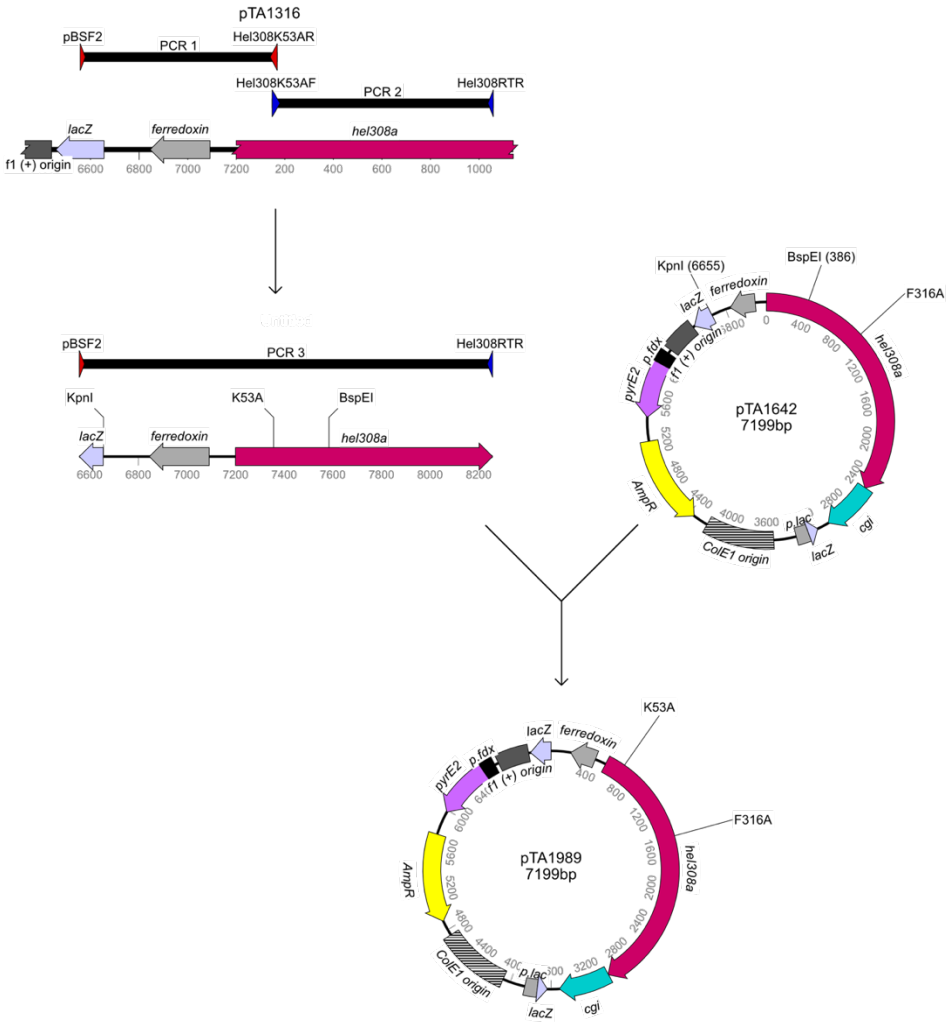


Figure 3.28: **Construction of *hel308-K53A-F316A* gene replacement construct, pTA1989.** pTA1316 was used as template DNA to PCR amplify from pBSF2 to Hel308K53AR to yield the PCR 1 product, and from Hel308K53AF to Hel308RTR to yield the PCR 2 product. PCR 1 and PCR 2 products were mixed in approximate equal proportion and underwent 10 cycles of PCR without primers. External primers, pBSF2 and Hel308RTR, were added for the last 20 cycles of PCR to generate a *hel308* fragment containing the K53A point mutation. The fragment was digested with the restriction enzymes *Kpn*I and *Bsp*EI, and ligated into the *Kpn*I and *Bsp*EI sites of pTA1642 to generate *hel308-K53A-F316A* gene replacement construct, pTA1989.

pTA1990 (*hel308-K53A-R743A* gene replacement construct)

The *hel308* genomic clone, pTA1316, was used as template DNA for overlap extension PCR with the primers pBSF2, Hel308K53AR, Hel308K53AF and Hel308RTR, to generate a *hel308* fragment containing the K53A point mutation. The fragment was digested with the restriction enzymes *Kpn*I and *Bsp*EI, and ligated into the *Kpn*I and *Bsp*EI sites of pTA1643 to generate *hel308-K53A-R743A* gene replacement construct, pTA1990. Construct confirmed by DNA sequencing (Figure 3.29).

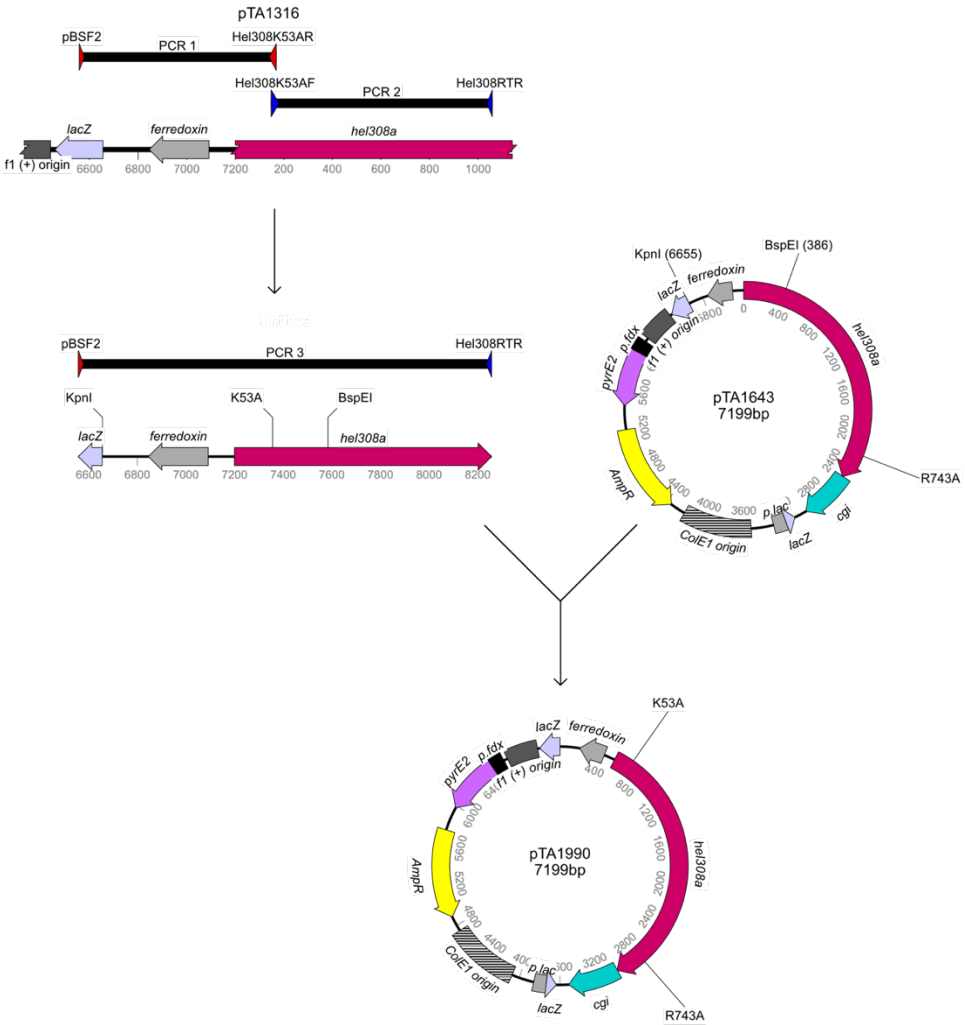


Figure 3.29: **Construction of *hel308-K53A-R743A* gene replacement construct, pTA1990.** pTA1316 was used as template DNA to PCR amplify from pBSF2 to Hel308K53AR to yield the PCR 1 product, and from Hel308K53AF to Hel308RTR to yield the PCR 2 product. PCR 1 and PCR 2 products were mixed in approximate equal proportion and underwent 10 cycles of PCR without primers. External primers, pBSF2 and Hel308RTR, were added for the last 20 cycles of PCR to generate a *hel308* fragment containing the K53A point mutation. The fragment was digested with the restriction enzymes *Kpn*I and *Bsp*EI, and ligated into the *Kpn*I and *Bsp*EI sites of pTA1643 to generate *hel308-K53A-R743A* gene replacement construct, pTA1990.

pTA1988, pTA1989, and pTA1990 were introduced into H2117 (Δ*pyrE2, bgaHa-Bb, leuB-Ag1,* Δ*trpA,* Δ*hel308::trpA*+). As with the K53R mutation, all colonies screened (200 each) reverted to a *trpA*+ phenotype, suggesting that the *hel308-K53A* mutation is more detrimental to the cell than the deletion of the entire *hel308* gene.

Lethality of the K53A/R mutants

Because neither the K53R nor the K53A point mutations in Hel308 appear to be tolerated by the cell, constructs for induced episomal expression were made to test the possibility that mutating the K53 residue is lethal to the cell. Whilst cloning attempts were first made in *E. coli, hel308-K53A/R* under control of the 'leaky' *p.tnaM3* promoter proved toxic to *E. coli* cells. Therefore, the episomal constructs were cloned directly into *H. volcanii.* Despite these efforts, the *p.tnaM3-hel308-K53R* expression vector could not be generated.

pTA2200 (*p.tnaM3-hel308-K53A* expression vector)

The *p.tnaM3-his*$_7$*-strepII-strepII-hel308* overexpression construct, pTA2121 (see Chapter 4: Biochemical analysis of Hel308 for more detail), was first digested with *Nde*I, blunt ended with Klenow, and then digested with *Nsi*I. The *hel308-K53A* gene replacement construct, pTA1988, was digested with *Psh*AI and *Nsi*I, and was ligated into the *Nde*I (blunt-ended) and *Nsi*I sites in pTA2121, to introduce the K53A point mutation whilst simultaneously removing the *his*$_7$*-strepII-strepII*-tandem tag. This generated the *p.tnaM3-hel308-K53A* overexpression construct, pTA2200, with *pyrE2*+ and *hdrB*+ markers.

The *p.tnaM3-hel308-K53A* overexpression construct, pTA2200, could not be generated in *E. coli*, the preferred organism for plasmid construction. Instead, pTA2200 was generated directly in H1209 (Δ*pyrE2,* Δ*hdrB, Nph-pitA,* Δ*mrr*) to produce the strain H4399. The strain was verified by DNA sequencing of the *p.tnaM3-hel308-K53A* episome.

pTA2200, extracted from H4399, was transformed into H4181 (Δ*pyrE2, bgaHa-Bb, leuB-Ag1,* Δ*hdrB,* Δ*hel308*) to generate the strain H4448 for analysis via a tryptophan gradient plate. Tryptophan is the inducer for the *p.tnaM3* promotor. A high concentration of tryptophan leads to high levels of expression of the *hel308-K53A* gene, if synthetically lethal this in turn leads to cell death. To generate a gradient of tryptophan across a plate, with the desired tryptophan concentration at one side fading to no tryptophan on the other side, plates were first poured with Hv-Ca +Ura +Trp (1.5 mM) on a slant to form a tapered wedge, once set the plate was placed flat and the wedge was covered with Hv-Ca + Ura agar, lacking tryptophan (Adapted from Bryson and Szybalski 1952). Strains

were grown over two consecutive overnights to form a vigorously growing culture until an A_{650} of 1.0. Serial dilutions of H26 (Δ*pyrE2*), H53 (Δ*pyrE2, ΔtrpA*), H4448 (Δ*pyrE2, bgaHa-Bb, leuB-Ag1, ΔhdrB, Δhel308, {p.tnaM3-hel308-K53A, pyrE2+, hdrB+}*), H4195 (Δ*pyrE2, bgaHa-Bb, leuB-Ag1, ΔhdrB, Δhel308, {p.tnaM3-his*$_7$*-strepII-strepII-hel308, pyrE2+, hdrB+}*), H4196 (Δ*pyrE2, bgaHa-Bb, leuB-Ag1, ΔhdrB, Δhel308, {p.tnaM3-his*$_7$*-strepII-strepII-hel308-D145N, pyrE2+, hdrB+}*), H4400 (Δ*pyrE2, ΔhdrB, Nph-pitA, Δmrr, {p.tnaM3-his*$_7$*-strepII-strepII-hel308-F316A, pyrE2+, hdrB+}*) in 18% SW to 10^{-4} were prepared. Autoclaved paintbrushes were first dipped in 18% SW. Using a fresh paintbrush for each strain, the paintbrush was dipped into the diluted culture and then painted in one direction across the gradient plate. Using the same paintbrush dipped again in the diluted culture, a second line was painted over the first in the opposite direction (Figure 3.30).

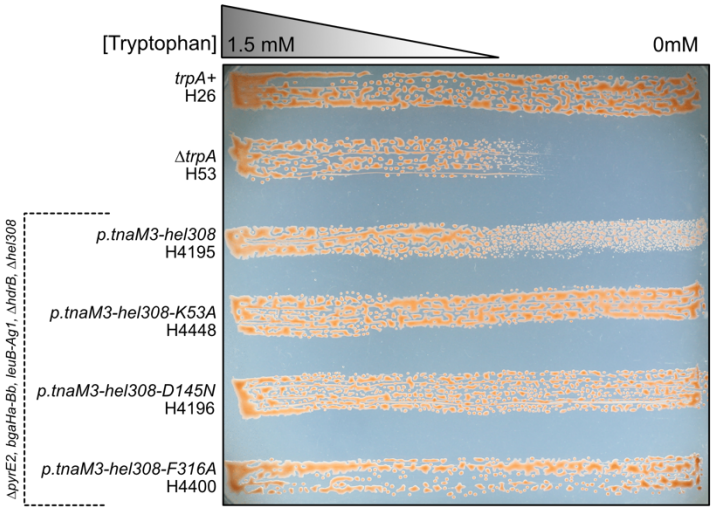


Figure 3.30: ***p.tnaM3-hel308*** **strains on tryptophan gradient plates.** H4195 (*p.tnaM3-his*$_7$*-strepII-strepII-hel308),* H4448 (*p.tnaM3-hel308-K53A*), H4196 (*p.tnaM3-his*$_7$*-strepII-strepII-hel308-D145N*), and H4400 (*p.tnaM3-his*$_7$*-strepII-strepII-hel308-F316A*) were painted onto tryptophan gradient plates with a tryptophan concentration of 1.5mM. Cells are viable across the whole gradient in all cases. H26 (*trpA+*) and H53 (Δ*trpA*) were painted onto tryptophan gradient plates to aid in visualisation of the gradient.

H26 (Δ*pyrE2*) and H53 (Δ*pyrE2, ΔtrpA*) confirmed the tryptophan gradient was present, with H53 unable to grow in the absence of tryptophan. H4448 {K53A}, H4195 {WT}, H4196 {D145N}, H4400 {F316A} all grew similar to H26 suggesting that the K53A mutation is not lethal to the cell. To confirm that the *p.tnaM3-hel308-K53A* overexpression construct was present, pTA2200 was extracted from H4448 isolated from the tryptophan gradient plate and DNA sequenced revealing that the episome had spontaneously lost ~1000bp of the *hel308* CDS including the K53A mutation. Due to time constraints the correct strain for this episomal expression analysis was not generated during this study.

Combination of ATPase-null and hyper-recombinant point mutations in *hel308.*

The levels of recombination in strains containing either *hel308-F316A* or *hel308-R743A* alleles are increased a hundred-fold compared to wild-type and are therefore referred to as "hyper-recombinant mutations". However, these point mutants are not deficient in DNA repair or normal growth. By contrast, ATPase-null mutants are slow-growing and sensitive to DNA cross-linking agents, similar to Δ*hel308* strains (Gamble-Milner, 2016). This suggests that the hyper-recombinant point mutations identify a separation of function of Hel308, where the ability to bind D-loop intermediates of recombination is affected, but the helicase activity involved in DNA repair is not. To further explore this, the ATPase-null mutation D145N was combined with the hyper-recombinant mutations.

pTA1947 (*hel308-D145N-R743A* gene replacement construct)

The *hel308-R7423A* gene replacement construct, pTA1643, was digested with the restriction enzymes *Blp*I and *Eco*RI, and ligated into the *Blp*I and *Eco*RI sites of pTA1335 to generate the *hel308-D145N-R743A* gene replacement construct, pTA1947. Construct confirmed by DNA sequencing (Figure 3.31).

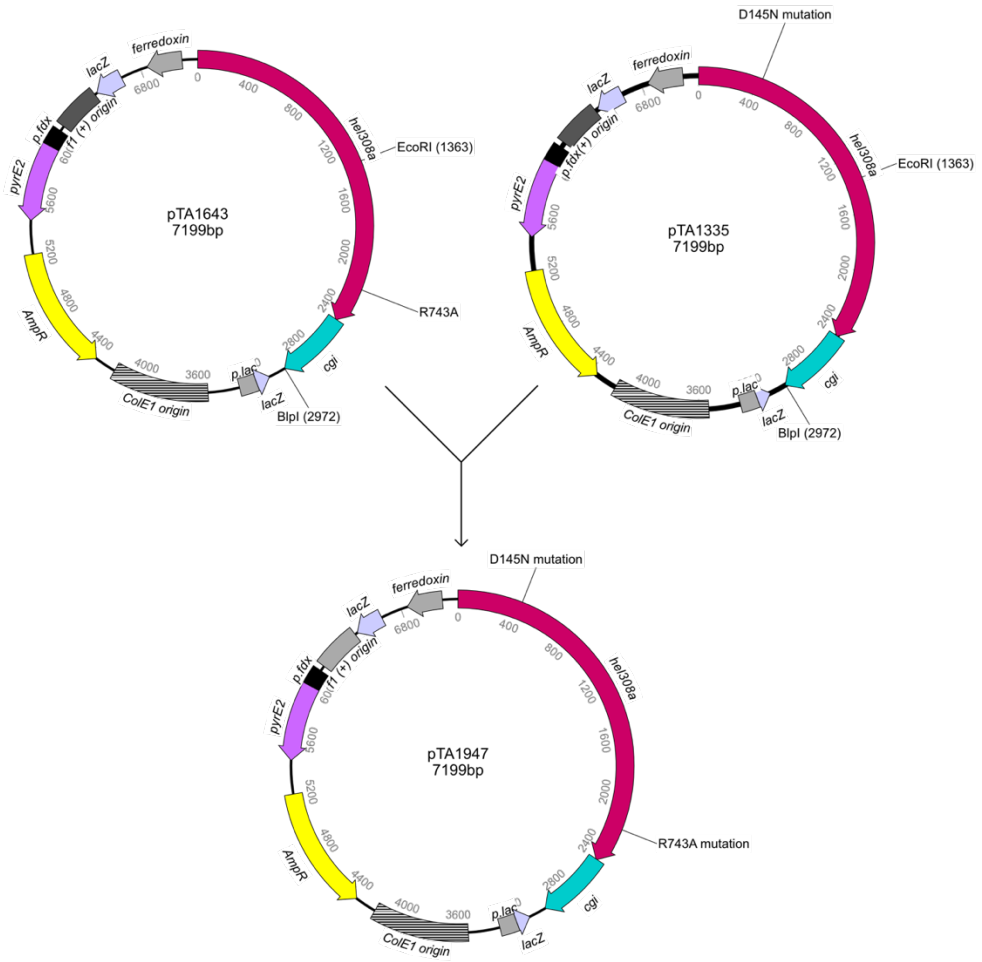


Figure 3.31: **Construction of *hel308-D145N-R743A* gene replacement construct, pTA1947.** pTA1643 was digested with *Blp*I and *Eco*RI, and ligated into the *Blp*I and *Eco*RI sites of pTA1335 to generate the *hel308-D145N-R743A* gene replacement construct, pTA1947.

The *hel308-D145N-R743A* gene replacement construct, pTA1947, was introduced into H2117 (Δ*pyrE2, bgaHa-Bb, leuB-Ag1,* Δ*trpA,* Δ*hel308::trpA+*) to generate the pop-in strain H3833. Strains were grown over three consecutive overnights in YPC to allow the pop-out to occur. Strains were plated on 5-FOA in order to select for the pop-out of pTA1947. Colonies were screened via colony hybridisation using a radiolabelled probe. H3927 (Δ*pyrE2, bgaHa-Bb, leuB-Ag1,* Δ*trpA, hel308-D145N-R743A*) was verified via Southern blot (Figure 3.32).

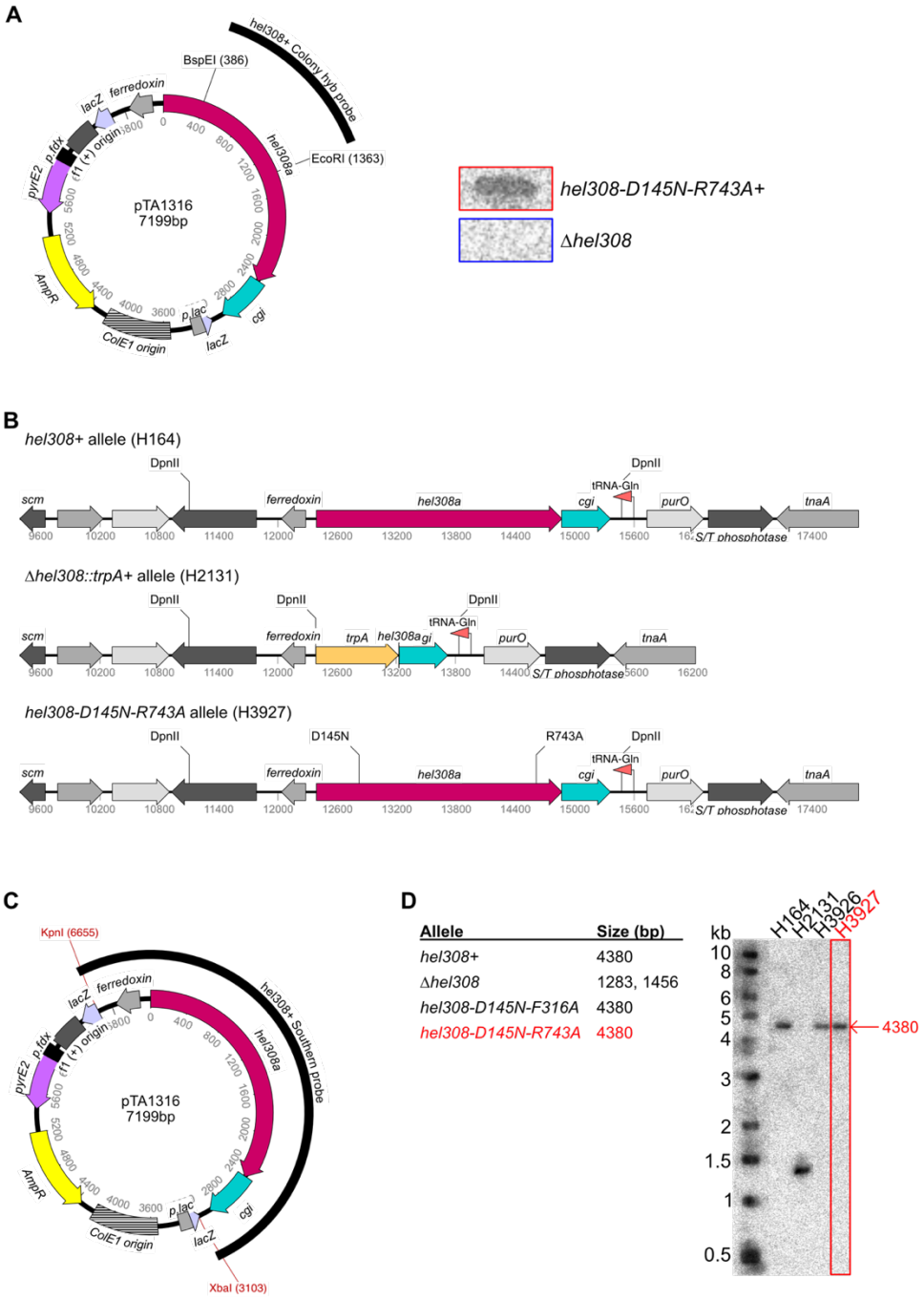


Figure 3.32 **Gene replacement of Δ*hel308::trpA*+ with *hel308-D145N-R743A* in H2131 (Δ*pyrE2, Nph-pitA,* Δ*mrr, cdc48d-ct,* Δ*trpA,* Δ*hel308::trpA*+).** A) The 977bp *hel308*+ probe for colony hybridisation was made by digesting pTA1316 with *Bsp*EI and *Eco*RI, and the corresponding hyb. B) Allele for *hel308+,* Δ*hel308::trpA*+ and *hel308-D145N-R743A*. The *Dpn*II sites used to digest genomic DNA are shown. C) A 3647bp *hel308*+ probe for southern hybridisation was made by digesting pTA1316 with *Kpn*I and *Xba*I. D) Predicted fragment size and Southern blot, highlighted and boxed in red, showing H3927 is *hel308-D145N-R743A*+.

pTA1952 (*hel308-D145N-F316A* gene replacement construct)

Due to the proximity of the D145N and F316A point mutations in *hel308*, there were no unique restriction sites in between to carry out a simple fragment swap. Instead, overlap extension PCR was used to introduce the F316A point mutation using the *hel308-D145N* gene replacement construct, pTA1335, as a DNA template. The primers used were Hel308FInt, Hel308F316AR, Hel308F316AF and Hel308EcoR. The resultant *hel308* fragment, containing both the D145N and F316A point mutations, was digested with the restriction enzymes *Eco*RI and *Bsp*EI, and ligated into the *Eco*RI and *Bsp*EI sites of pTA1335 to generate *hel308-D145N-F316A* gene replacement construct, pTA1952. Construct confirmed by DNA sequencing (Figure 3.33).

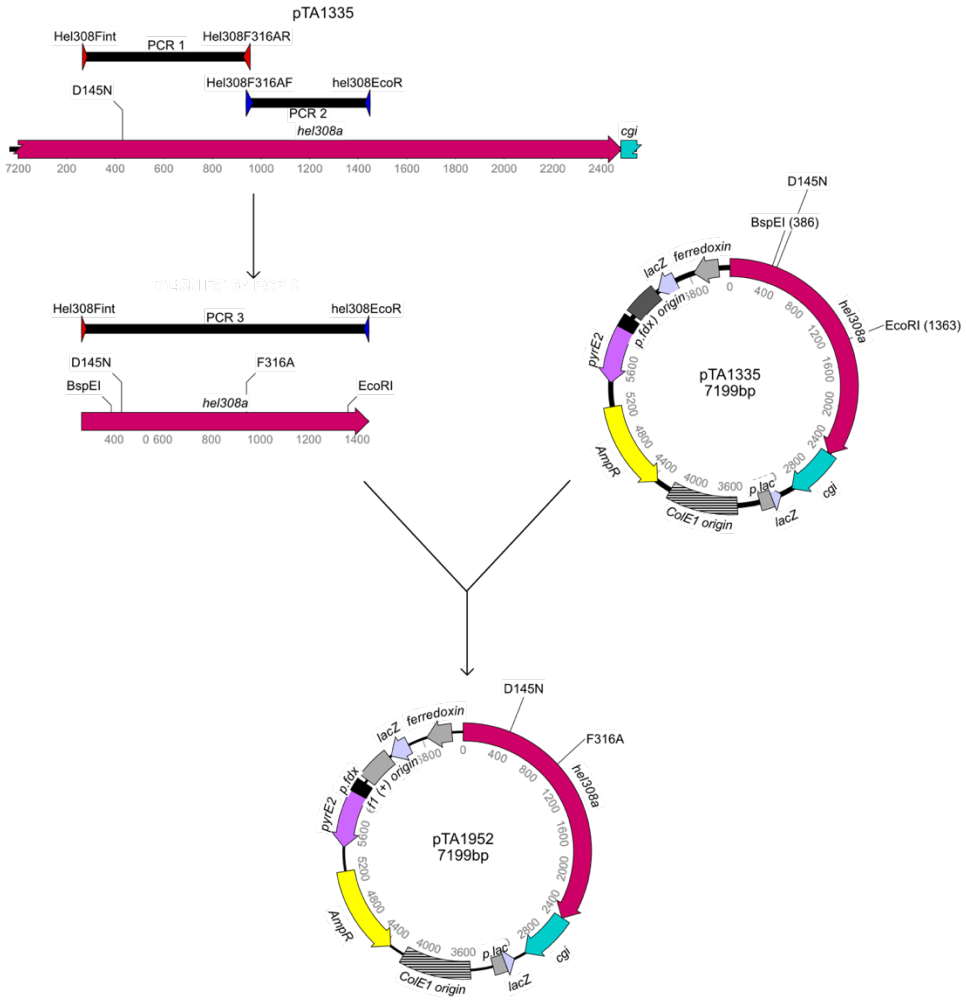


Figure 3.33: **Construction of *hel308-D145N-F316A* gene replacement construct, pTA1952.** pTA1335 was used as template DNA to PCR amplify from Hel308FInt to Hel308F316AR to yield the PCR 1 product, and from Hel308F316AF to Hel308EcoR to yield the PCR 2 product. PCR 1 and PCR 2 products were mixed in approximate equal proportion and underwent 10 cycles of PCR without primers. External primers, Hel308FInt and Hel308EcoR, were added for the last 20 cycles of PCR to generate a *hel308* fragment containing the both the D145N and F316A point mutations. The fragment was digested with the restriction enzymes *Eco*I and *Bsp*EI, and ligated into the *Eco*I and *Bsp*EI sites of pTA1335 to generate *hel308-D145N-F316A* gene replacement construct, pTA1952.

The *hel308-D145N-F316A* gene replacement construct, pTA1952, was introduced into H2117 (Δ*pyrE2, bgaHa-Bb, leuB-Ag1,* Δ*trpA,* Δ*hel308::trpA+*) to generate the pop-in strain H3835. Strains were grown over three consecutive overnights in YPC to allow the pop-out to occur. Strains were plated on 5-FOA in order to select for the pop-out of pTA1952. Colonies were screened via colony hybridisation using a radiolabelled probe. H3926 (Δ*pyrE2, bgaHa-Bb, leuB-Ag1,* Δ*trpA, hel308-D145N-F316A*) was verified via Southern blot (Figure 3.34).

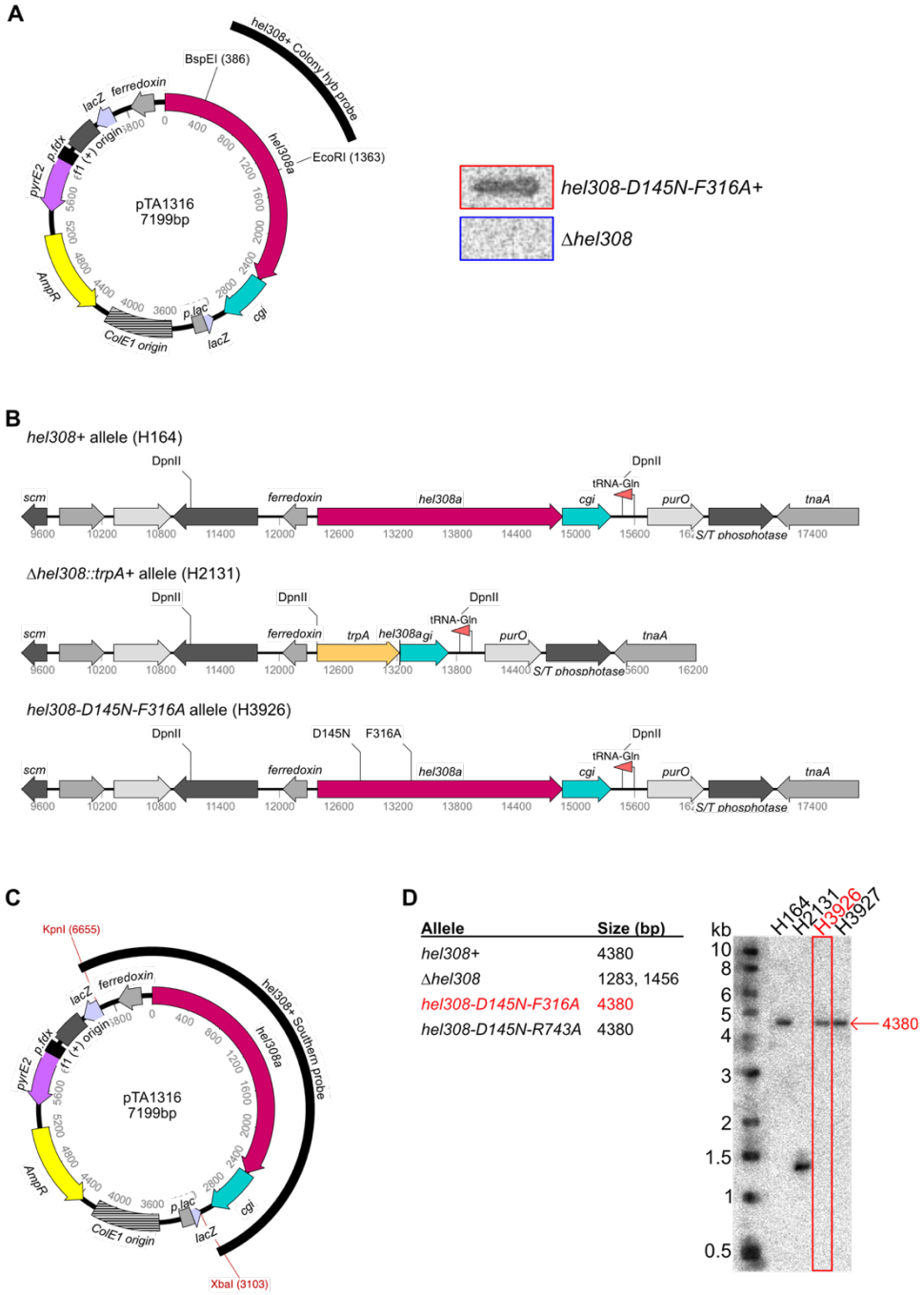


Figure 3.34 **Gene replacement of Δ*hel308::trpA*+ with *hel308-D145N-F316A* in H2131 (Δ*pyrE2, Nph-pitA*, Δ*mrr, cdc48d-ct*, Δ*trpA*, Δ*hel308::trpA*+).** A) The 977bp *hel308*+ probe for colony hybridisation was made by digesting pTA1316 with *Bsp*EI and *Eco*RI, and the corresponding hyb. B) Allele for *hel308*+, Δ*hel308::trpA*+ and *hel308-D145N-F316A*. The *Dpn*II sites used to digest genomic DNA are shown. C) A 3647bp *hel308*+ probe for southern hybridisation was made by digesting pTA1316 with *Kpn*I and *Xba*I. D) Predicted fragment size and Southern blot, highlighted and boxed in red, showing H3926 is *hel308-D145N-F316A*+.

3.4.2.2 Characterisation of ATPase-null and hyper-recombinant mutant *hel308* strains

Growth Rate

Standard growth rates in liquid media were determined for the ATPase-null and hyper-recombinant mutant *hel308* strains. Strains were grown over two consecutive overnights to form a vigorously growing culture. The generation time in exponential phase of cell growth was calculated using an Epoch2 Microplate Spectrophotometer (BioTek) by measuring the A_{600} of the culture at 15 minute intervals over a 48-hour time course. A growth curve was plotted and the generation time during exponential phase for each strain was calculated (Figure 3.35).

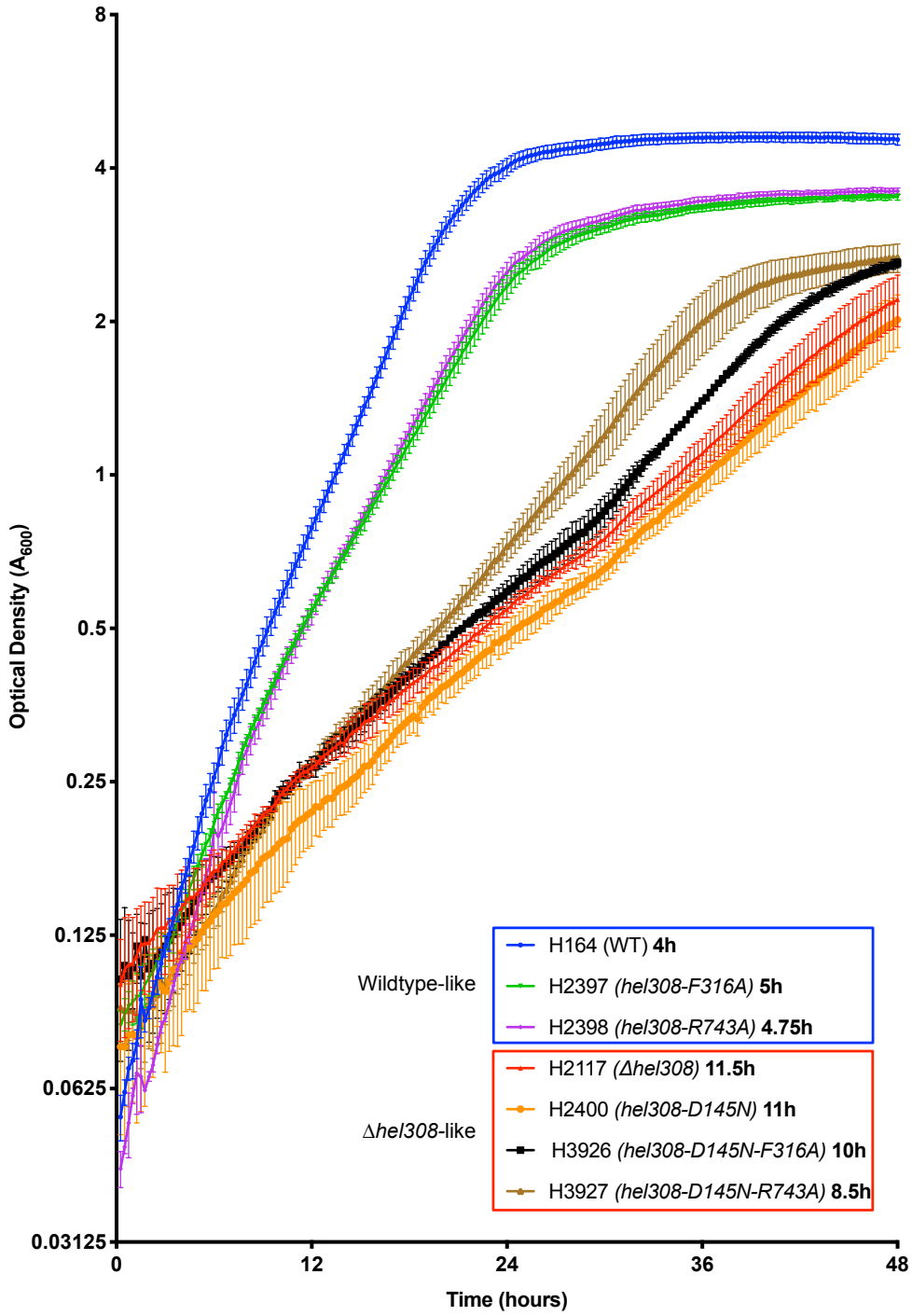


Figure 3.35: **Exponential growth rate of ATPase-null and hyper-recombinant mutant *hel308* strains.** H164 (*hel308*+), H2397 (*hel308-F316A*), H2398 (*hel308-R743A*), H2117 (Δ*hel308*), H2400 (*hel308-D145N*), H3926 (*hel308-D145N-F316A*), and H3927 (*hel308-D145N-R743A*) were grown in liquid culture for two consecutive overnights. The strains were then diluted and grown for 48 hours. Growth was measured by A_{600}. Generation time is indicated next to each strain. Boxes indicate strain groups, wildtype-like (blue) and Δ*hel308*-like (red). All strains were incubated on the same 96-well plate and measured simultaneously using an Epoch2 Microplate Spectrophotometer (BioTek). Data points are a mean of two technical repeats within two biological repeats. Standard deviation is shown.

On this occasion, the wildtype strain H164 had a slower than expected generation time of 4h, demonstrating the importance of making comparisons between sets of strains within the same experiment only. The generation times can be separated into two groups, wildtype-like and Δ*hel308*-like. The Hel308-F316A and Hel308-R743A strains group with the wildtype strain, with slightly slower generation times of 5h and 4.75h, respectively. However, all strains containing the D145N point mutation group with the Δ*hel308* strain. On this occasion, H2117 (Δ*hel308*) had a generation time of 11h. H2400 (*hel308-D145N*), H3926 (*hel308-D145N-F316A*), and H3927 (*hel308-D145N-R743A*) have marginally faster generation times of 11h, 10h, and 8.5h respectively.

Survival Following Treatment with DNA-damaging Agents

The ATPase-null Hel308 strain, H2400, exhibits a Δ*hel308*-like phenotype following UV irradiation or treatment with MMC. By contrast the hyper-recombinant Hel308 strains, H2397 (*hel308-F316A*) and H2398 (*hel308-R743A*), showed a wildtype-like phenotype following UV irradiation or treatment with MMC. Therefore, it was of interest to determine the DNA damage phenotype for the combined ATPase-null and hyper-recombinant mutant strains (Figure 3.36).

Similar to the growth rate analysis, the strains can be categorised into the same two groups, wildtype-like and Δ*hel308*-like. Once more, the Hel308-F316A and Hel308-R743A strains group with the wildtype strain with no observed difference in survival fraction for H2397 (*hel308-F316A*) and H2398 (*hel308-R743A*) compared to H164 (WT) following UV irradiation or MMC treatment. Again similar to the growth rate analysis, all strains containing the D145N point mutation group with the Δ*hel308* strain. These data suggest that the ATPase-null mutation is dominant to the hyper-recombinant mutations.

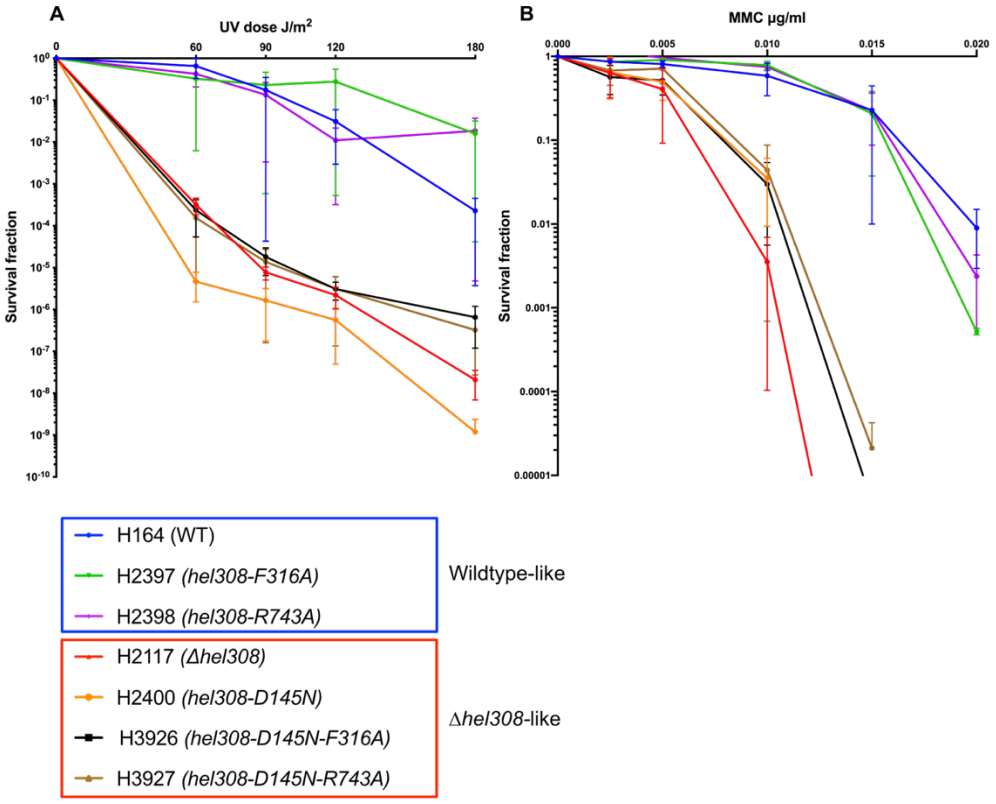
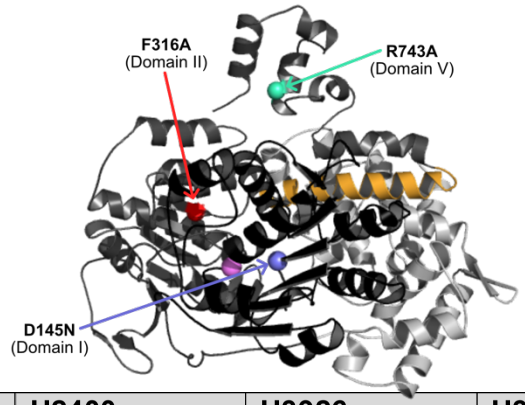


Figure 3.36: **Survival frequency of ATPase-null and hyper-recombinant mutant *hel308* strains following treatment with DNA damaging agents.** No observed difference in survival fraction for H2397 (*hel308-F316A*) and H2398 (*hel308-R743A*) compared to wildtype H164 (*hel308+*) following UV irradiation or MMC treatment. H2117 (Δ*hel308*), H2400 (*hel308-D145N*), H3926 (*hel308-D145N-F316A*), and H3927 (*hel308-D145N-R743A*) have a significant decrease in survival fraction compared to wildtype H164 after UV irradiation and treatment with MMC. Survival fraction is calculated relative to un-treated control. Each data point is generated as the mean of at least 3 independent trials. Boxes indicate strain groups, wildtype-like (blue) and Δ*hel308*-like (red). Standard deviation is shown.

Recombination Frequency

Hyper-recombinant strains are so named because the levels of recombination in these strains are increased a hundred-fold compared to wild-type (Gamble-Milner, 2016). However, these point mutants are not deficient in DNA repair or normal growth. By contrast, the ATPase-null mutant strain, H2400 (*hel308-D145N*), is slow-growing and sensitive to DNA cross-linking agents; the recombination phenotype had not previously been determined. These data suggest a separation of function of Hel308 between its ability to bind D-loop intermediates of recombination and its helicase activity involved in DNA repair. Therefore, it was of interest to analyse the recombination frequency of the ATPase-null mutation strain and the combined ATPase-null and hyper-recombinant mutation strains (Table 3.5 and Figure 3.37).



Strain	**H164**	**H2117**	**H2397**	**H2398**	**H2400**	**H3926**	**H3927**
	hel308+	*Δhel308*	*hel308-F316A*	*hel308-R743A*	*hel308-D145N*	*hel308-D145N-F316A*	*hel308-D145N-R743A*
Recombination Frequency (RF)	$1.92x10^{-5}$ ($+/- 6.71x10^{-6}$)	$4.13x10^{-4}$ ($+/- 2.73x10^{-4}$)	$1.81x10^{1}$ ($+/- 4.59x10^{0}$)	$1.31x10^{1}$ ($+/- 3.77x10^{0}$)	$3.07x10^{-5}$ ($+/- 2.05x10^{-5}$)	$9.35x10^{-4}$ ($+/- 3.36x10^{-4}$)	$3.45x10^{-4}$ ($+/- 1.31x10^{-4}$)
Transformation Efficiency (TE)	$1.34x10^{-3}$	$4.85x10^{-3}$	$7.63x10^{-3}$	$2.58x10^{-2}$	$4.23x10^{-3}$	$1.28x10^{-1}$	$4.63x10^{-2}$
Relative RF normalised by TE	$1.43x10^{-2}$	$7.85x10^{-2}$	$2.37x10^{3}$	$5.07x10^{2}$	$7.26x10^{-3}$	$7.32x10^{-3}$	$7.45x10^{-3}$
	1	**5.5x**	**166000x**	**35500x**	**0.51x**	**0.51x**	**0.52x**
CO fraction	16.5% (200)	21.9% (64)	<0.5% (200)	<0.5% (200)	35.7% (14)	36.8% (19)	40.7% (199)
NCO fraction	83.5% (200)	78.1% (64)	>99.5% (200)	>99.5% (200)	64.3% (14)	63.2% (19)	59.3% (199)

Table 3.5: **Recombination frequencies of HVO *hel308-D145N, hel308-F316A* and *hel308-R743A*.** Values in bold indicate the amount of recombination relative to wildtype H164 (*hel308+*). Values are generated as an average of 5 independent trials, +/- standard error is shown in brackets. Blue shaded cells indicate a recombination defect, whilst red indicates hyper-recombination. Proportion of CO and NCO events are represented as a percentage, with the number of assayed colonies indicated in brackets underneath. Significant differences from the wildtype (P=0.05) are shaded in blue for a decrease and red for an increase (with one degree of freedom with a chi-squared test).

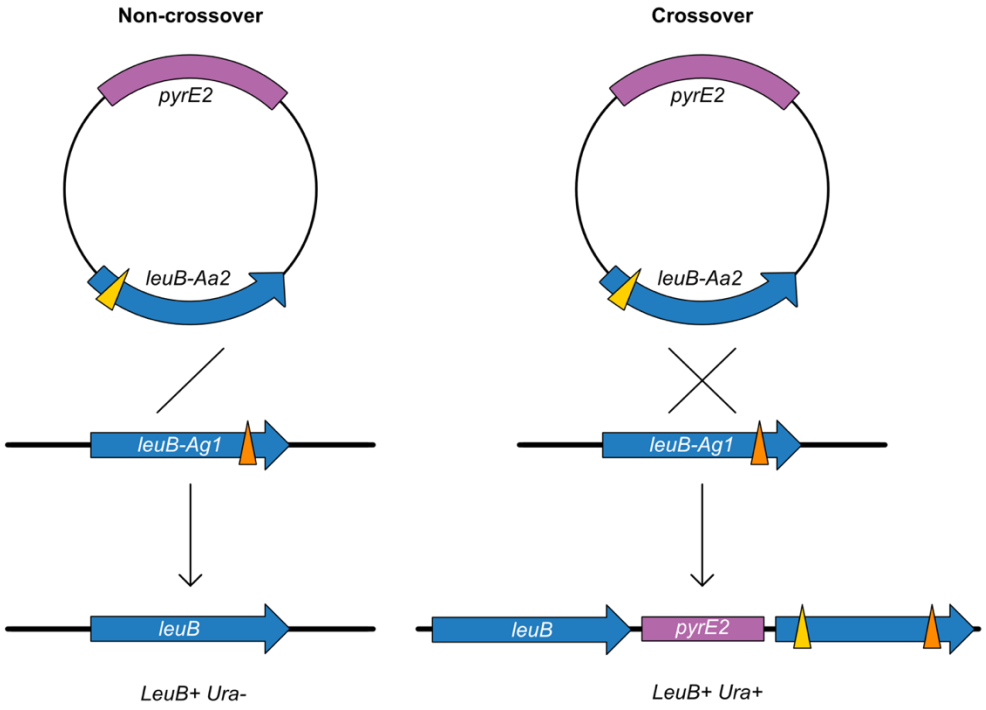


Figure 3.37: **Recombination frequencies of HVO *hel308-D145N, hel308-F316A* and *hel308-R743A.*** Δ*pyrE2* strains with a chromosomal *leuB-Ag1* allele (*leu-*) are transformed with pTA163, containing *pyrE2* and *leuB-Aa2* allele (*leu-*). A recombination event between the plasmid and chromosomal mutant *leuB* alleles generates a wildtype *leu+* allele and therefore strains are able to grow on media lacking leucine. The proportion of crossover (CO) and non-crossover (NCO) events are determined by the retention or loss of the *pyrE2* marker found on pTA163: CO are *pyrE2+ (ura+),* and NCO are Δ*pyrE2 (ura-)*.

H2117 (Δ*hel308*) has a hyper-recombinant phenotype with recombination levels 5.5x that of wild type H164 (*hel308+*). Crossover recombination events were marginally higher that wild type. These data support the proposal that Hel308 is an anti-recombinase. Both H2397 (*hel308-F316A*) and H2398 (*hel308-R743A*) have an extreme hyper-recombinant phenotype with recombination levels of 166,000x and 35,500x, respectively. This result is due exclusively to increased non-crossover recombination, since no crossover events were observed for these strains. A decrease in the levels of recombination was observed for the ATPase-null strain, H2400 (*hel308-D145N*), at 0.51x that of wild type. The combined ATPase-null and hyper-recombinant mutations in the strains H3926 (*hel308-D145N-F316A*) and H3927 (*hel308-D145N-R743A*) resulted in a recombination phenotype identical to the ATPase-null strain H2400 (*hel308-D145N*). Therefore, the ATPase-null (D145N) mutation is dominant to the hyper-recombinant (F316A and R743A) mutations.

DNA Content and Cell Size

Flow cytometry was used to analyse the ATPase-null and hyper-recombinant strains, in order to determine the effect that these point mutations have on DNA content and cell size (Figure 3.38).

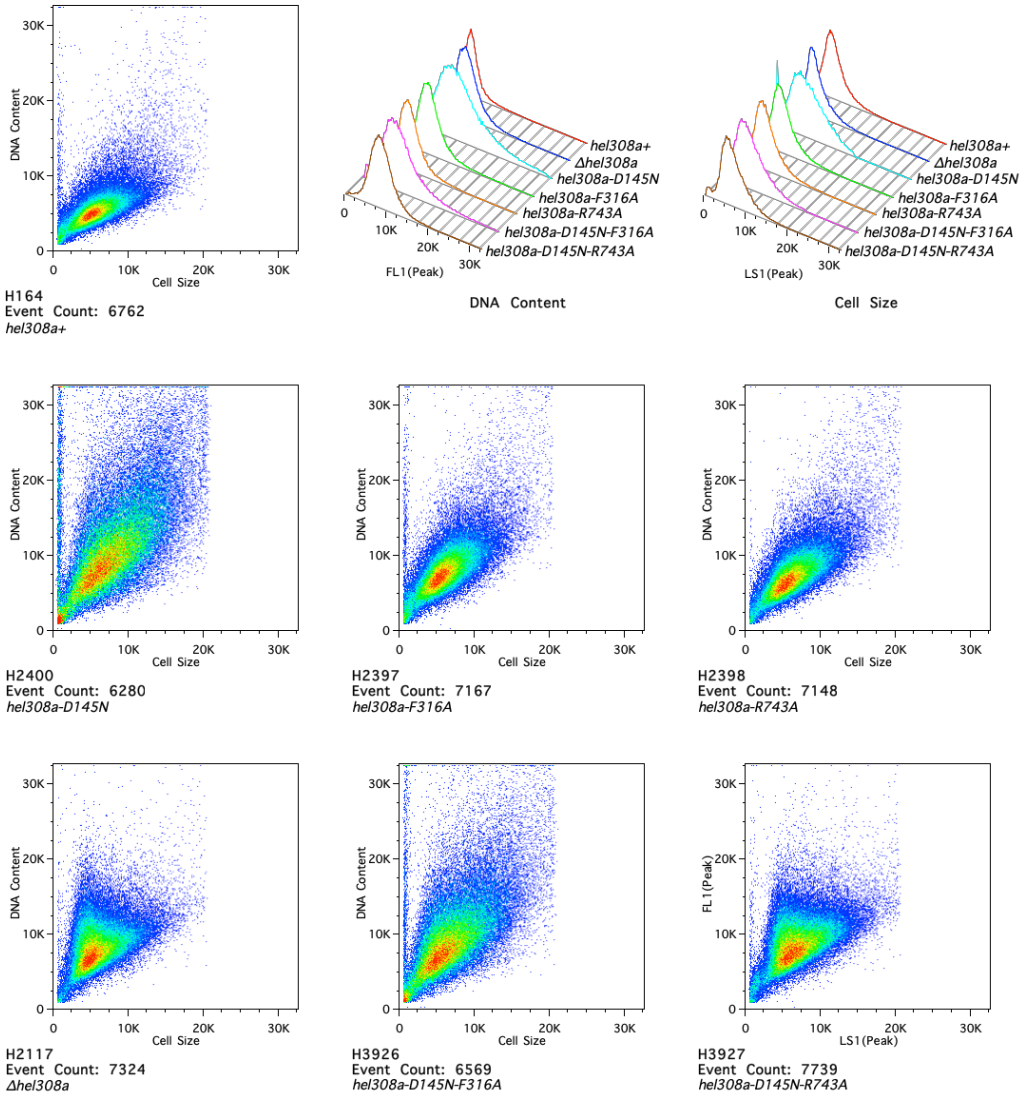


Figure 3.38: **Flow cytometry analysis of ATPase-null and hyper-recombinant mutant *hel308* strains.** Cell size vs DNA content (Arbitrary units) (Alvarez *et al.*, 2010). Determination of cell size, H2400 (*hel308-D145N*) has a broad range of cell sizes. H2397 (*hel308-F316A*), H2398 (*hel308-R743A*), H2117 (Δ*hel308*), H3926 (*hel308-D145N-F316A*), and H3927 (*hel308-D145N-R743A*) have a similar range of cell sizes compared to wildtype H164 (*hel308+*). Determination of DNA content, H2117 (Δ*hel308*) as a higher DNA content compared to wildtype H164 (*hel308+*). H2397 (*hel308-F316A*), H2398 (*hel308-R743A*), H3926 (*hel308-D145N-F316A*), and H3927 (*hel308-D145N-R743A*) have a similar DNA content compared to H2117 (Δ*hel308*). H2400 (*hel308-D145N*) has the highest DNA content.

There was little difference observed between the range of cell sizes of H164 (WT) and H2117 (Δ*hel308*). However, there was a higher DNA content in Δ*hel308* cells compared to wildtype. The hyper-recombinant strains, H2397 (*hel308-F316A*) and H2398 (*hel308-R743A*), have comparable cell size ranges to wild-type and a DNA content similar to Δ*hel308* cells. H2400 (*hel308-D145N*) has the broadest range of cell sizes and highest DNA content. The combined ATPase-null and hyper-recombinant strains, H3926 (*hel308-D145N-F316A*) and H3927 (*hel308-D145N-R743A*), also have comparable cell size ranges to wild-type. The DNA content of H3926 (*hel308-D145N-F316A*) and H3927 (*hel308-D145N-R743A*) is higher than H2397 (*hel308-F316A*) and H2398 (*hel308-R743A*) but lower that H2400 (*hel308-D145N*).

3.5 Discussion

A summary of the DNA damage, growth and recombination phenotypes of all Hel308 point mutations as well as Δ*radA* and Δ*hel308* strains is shown in Table 3.6 below.

Mutation	UV repair	MMC repair	Generation time	Recombination frequency	CO	NCO
Δ*hel308*	↓	↓	↓	↑	WT	WT
Δ*radA*	↓	↓	↓↓↓	↓↓↓ = 0*	-	-
Δ*hel308*, Δ*radA*	WT	WT	WT	↓↓↓ = 0*	-	-
hel308-D145N	↓ (=Δ*hel308*)	↓ (=Δ*hel308*)	↓ (=Δ*hel308*)	↓	↑	↓
hel308-F316A	WT	WT	WT	↑↑↑↑↑	↓↓↓ (<0.5%)	↑↑↑ (>99.5%)
hel308-R743A	WT	WT	WT	↑↑↑	↓↓↓ (<0.5%)	↑↑↑ (>99.5%)
hel308-D145N-F316A	↓ (=Δ*hel308*)	↓ (=Δ*hel308*)	↓ (=Δ*hel308*)	↓ (=D145N)	↑ (=D145N)	↓ (=D145N)
hel308-D145N-R743A	↓ (=Δ*hel308*)	↓ (=Δ*hel308*)	↓ (=Δ*hel308*)	↓ (=D145N)	↑ (=D145N)	↓ (=D145N)

Table 3.6 **Summary of phenotypes of Δ*hel308*, Δ*radA*, and Hel308 point mutants.** Cells are shaded blue to indicate a defect in survival following DNA damage, generation time, or recombination. Defects equivalent to the Δ*hel308* or *hel308-D145N* strains are indicated. Cells shaded red indicate hyper-recombination.

Separation of functions in recombination and DNA repair

Deleting the *hel308* gene increases recombination but decreases DNA repair and growth. This suggests that Hel308 actively targets and unwinds D-loops, supporting the proposal that Hel308 functions as an anti-recombinase. Mutating residues F316 and R743 increases recombination without affecting DNA repair or growth; by contrast, replacing *hel308* with the ATPase-null *hel308-D145N*, decreases recombination levels to half that of wild type. The combined ATPase-null and hyper-recombinant mutations, *hel308-D145N-F316A* and *hel308-D145N-R743A*, also decrease recombination levels to half that of wild type. Therefore, the ATPase-null (D145N) mutation is dominant to the hyper-recombinant (F316A and R743A) mutations. These data suggest that domains 2 and 5 play a significant role in the regulation of homologous recombination. The decrease in recombination when the D145N mutation is present implies that helicase activity is important for homologous recombination; conversely, the increase in recombination in the Δ*hel308* strain contradicts this. An alternative explanation arises when comparing the Walker A and Walker B mutations. In this study, the Walker A mutations (K53A/R) were not viable and possibly lethal to the cell, whereas the Walker B mutation (D145N) was viable and showed a Δ*hel308*-like phenotype in terms of DNA repair and growth. The Walker A motif is responsible for binding ATP and the Walker B is responsible for ATP hydrolysis. So why is the Walker B mutation tolerated but not the Walker A mutation? This may be answered by considering whether Hel308 first binds ATP or DNA. If ATP is bound prior to DNA binding, then the Walker A mutation should give rise to a Δ*hel308*-like phenotype in terms of DNA repair and growth, which it does not. In fact, the K53A/R mutations are more detrimental to the cell than the deletion of the entire gene. This proposes a model of Hel308 binding in *H. volcanii* whereby Hel308 first binds DNA tightly. On binding ATP, Hel308 changes conformation and loosens its grip on the DNA but remains static. Hydrolysis of the ATP causes a second conformational change and a release of energy resulting in the paddling motion and translocation along the DNA (Figure 3.39).

.

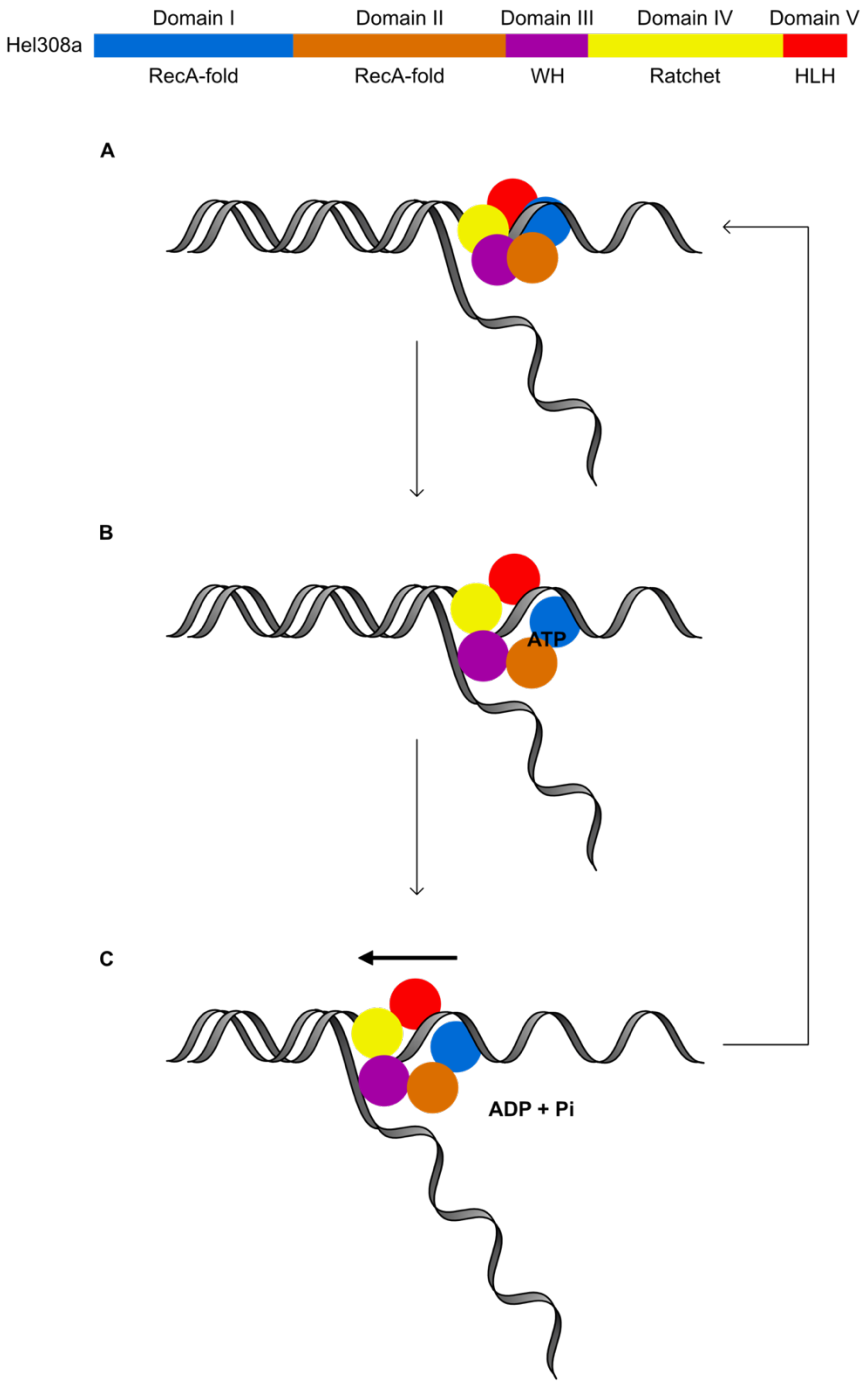


Figure 3.39: **Proposed DNA binding and translocation mechanism of Hel308.** A) Hel308 first binds DNA tightly. B) On binding ATP, Hel308 changes conformation and loosens its grip on the DNA but remains static. C) Hydrolysis of the ATP causes a second conformational change and a release of energy resulting in a paddling motion and translocation along the DNA in the direction of the arrow. In the absence of ATP, Hel308 binds DNA tightly once more completing a conformational change cycle.

Unable to bind ATP or translocate, the Hel308-K53A/R effectively creates a "road-block" by binding tightly to DNA and stalling replication and repair. This results in catastrophic genome arrest, hence the cells are not viable. Conversely, the Hel308-D145N protein is able to successfully bind DNA followed by ATP

but remains static on the ssDNA, blocking access for DNA replication and repair proteins. Due to the relatively loose grip that Hel308-D145N has on the DNA, the helicase can be dissociated granting access to DNA replication, recombination, and repair proteins. Therefore, the Walker B mutation acts as a 'speed-bump' slowing down recombination by blocking access to the ssDNA for RadA nucleofilament formation. Whilst this proposed model provides an explanation for the difference between the *H. volcanii* Walker A and Walker B mutants, existing ATP-dependent helicase crystal structures, with and without the ADP and AMPPCP (ATP mimic) ligands, show no overall change in helicase structure. Therefore, if the proposed conformational changes in this model of Hel308a do exist, the changes are likely to be subtle at a base level as opposed to large scale domain changes. The hyper-recombinant and ATPase mutations have identified a separation of functions between helicase-independent recombination and helicase-dependent DNA repair.

Hel308 is a DNA repair regulator

In *Haloferax volcanii*, mutants deleted for the RadA paralogue *radB* exhibit a decrease in the level of homologous recombination to 5% of that in wild type cells. However, a double deletion of *hel308* with *radB* restores, in part, homologous recombination. Therefore it was proposed that Hel308 functions as an anti-recombinase (Gamble-Milner, 2016). To further explore the role of Hel308 in homologous recombination, *hel308* was successfully deleted in combination with the homologous recombination gene, *radA*. Strains deleted for *radA* usually have a severe growth defect (Woods and Dyall-Smith, 1997). As expected, the Δ*radA* strain (H298) had a severe fitness defect with a generation time of 7h compared to the wildtype generation time of 2.75h in this study. Strains deleted for *hel308* also exhibit significant fitness defects but not to the same severity as the Δ*radA* strains; the Δ*hel308* strain had a generation time of 5.25h. Most interestingly, the strain deleted for both *hel308* and *radA* (H3515) was healthier than either single deletion, with a generation time of 3.75h, only ~30% slower than the wildtype. A similar phenotype was observed in survival fraction following UV irradiation and MMC treatment. Strains deleted for *radA* are sensitive to both UV irradiation and MMC treatment, as was seen in this study. Similarly, strains deleted for *hel308* are sensitive to MMC treatment and sometimes UV irradiation depending on the genotypic background. Again, the strain deleted for both *hel308* and *radA* (H3515) was healthier than either single deletion, exhibiting a wildtype-like response to both UV irradiation and MMC treatment (Table 3.6). These data suggest that in the absence of Hel308-mediated repair and homologous recombination, a third distinct DNA-repair pathway is activated. The fitness defects observed in the single *radA* and *hel308* delete strains suggest that this third DNA repair pathway is supressed by both Hel308-mediated repair and homologous recombination (Figure 3.40).

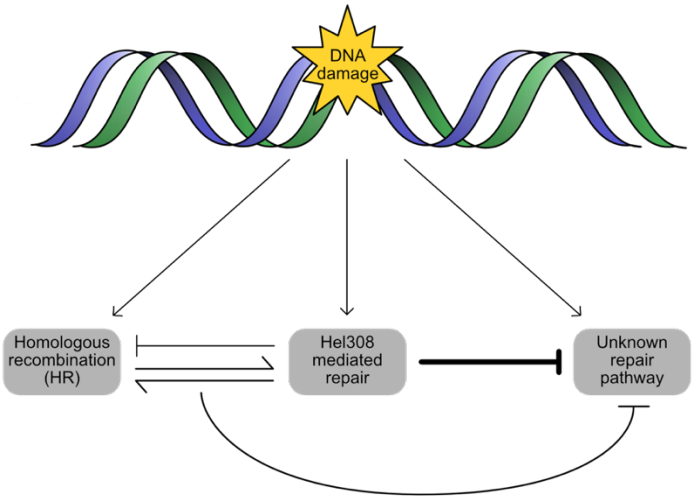


Figure 3.40: **Hel308 is a regulator of DNA repair pathways.** In the absence of Hel308-mediated repair and homologous recombination, a third distinct DNA-repair pathway is activated. Hel308 strongly suppresses this third DNA repair pathway and partially suppresses the homologous recombination pathway. The unknown repair pathway is also suppressed by homologous recombination.

An alternative explanation is that in deleting both the *hel308* and *radA* genes, a suppressor mutation has arisen. In *E. coli,* the RecBCD complex processes DNA ends in pre-synapsis in order to generate ssDNA onto which RecA is loaded. The RecA nucleoprotein filament then binds dsDNA molecules in order to search for a region of homology. Once found, RecA catalyses strand invasion and exchange. Like archaeal Δ*radA* mutants, Δ*recA* mutants are totally deficient in recombination. The single and double Δ*recB* and Δ*recC* mutants are partially deficient in recombination, similar to the archaeal Δ*radB* phenotype. Two classes of suppressor mutations that suppress the Δ*recBC* phenotype have been identified in *E. coli* (Clark and Margulies, 1965). The two classes are activation of dormant genes in order to fulfil the missing role, and deactivation of processing enzymes, which degrade recombination intermediates. The *sbcA* mutation activates the quiescent *recE* and *recT* genes. *recE* encodes Exonuclease VIII, a double-stranded exonuclease with the capability of processing DNA ends in the absence of *recB* and *recC*; RecT promotes DNA annealing and strand exchange. The deactivation of processing enzymes requires mutations in both *sbcB* and *sbcC/D.* The *sbcB* mutation inactivates the 3' ssDNA exonuclease, which would otherwise degrade the 3' tail preventing RecA nucleoprotein filament formation. The *sbcC/D* mutations inactivate the SbcCD nuclease, which degrades long DNA palindromes and processes convergent replication forks (Wendel *et al.*, 2017). The deactivation of these processing enzymes allows time for alternative pre-synaptic processing by RecQ and RecJ. It is possible that in deleting both *hel308* and *radA*, suppressor mutations such as these have arisen in the H3515 strain in order to restore wildtype-like fitness to the cell. Only a singular clone with the Δ*hel308*Δ*radA* genotype was generated, which supports

the suppressor-mutation theory. However, a suppressor mutation or mutations would need the ability to bypass two distinct repair pathways in order to restore wildtype-like fitness. This seems unlikely considering that the *E. coli* recBC mutants, which contain the suppressor mutations, are never as healthy or as proficient in homologous recombination as the wild-type.

3.6 Future Perspectives

Is the Walker A mutant synthetically lethal?

In this study, the Walker A mutations (K53A/R) were not viable and possibly lethal to the cell, whereas the Walker B mutation (D145N) was viable and had a Δ*hel308*-like phenotype in terms of DNA repair and growth. If ATP is bound prior to DNA binding, then the Walker A mutation should also give rise to a Δ*hel308*-like phenotype in terms of DNA repair and growth. The K53A/R mutations are more detrimental to the cell than the deletion of the entire gene suggesting that Hel308 first binds DNA tightly. On binding ATP, Hel308 changes conformation and loosens its grip on the DNA but remains static. Hydrolyses of the ATP causes a second conformational change and a release of energy resulting in the paddling motion and translocation along the DNA. Unable to bind ATP nor translocate, the Hel308-K53A/R affectively creates a 'road-block' by binding tightly to DNA and stalling replication and repair. This results in catastrophic genome arrest, hence the cells are not viable. Whilst this study attempted to explore this, the correct strain for this episomal expression analysis was not generated. Therefore, it would be of interest to repeat the episomal expression assay with the correct strain in order to determine whether or not the Walker A mutant is synthetically lethal.

Hel308 Point mutations

Further point mutations could be made to aid in the understanding of the function of Hel308. Due to the hyper-recombinant phenotype of the central arginine residue of the RXRAR motif in domain 5 (R743A), it would be of interest to mutate the other key arginine residues, R741 and R745. Genetic analysis of the flanking arginine point mutations would complement the biochemical analyses completed in *M. thermautotrophicus* and *Sulfolobus solfataricus.* Comparison of the genetic and biochemical data should give insight into the ATPase and helicase functions of Hel308. Similarly, due to the hyper-recombinant phenotype of the F316A mutation in domain 2, it would be of interest to generate further mutations in this region in order to locate the specific residues involved in the regulation of homologous recombination.

The function of domain 4, also known as the winged helix (WH) domain, remains unknown in *H. volcanii.* The Hel308 winged helix domain is non-canonical. In *M. thermautotrophicus,* mutagenesis of conserved aromatic residues within the WH domain have a profound effect on helicase unwinding however, mutations in the recognition helix have little effect (Woodman and Bolt, 2011b). It has been proposed that the WH domain may contribute to ssDNA translocation by binding dsDNA, distinct to the ssDNA binding by the core helicase, in order to directly promote activity of Hel308 or the removal of other DNA bound proteins by 'reeling' ssDNA. The F434V and F439A mutations abolish this function in *M. thermautotrophicus* (Northall *et al.*, 2017). It would be of interest to mutate the corresponding residues, F456V, E494G and Q497G, for analysis in *H. volcanii.* The second of the two phenylalanine residues in *M. thermautotrophicus*, F439, is not conserved in *H. volcanii* however, the adjacent threonine and tyrosine residues are. Therefore, it would be of interest to mutate T460, L461 (the equivalent position of Mth F439), and Y462.

Combining hyper-recombinant and ATPase-null point mutants with the deletion of recombination genes

During this study, several attempts to combine the hyper-recombinant and ATPase-null point mutations in Hel308 with either the deletion of *radB* alone or the 5-gene *radB* neighbourhood deletion (Δ*rcrA*Δ*radB*Δ*ndnR*Δ*hrp*Δ*cdc48a*) were made but remained unsuccessful. In *H. volcanii*, a low level of recombination, ~5% that of wildtype, still occurs within *radB* deleted strains. Therefore, introducing the Hel308 hyper-recombinant and ATPase-null point mutations into Δ*radB* and Δ*rcrA*Δ*radB*Δ*ndnR*Δ*hrp*Δ*cdc48a* genetic backgrounds, would indicate whether Hel308 is regulating homologous recombination via direct interaction with RadB or by other means. It would also be of interest to couple the hyper-recombinant and ATPase-null point mutations in Hel308 with the deletion of *radA.*

Why is Δ*hel308*Δ*radA* healthy?

In order to determine whether a suppressor mutation or mutations have arisen during the deletion of *hel308* and *radA*, analysis of the H3515 strain via Next Gen sequencing needs to be completed. Complementation assays, reintroducing *hel308* and *radA* at the respective chromosomal loci, would also aid in the understanding of the complex interplay between these pathways. In order to achieve this, the *radA in trans* expression construct, pTA635 (or the *radA-S101P in trans* expression construct, pTA2080), must be introduced before the normal transformation protocol can be attempted.

In *Bacillus subtilis,* the RecA nucleoprotein filament is unable to catalyse DNA strand exchange *in vitro*. To overcome this, a two-component mediator comprising single-stranded binding protein (SsbA) and DNA processing protein (DprA) (RecO can replace DprA in a Δ*dprA* strain) is required. The two-component mediator promotes RecA nucleoprotein filament formation onto SsbA/B coated ssDNA. The RecA nucleoprotein filament, with the help of the two-component mediator, then binds dsDNA molecules in order to search for a region of homology. Once found, the two-component mediator activates RecA in order to catalyse strand invasion and D-loop formation. *Bacillus subtilis* encodes three branch migration translocases, RuvAB, RecG, and RadA (encoded by the *radA* gene*,* the protein is also known as Sms and is analogous to archaeal RadB). RadA/Sms unwinds forked DNA in the 5′-3′ direction, in the absence of RadA/Sms the RecG translocase is required for natural chromosomal transformation (Torres, Serrano and Alonso, 2019). This interaction between the RadA homologue, RadB analogue, and the RecG helicase in order to regulate recombination suggests that a complex interplay between Hel308, RadA, and RadB may also exist. Following the analysis of the Δ*hel308* Δ*radA* strain via Next Gen sequencing to determine the presence of any suppressor mutations, it would be of interest to reattempt the triple deletion of *hel308, radB,* and *radA* in order to explore any complex interplay between Hel308, RadA, and RadB. Additional *in trans* expression of both *radB* and *hel308* may assist in the generation of the triple delete strain.

3.7 Conclusion

Hel308 acts as an anti-recombinase. Mutation of the F316 and R743 residues in Hel308 has identified a separation of functions between recombination and DNA repair. The ATPase-null mutation is dominant to the hyper-recombinant mutations. It is suggested that Hel308 binds DNA prior to ATP binding.

Chapter 4: Biochemical Analysis of Hel308

4.1 Background

Halophilic proteins

Halophiles must maintain an osmotic balance between intracellular and extracellular salt concentrations in order to survive. There are two alternative mechanisms that an organism can adopt to achieve this, a 'salt out' or a 'salt in' mechanism. 'Salt out' is the choice mechanism adopted by the various halophilic bacteria and eukaryotes present in these conditions. Here, an osmotic balance is maintained via two techniques. Firstly, the pumping out of salts from the cell and secondly, cytosolic packing with organic solutes such as glycerol (Christian and Waltho, 1962; Oren, 1999, 2008). Halophilic archaea, including *Haloferax volcanii*, adopt the 'salt in' mechanism whereby an osmotic balance is maintained by the accumulation of high levels of potassium cations in the cytosol, potassium hydrates less than sodium. (Oren *et al.*, 2002).

The proteins within halophilic archaea have adapted over time to function in this high salt to low water ratio. This protein adaptation has occurred via several different strategies (Ortenberg, Rozenblatt-Rosen and Mevarech, 2000). Firstly, the amino acid composition and distribution within halophilic proteins differ from mesophilic proteins. Halophilic proteins are rich in acidic residues such as aspartic acid and glutamic acid. These acidic residues are located predominantly on the surface of the protein creating an overall low isoelectric point (pI) and a high density of negative charges. These negative charges on the surface of the protein co-ordinate a network of hydrated cations enabling the protein to remain soluble (Lanyi, 1974). Some halophilic proteins may have additional acidic residue rich domains or peptide insertions, which are crucial for correct protein folding (Graziano and Merlino, 2014). The surface hydrophobicity of halophilic proteins is also reduced by replacing large hydrophobic side groups with small hydrophilic side groups.

Protein purification

Halophilic proteins require a high salt concentration in order to maintain structural stability and activity. Therefore, the use of mesophilic hosts such as *E. coli* for expression of halophilic proteins can be problematic. The low ionic conditions in *E. coli* often cause the expressed halophilic proteins to aggregate and therefore require recovery from inclusion bodies followed by denaturation and refolding in hypersaline solutions (Jaenicke, 2000; Ishibashi *et al.*, 2003).

Connaris and colleagues used an *E. coli* system to express and purify *H. volcanii* dihydrolipoamide dehydrogenase, a flavoprotein enzyme involved in energy metabolism. The protein was first purified using copper metal ion affinity chromatography in the presence of 2 M KCl, and then 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 4°C (Connaris *et al.*, 1999). Functionality assays, such as helicase or nuclease assays, can confirm the correct refolding of a protein. However, this is only possible for well-characterised proteins and therefore protein recovery is only reliable in these instances. Furthermore, proteins expressed in an exogenous host may not undergo the necessary post-translational modifications such as acetylation or SAMPylation that would occur in the native host (Altman-Price and Mevarech, 2009; Zhou *et al.*, 2010; Eichler and Maupin-Furlow, 2013).

Purification of halophilic proteins expressed in *H. volcanii*

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

Jolley and colleagues expressed dihydrolipoamide dehydrogenase in the native host *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 both positively-charged calcium ions and negatively charged-phosphate ions, which interact with carboxylate residues and basic residues on the protein surface, respectively. This dual interaction with an individual protein with specific properties, enables a particular elution pattern across a phosphate gradient (Schröder, Jönsson and Poole, 2003).

Humbard and colleagues purified the 20S proteasome from *H. volcanii* utilising tandem affinity tagging; the α1-subunit was His_6-tagged and the β-subunit was StrepII-tagged, both at their respective C-termini. Purification was accomplished first by nickel-affinity chromatography followed by application of the elutions to a StrepTactin column (streptavidin variant) (Humbard, Zhou and Maupin-Furlow, 2009). Due to the compatibility of these tags in high salt, this approach has been adopted as a standard method of expressing and purifying proteins in *H. volcanii.*

The protein expression constructs used in these studies were custom made and then tailored to Hel308. 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

Previous to this study, Allers and colleagues had developed generic episomal overexpression constructs that allow for native conditional over-expression of *H. volcanii* proteins that are N-terminally His_6-tagged (Allers *et al.*, 2010), N-terminally His_6-tagged and C-terminally StrepII-tagged, and N-terminally StrepII-tagged and C-terminally His_6-tagged (Gamble-Milner, 2016).

The expression plasmids pTA963 and pTA1228 feature a *his_6-tag* flanked by restriction sites that allow for in-frame insertion of a gene of interest (Figure 4.1). The original pTA963 contains the restriction sites *Bsp*HI, *Nco*I and *Pci*I to allow for insertion of genes that have the second codon starting with an A, G and T respectively. pTA1228 is the improved vector with an additional *Nsp*I restriction site to allow for insertion of genes that have the second codon starting with a C, *Nsp*I has the additional advantage of being compatible with *Sph*I.

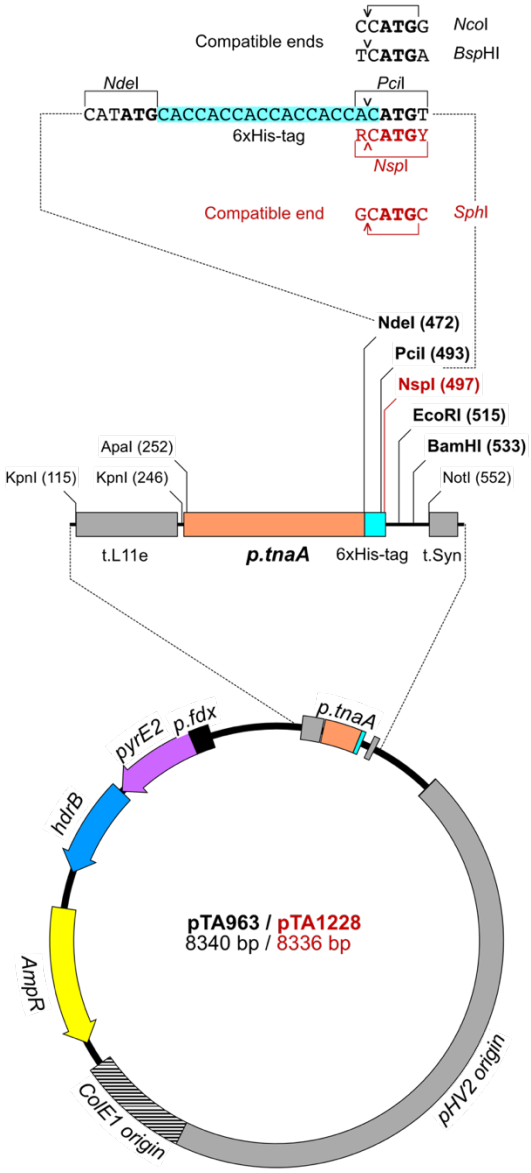


Figure 4.1: **Plasmids for conditional overexpression of His_6-tagged proteins.** There are additional available sites in pTA1228 (red) compared with pTA963. Plasmids contain a *his_6-tag* downstream of the tryptophan inducible promoter *p.tnaA*. The genes *pyrE2* and *hdrB* allow for selection in *H. volcanii*. Restriction sites, and compatible enzymes, available for inserting genes of interest are displayed on the schematic. Abbreviations: *t.L11e* (terminator), *p.tnaA* (tryptophan inducible promoter), *his_6-tag* (hexahistidine tag), *t.Syn* (terminator), *pHV2* origin (*H. volcanii* origin of replication), *ColE1* origin (*E. coli* origin of replication), *AmpR* (ampicillin resistance gene, *E. coli*), *hdrB* (thymidine biosynthesis, *H. volcanii*), *pyrE2* (uracil biosynthesis, *H. volcanii*), *p.fdx* (ferredoxin promoter). Y = T/C and R = A/G. Adapted from Allers et al (2010).

The expression plasmids pTA1392 and pTA1403 both feature a *strepII-tag* in addition to the *his$_6$-tag*. pTA1392 allows for in-frame insertion of a gene of interest generating an N-terminally His$_6$-tagged and C-terminally StrepII-tagged recombinant protein. An *Nhe*I restriction site replaces the stop codon of the protein of interest and a new stop codon is introduced after the *strepII-tag* (Figure 4.2). pTA1403 allows for in-frame insertion of a gene of interest generating an N-terminally StrepII-tagged and C-terminally His$_6$-tagged recombinant protein. Similarly, an *Nhe*I restriction site replaces the stop codon of the protein of interest and a new stop codon is introduced after the *his$_6$-tag* (Figure 4.2).

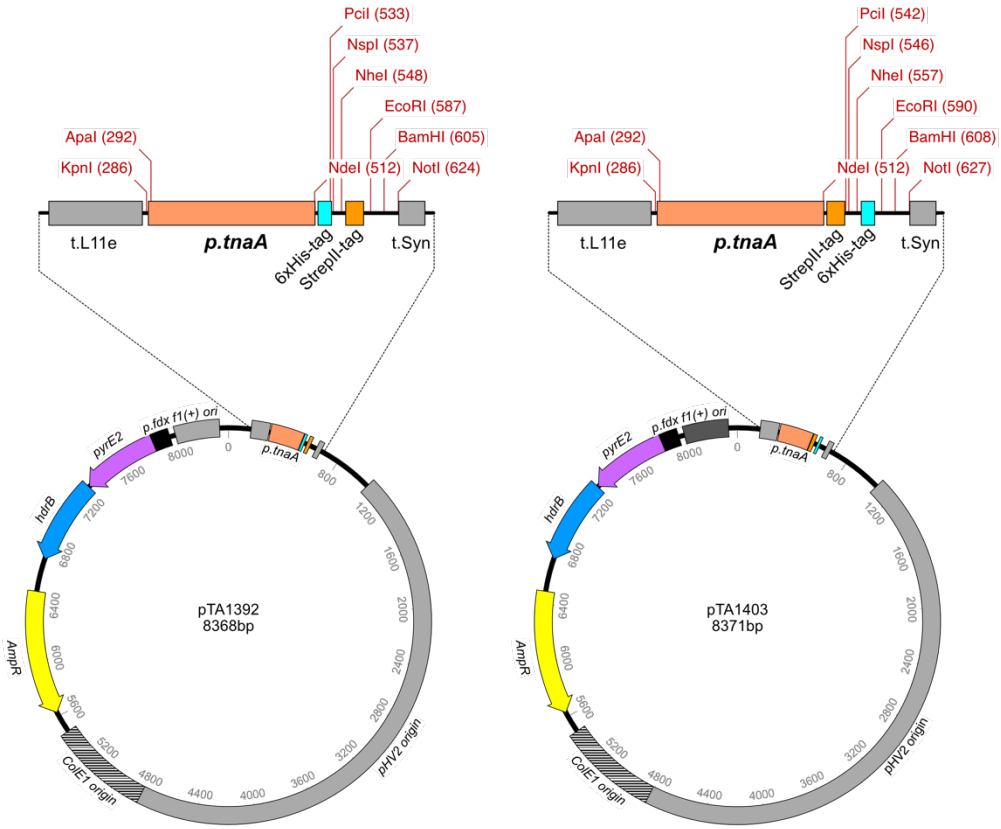


Figure 4.2: **Plasmids for conditional overexpression of His$_6$- and StrepII- tagged proteins.** pTA1392 (left) allows for in-frame insertion of a gene of interest generating an N-terminally His$_6$-tagged and C-terminally StrepII-tagged recombinant protein. pTA1403 (right) allows for in-frame insertion of a gene of interest generating an N-terminally StrepII-tagged and C-terminally His$_6$-tagged recombinant protein. Plasmids contain a *his$_6$-tag* and *strepII-tag* downstream of the tryptophan inducible promoter *p.tnaA*. An *Nhe*I restriction site replaces the stop codon of the protein of interest and a new stop codon is introduced after the C-terminal tag. The genes *pyrE2* and *hdrB* allow for selection in *H. volcanii*. Restriction sites available for inserting genes of interest are displayed on the schematic. Abbreviations: *t.L11e* (terminator), *p.tnaA* (tryptophan inducible promoter), *his$_6$-tag* (hexahistidine tag), *t.Syn* (terminator), *pHV2* origin (*H. volcanii* origin of replication), *ColE1* origin (*E. coli* origin of replication), *AmpR* (ampicillin resistance gene, *E. coli*), *hdrB* (thymidine biosynthesis, *H. volcanii*), *pyrE2* (uracil biosynthesis, *H. volcanii*), *p.fdx* (ferredoxin promoter).

A version of pTA1228 has also been constructed for integration into the genome (Braun *et al.*, 2019). Like pTA1228, pTA1369 features a *his$_6$-tag* flanked by restriction sites that allow for in-frame insertion of a gene of interest downstream of the *p.tnaA* promoter (Figure 4.3). Multicloning sites enable the insertion of the upstream and downstream regions of a gene of interest either side of the *p.tnaA* promoter and the *his$_6$-tagged* gene of interest. This allows for integration of the plasmid and tryptophan induced expression at the gene locus.

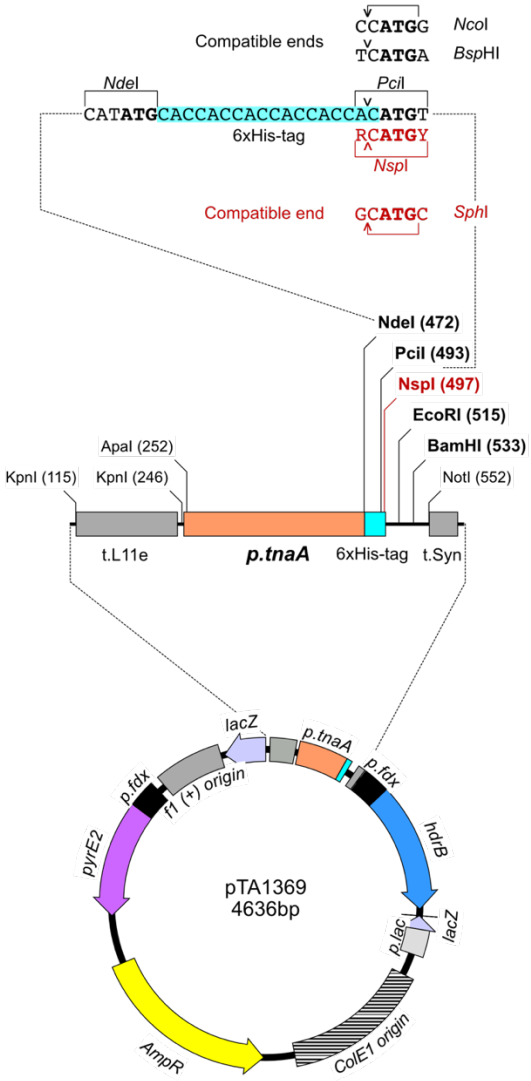


Figure 4.3: **pTA1369 for conditional overexpression of His$_6$-tagged proteins at the gene locus**. pTA1369 contains a *his$_6$-tag* downstream of the tryptophan inducible promoter *p.tnaA*. The genes *pyrE2* and *hdrB* allow for selection in *H. volcanii*. Restriction sites, and compatible enzymes, available for inserting genes of interest are displayed on the schematic. Multicloning sites either side of the *p.tnaA* promoter and the *his$_6$-tag* enable the insertion of the upstream and downstream regions of a gene of interest. Abbreviations: *t.L11e* (terminator), *p.tnaA* (tryptophan inducible promoter), *his$_6$-tag* (hexahistidine tag), *t.Syn* (terminator), *p.fdx* (ferredoxin promoter), *hdrB* (thymidine biosynthesis, *H. volcanii*), *lacZ* (gene containing multicloning site), *p.lac* (*lacZ* promoter), *ColE1* origin (*E. coli* origin of replication), *AmpR* (ampicillin resistance gene, *E. coli*), *pyrE2* (uracil biosynthesis, *H. volcanii*). Y = T/C and R = A/G.

The *p.tnaA* promoter utilised in the expression vectors pTA963 and pTA1228 is leaky in *E. coli* but not in *H. volcanii* (Large *et al.*, 2007; Allers *et al.*, 2010). Therefore, the construction of plasmids containing genes encoding halophilic proteins that are toxic to *E. coli* may need to be carried out directly in *H. volcanii*. Alternatively, utilisation of a less leaky promoter may be able to bypass this necessity. Two alternative promoters have been designed and utilised in protein expression from *H. volcanii*. The introduction of a point mutation in the *p.tnaA* promoter generates a more tightly-regulated and therefore less leaky promoter in *E. coli*, known as the *p.tnaM3* promoter (Figure 4.4). In addition, the *p.tnaM3* promoter is half as active than the *p.tnaA* promoter. The generic expression vector pTA1451 (like pTA1369 but with the *p.tnaA* promoter replaced with the *p.tnaM3* promoter) was constructed for the expression of proteins under the *p.tnaM3* promoter (Braun *et al.*, 2019). In *E. coli,* it is thought that the addition of tryptophan inhibits both TnaC peptidyl-tRNAPro hydrolysis and TnaC peptidyl transfer in ribosomes bearing nascent TnaC-peptidyl-tRNAPro. Consequently, the associated translating ribosome remains attached to the leader transcript. The bound ribosome blocks Rho factor binding and prevents subsequent transcription termination. Therefore, the tryptophan operon remains 'open' for further rounds of transcription (Gong *et al.*, 2001). Both the *p.tnaA* and *p.tnaM3* promoters are inducible by the addition of tryptophan, the third promoter utilised in protein expression from *H. volcanii* is constitutively active. The *p.syn* promoter is a highly active synthetic promoter useful for overexpression of protein for biochemical assays. The generic expression vector pTA1992 (like pTA1392 but with the *p.tnaA* promoter replaced with the *p.syn* promoter) was constructed for the expression of proteins under the *p.syn* promoter (Haque, Paradisi and Allers, 2019).

```
p.tnaA TGATCAGCTCGTGGCGAGAACGGAACAGCCGGCGACACCGATGCACACACCAGTCCACGAGCGCCGAA
p.tnaM3 TGATCAGCTCGTGGCGAGAACGGAACAGCCGGCGACACCGATGCACACACCAGTCCACGAGCGCCGAA

                                  *
AACCGGGCGTAGCCCCTCGATTTTCCGCCTGCCGATTACTTCACATTCGCGGACCTATTGCGCATATG
AACCGGGCGTAGCCCCTCGATTTTCCGCCTGCCGATGACTTCACATTCGCGGACCTATTGCGCATATG
```

Figure 4.4: **Sequence alignment of the *p.tnaA* and *p.tnaM3* promoters.** Point mutation of T to G within the TATA box (underlined) is indicated in **bold** and by * (red). This point mutation reduces promoter activity by 50%.

Overexpression Strains

Prior to the commencement of this study, the *H. volcanii* strains H1424 *(ΔpyrE2, ΔhdrB, Δmrr, Nph-pitA, cdc48d-Ct*) and H2047 *(ΔpyrE2, Δmrr, Nph-pitA, cdc48d-Ct, ΔtrpA*) were available for protein expression (Allers *et al.*, 2010; Stroud, Liddell and Allers, 2012; Wardell *et al.*, 2017). The *Δmrr, Nph-pitA,* and

cdc48d-Ct alleles were specifically generated for this technique and will be discussed in more detail below.

Mrr is a restriction enzyme encoded by *H. volcanii* that targets and cuts methylated 5'-GATC-3' sequences. Dam methylase encoded by *E. coli* methylates 5'-GATC-3' sequences. Transformation into a *mrr*+ strain of *H.volcanii* with *dam*+ DNA decreases the efficiency by 50x compared with transforming the same strain with *dam*- DNA. Therefore, for efficient transformation of plasmids constructed in *dam*+ *E. coli* into *H. volcanii*, the plasmid DNA must first pass through a *dam*- strain of *E.coli* (Holmes, Nuttall and Dyall-smith, 1991). Alternatively, efficient *dam*+ DNA transformations can be achieved (>10^7 transformants/µg DNA) by first deleting the *mrr* gene from the *H. volcanii* strain. To eliminate the need for passing protein over-expression plasmids through *dam- E.coli* (which are mutators), the *mrr* gene was deleted from the protein expression strain of *H. volcanii* (Allers, 2010).

Naturally histidine-rich proteins are major co-contaminants when purifying proteins via nickel-affinity chromatography. *H. volcanii*, like most halophilic organisms, encodes many histidine-rich proteins. The naturally-occurring histidine rich contaminants hinder protein purification in *H. volcanii* and interfere with the detection of meaningful protein:protein interactions by protein pull down assays.

One of the main co-contaminants is PitA, an essential protein that contains a histidine rich linker region (Bab-Dinitz *et al.*, 2006; Humbard, Zhou and Maupin-Furlow, 2009). To overcome this problem, the *H. volcanii* histidine-rich linker was replaced with a non-histidine-rich orthologue from *Natronomonas pharaonis* (Allers *et al.*, 2010). The elimination of the PitA contaminant exposed Cdc48d (HVO_1907), another histidine-rich contaminant of the same size. A truncated version of the *cdc48d* gene removes the histidine-rich C-terminal domain and reduces the issue of co-contamination during nickel affinity chromatography (Stroud, Liddell and Allers, 2012). Complete deletion of these two genes from *H. volcanii* is not possible as both *pitA* and *cdc48d* are essential.

Despite the removal of the two major histidine-rich contaminants PitA and Cdc48d, many other histidine-rich proteins still persist during nickel affinity chromatography (Stroud, Liddell and Allers, 2012) (Figure 4.5). For this reason, a second high-salt compatible tag, the StrepII-tag, is often used in conjunction with the His_6-tag.

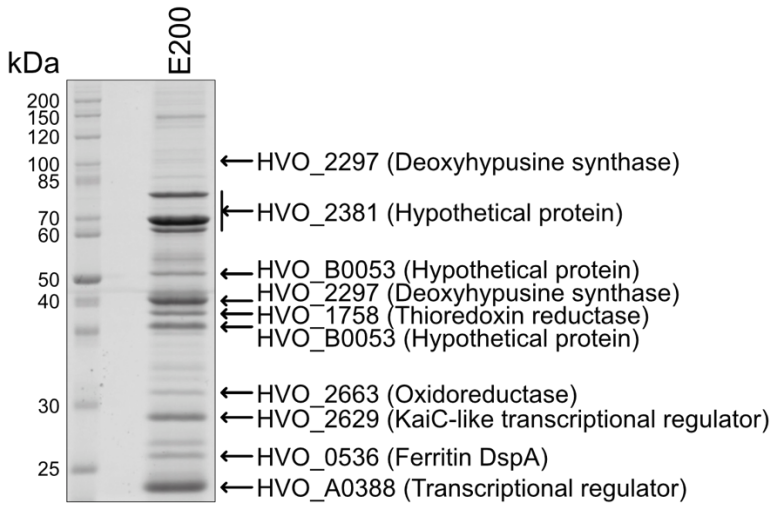


Figure 4.5: **Histidine-rich contaminant proteins in nickel affinity chromatography.** A mock protein overexpression expressing the empty vector protein expression plasmid pTA963. Histidine-rich cellular proteins were purified from soluble lysate fraction by affinity chromatography on a Ni^{2+}chelating column. Bound proteins were eluted using 200 mM imidazole. Proteins were precipitated using trichloroacetic acid and deoxycholate to aid visualisation. Proteins were identified via mass spectrometry, *H. volcanii* gene numbers indicated. Adapted from Stroud et al (2012).

The development of the generic episomal overexpression constructs and the expression strains has allowed for the expression and purification of His_6-tagged and StrepII-tagged recombinant proteins in *H. volcanii*. Due to the high abundance of naturally histidine-rich proteins in *H. volcanii*, contamination during nickel-affinity chromatography remains problematic. To minimise contaminants, application to a StrepTactin column for a second purification step may be employed. However, the StrepTactin column cannot tolerate imidazole, the compound used to elute protein from the nickel-column, and therefore purifying a protein in this sequential manner requires a change of buffer between columns. Each purification step decreases the yield of the protein of interest which can be problematic, especially if the protein of interest is toxic at high concentrations and therefore difficult to express. Additionally, the presence of either tag at the C-terminus of the protein of interest often has a detrimental effect on the growth of the strain, suggesting that the tag is interfering with protein function. If the protein of interest is not functional during pulldown assays, no meaningful protein: protein interactions can be identified.

Prior to the commencement of this study, 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 explore protein:protein interaction partners of Hel308 within the cell. The chromosomal N-terminally *his_6-tagged* and C-terminally *strepII-tagged* hel308 strain was generated in the H2047 *(ΔpyrE2, Nph-pitA, Δmrr, cdc48d-ct, ΔtrpA*) background. The chromosomal *his_6-hel308-*

strepII is under control of the natural promoter of *hel308* resulting in native levels of expression. However, the presence of the StrepII-tag at the C-terminus of the Hel308 has a detrimental effect on the growth of the strain. H2418 (Δ*pyrE2, Nph-pitA,* Δ*mrr, cdc48d-ct,* Δ*trpA, his*$_6$*-hel308-strepII)* has an average generation time of ~8 hours, in contrast with ~2 hours for H2047, suggesting that the StrepII-tag is interfering with protein function.

4.2 Aims and Objectives

It is of interest to continue to develop methods for protein expression and purification in *H. volcanii*. The objectives of this chapter were to:

- Determine which His$_x$-tag and StrepII-tag combination for Hel308 has little to no detrimental effect on growth of the strain and therefore will not interfere with protein function.
- Determine the most efficient method of purification, with minimal contaminants, for the natively expressed His$_x$-tagged and StrepII-tagged Hel308 in *H. volcanii*.
- Screen for *in vivo* protein:protein interactions of Hel308 by expressing His$_x$-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.
- Develop generic strains, plasmids and techniques for the tightly-controlled overexpression and purification of His$_x$-tagged and StrepII-tagged recombinant proteins that are toxic to the cell in high concentrations.

4.3 Materials

All strains and plasmids were constructed by the author unless stated otherwise.

4.3.1 Strains

Table 4.1 ***Haloferax volcanii*** **strains used in this chapter.**

Parent: H2047 (*ΔpyrE2, Nph-pitA, Δmrr, cdc48d-Ct, ΔtrpA*)			
Strain	**Parent**	**Genotype**	**Notes**
H2131	H2130	*Δhel308::trpA+, pyrE2+*	Constructed by RG-M
H2418	H2372	*his6 tag-hel308-strepII tag*	Constructed by RG-M
H3360	H3045	*hel308-strepII tag*	
H3362	H3047	*his6 tag-hel308*	
H3363	H3048	*strepII tag-hel308-his6 tag*	
H3365	H3050	*his7-strepII-strepII-tag-hel308*	

H3367	H3052	*strepII-strepII tag-hel308*
H3369	H3054	*hel308-his6 tag*
H3371	H3056	*strepII tag-hel308 tag*

4.3.2 Plasmids

Table 4.2 **Plasmids used in this chapter.**

Name	Use	Notes
pTA1273	*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-D145N* gene replacement construct	Constructed by TA
pTA1425	N-terminally his6-tagged and a C-terminally StrepII-tagged *hel308* gene replacement construct, under the control of p.tnaA	Constructed by RG-M
pTA1642	*hel308-F316A* gene replacement construct	Constructed by RG-M
pTA1643	*hel308-R743A* gene replacement construct	Constructed by RG-M
pTA1647	*hel308-F316A* gene replacement construct	Constructed by RG-M
pTA1648	*hel308-R743A* gene replacement construct	Constructed by RG-M
pTA1661	Intermediate construct for creating pTA1662	Constructed by RG-M
pTA1662	N-terminally his6-tagged and a C-terminally StrepII-tagged *hel308* gene replacement construct	Constructed by RG-M
pTA1718	C-terminally StrepII-tagged hel308 gene replacement construct	
pTA1719	N-terminally his6-tagged hel308 gene replacement construct	
pTA1771	Intermediate construct used to create pTA1849	Constructed by TA
pTA1797	Intermediate construct used to create pTA1850	Constructed by TA
pTA1849	Intermediate construct used to create pTA1857	
pTA1850	Intermediate construct used to create pTA1857	

pTA1855	N-terminally StrepII-tagged and a C-terminally His6-tagged *hel308* gene replacement construct
pTA1856	N-terminally His7-2xStrepII-tagged *hel308* gene replacement construct
pTA1857	N-terminally 2xStrepII-tagged *hel308* gene replacement construct
pTA1859	C-terminally His6-tagged *hel308* gene replacement construct
pTA1860	N-terminally StrepII-tagged *hel308* gene replacement construct
pTA1945	*hel308-K53R-F316A* gene replacement construct
pTA1946	*hel308-K53R-R743A* gene replacement construct
pTA1947	*hel308-D145N-R743A* gene replacement construct
pTA1952	*hel308-D145N-F316A* gene replacement construct
pTA1989	*hel308-K53A-F316A* gene replacement construct
pTA1990	*hel308-K53A-R743A* gene replacement construct
pTA2010	*hel308-D145A* gene replacement construct
pTA2096	Intermediate construct for creating pTA2118
pTA2097	Intermediate construct for creating pTA2127
pTA2112	Intermediate construct for creating pTA2122
pTA2118	p.tnaM3-His7-2xStrepII expression vector
pTA2121	p.tnaM3-His7-2xStrepII-Hel308 expression vector
pTA2122	p.tnaM3-His7-2xStrepII-Hel308-D145N expression vector
pTA2127	p.syn-His7-2xStrepII expression vector
pTA2184	Intermediate construct for creating pTA2315

pTA2200	p.tnaM3-Hel308-K53A expression vector
pTA2201	p.tnaM3-His7-2xStrepII-Hel308-F316A expression vector

TA = Thorsten Allers, RG-M = Rebecca Gamble-Milner

pTA1316 *hel308* (HVO_0014)

The *hel308* genomic clone, pTA1316, was constructed prior to this study by Thorsten Allers in 2011 (Gamble-Milner, 2016) (Figure 4.6).

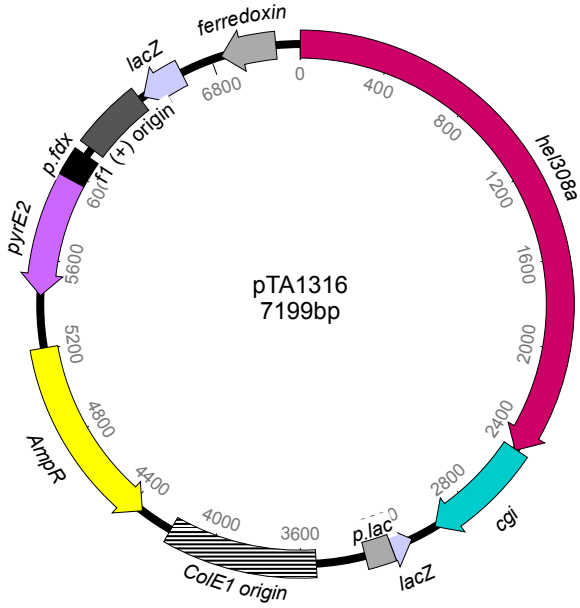


Figure 4.6: **pTA1316.** *hel308, cgi* genomic clone. Constructed by Thorsten Allers in 2011 (Gamble-Milner, 2016)

pTA1425 (*strepII-hel308-his$_6$* episomal over-expression construct)

The *strepII -hel308-his$_6$* episomal over-expression construct, pTA1425, was constructed prior to this study by Rebecca Gamble-Milner in 2013 (Gamble-Milner, 2016) (Figure 4.7).

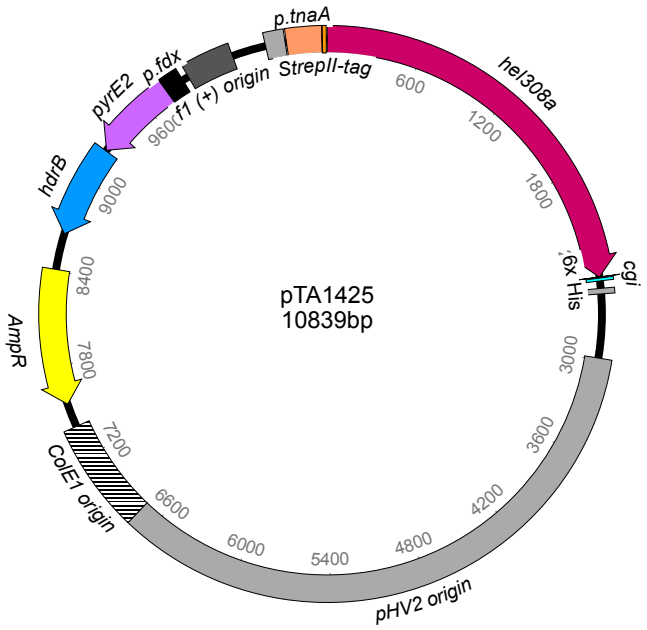


Figure 4.7: **pTA1425.** *strepII-hel308-his$_6$* episomal over-expression construct. Constructed by Rebecca Gamble-Milner in 2013 (Gamble-Milner, 2016)

pTA1451 (*p.tnaM3-his$_6$*-tagged gene episomal over-expression construct)

The generic gene tagging construct, pTA1451, N-terminally tags the gene of interest with the *his$_6$- tag* and also allows for the introduction of the upstream and downstream regions of the gene for chromosomal integration. pTA1451 was constructed prior to this study by Hannah Marriott in 2013 (Braun *et al.*, 2019) (Figure 4.8).

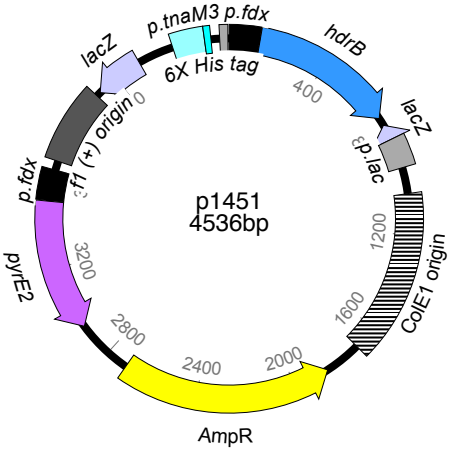


Figure 4.8: **pTA1451.** Generic *p.tnaM3-his$_6$-strepII*-tagged gene over-expression construct. Constructed by Hannah Marriott in 2013 (Braun *et al.*, 2019)

pTA1661 (*hel308* gene deletion construct)

The *hel308* gene deletion construct, pTA1661, was constructed prior to this study by Rebecca Gamble-Milner in 2015 (Gamble-Milner, 2016) (Figure 4.9).

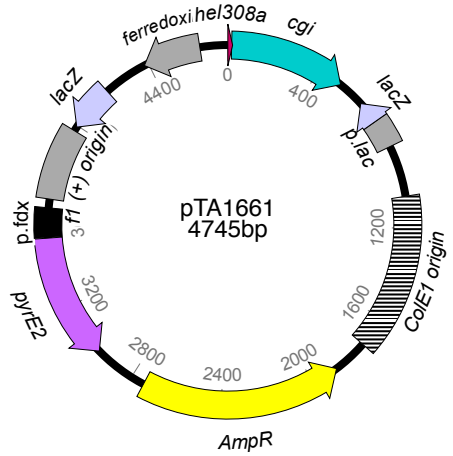


Figure 4.9: **pTA1661.** *hel308* deletion construct. Constructed by Rebecca Gamble-Milner in 2015 (Gamble-Milner, 2016)

pTA1662 (*his$_6$-hel308-strepII* gene replacement construct)

The *his$_6$-hel308-strepII* gene replacement construct, pTA1662, was constructed prior to this study by Rebecca Gamble-Milner in 2015 (Gamble-Milner, 2016) (Figure 4.10).

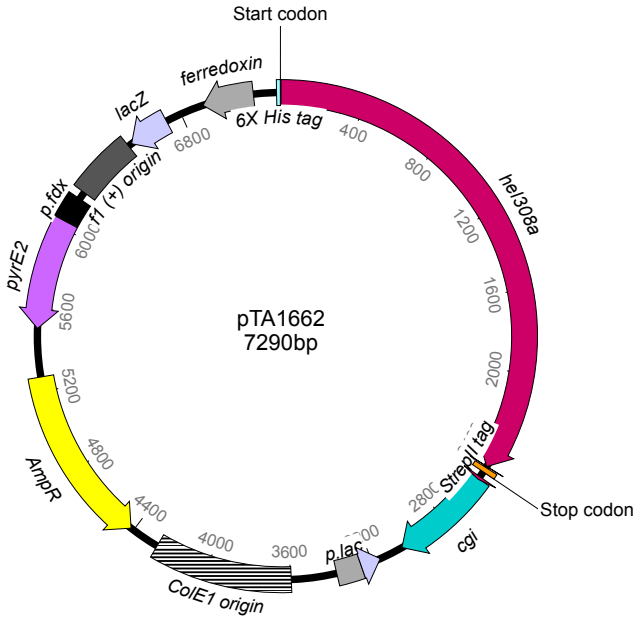


Figure 4.10: **pTA1662.** *his$_6$-hel308-strepII* gene replacement construct. Constructed by Rebecca Gamble-Milner in 2015 (Gamble-Milner, 2016)

pTA1771 (N-terminally *his$_7$-strepII-strepII*-tagged gene construct)

The generic gene tagging construct, pTA1771, N-terminally tags the gene of interest with the tandem *his$_7$-strepII-strepII-tag* and also allows for the introduction of the upstream and downstream regions of the gene for chromosomal integration. pTA1771 was constructed by Thorsten Allers in 2016 (Figure 4.11).

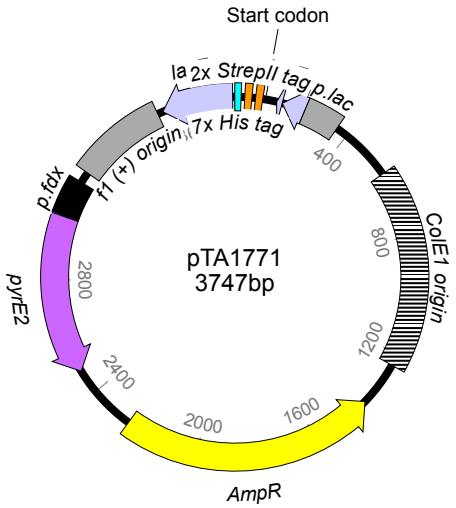


Figure 4.11: **pTA1771.** Generic N-terminally *his$_7$-strepII-strepII*-tagged gene construct. Constructed by Rebecca Thorsten Allers in 2016

pTA1797 (N-terminally *strepII-strepII*-tagged gene construct)

The generic gene tagging construct, pTA1797, N-terminally tags the gene of interest with the tandem *strepII-strepII-tag* and also allows for the introduction of the upstream and downstream regions of the gene for chromosomal integration. pTA1797 was constructed by Thorsten Allers in 2016 (Figure 4.12).

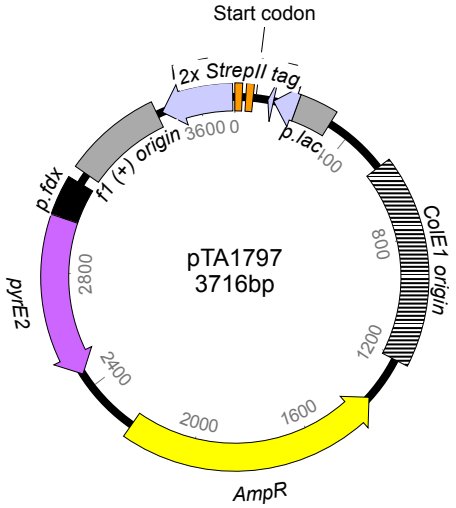


Figure 4.12: **pTA1771.** Generic N-terminally *strepII-strepII*-tagged gene construct. Constructed by Rebecca Thorsten Allers in 2016

4.3.3 Oligonucleotides

Table 4.3 **Oligonucleotides used in this chapter.**

Name	Sequence (5'-3')	Description	Used to construct
pBSF2	TTAAGTTGGGTAACGCCAGGG	Amplification of the upstream region to multiple genes: *recJ2, recJ4;* and *hel308a*. Sequencing primer	pTA1923, pTA1951, pTA1991, K53A plasmids
Hel308 NdeI_F	GGGAGATcatATGCGAACTGCGGACCTGACGGGC	Incorporation of a *Nde*I site upstream of h*el308a* start codon: to clone downstream of a tryptophanase promoter	pTA1854
Hel308SphF	GGGAGATGgcATGCGAACTGCGGACCTGACGGGC	Incorporation of a *Sph*I site upstream of h*el308a* start codon: to clone downstream of a tryptophanase promoter	pTA1849, pTA1850
helQ(R) BamHI 02	CGAGGAtCCTCATTCGAAATCACCCAGACTGG	Amplification of *hel308a* and introduction of a *Bam*HI site	pTA1849, pTA1850, pTA1854
oriC-1trp-R	CGGCCGCTCTAGAACTAGTGG	Amplification of 2626bp region of pTA1425 containing *StrepII-hel308-6His* CDS	pTA1855
PtnaAFint	GCCTGCCGATTACTTCACATTCGC	Amplification of 2626bp region of pTA1425 containing *StrepII-hel308-6His* CDS	pTA1855
hel308EcoR	AGGTAGTCGAGCACGCGGTCC	Amplification of internal region of *hel308a*	pTA1952, pTA1965
Hel308 Fint	AGCGCTGGGAGGAGTACGGC	Amplification of internal region of *hel308a*	pTA1952, pTA1965
Hel308 F316AF	CGCGGCGgcCCACCACGCGGGACTCGCCG	Generation of a *hel308-F316A* mutation	pTA1952, pTA1965
Hel308 F316AR	CGTGGTGGgcCGCCGCGCCTTTGGCGACC	Generation of a *hel308-F316A* mutation	pTA1952, pTA1965
Hel308 RTR	GTGTTGACGCCGGCGGCGAGCG	Amplification of internal region of *hel308a*	K53A plasmids
Hel308 K53AF	CGAGCGGgccGACGCTCATCGCCGAACTGGCG	Generation of a *hel308-K53A* mutation	K53A plasmids

Hel308 K53AR	CGATGAGCGTCg gcCCGCTCGCCG TCGGG	Generation of a *hel308-K53A* mutation	K53A plasmids

4.4 Results

4.4.1 Development of Chromosomally-Tagged Expression Strains

A variety of strains were generated to explore different combinations and placements of the *his$_x$*- and *strepII-tags* on *hel308*, in place of the genomic copy of *hel308* under the natural promoter for *hel308*.

pTA1718 (*hel308-strepII* gene replacement construct)

pTA1316 was digested with the restriction enzymes *Nsi*I and *Xba*I to remove the 2082 bp fragment at the 3' end of *hel308.* pTA1662 was digested with the restriction enzymes *Nsi*I and *Xba*I to produce a 2152 bp fragment containing the *strepII-tagged* 3'-portion of *hel308* (and *cgi*). The *3'hel308-strepII* fragment was ligated into the pTA1316 vector to generate *hel308-strepII* gene replacement construct, pTA1718 (Figure 4.13).

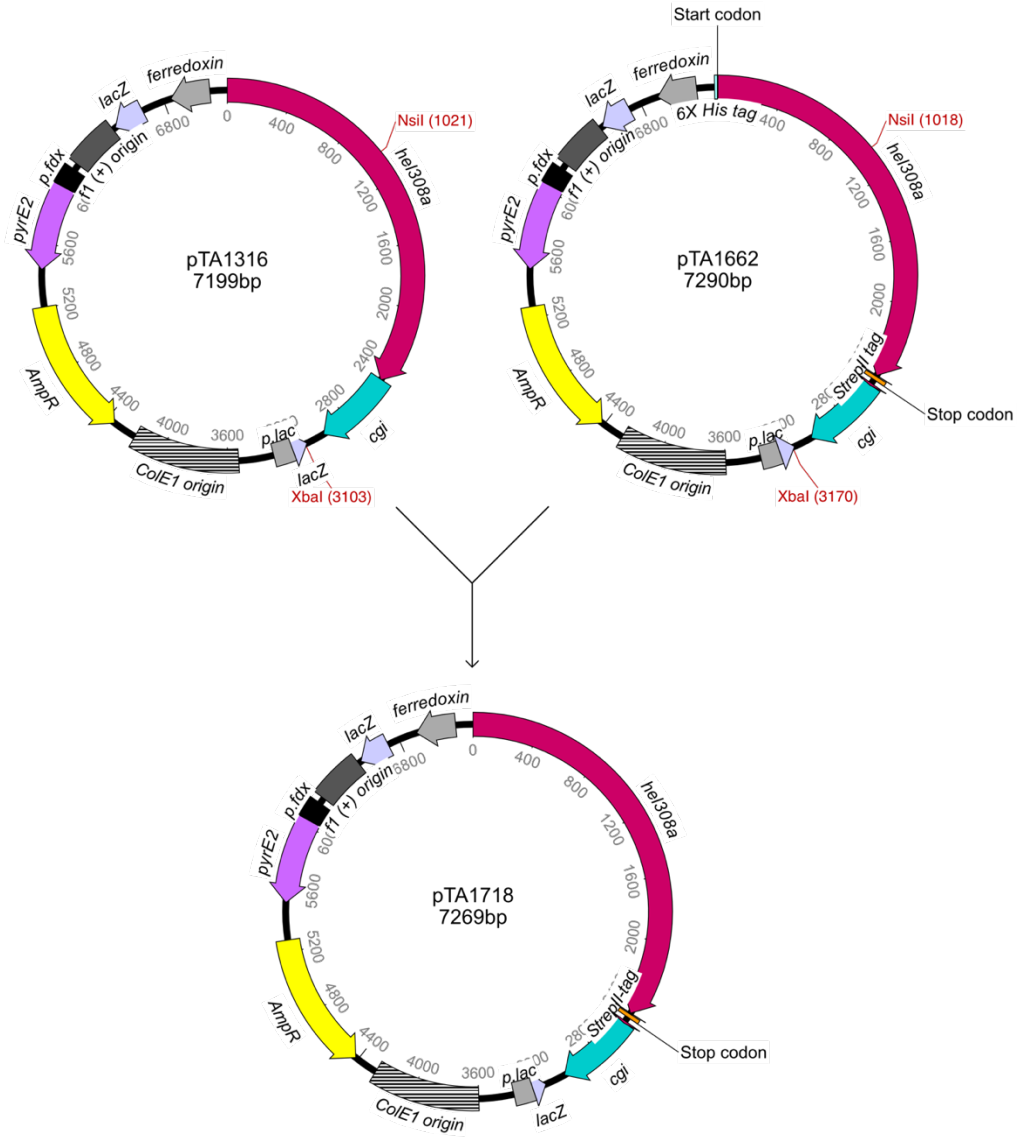


Figure 4.13: **Construction of C-terminally *strepII-tagged hel308* gene replacement construct, pTA1718.** 3' *hel308-strepII* was digested from pTA1662 using the *Nsi*I and *Xba*I restriction sites and ligated into pTA1316 at the *Nsi*I and *Xba*I restriction sites, to generate the *hel308-strepII* gene replacement construct , pTA1718.

The strain H2131 (Δ*pyrE2, Nph-pitA,* Δ*mrr, cdc48d-ct,* Δ*trpA,* Δ*hel308::trpA+*) was constructed by Rebecca Gamble-Milner in 2015, prior to this study (Gamble-Milner, 2016). The plasmid pTA1718 was used to replace Δ*hel308::trpA+* with a C-terminally *strepII-tagged hel308* to generate the strain H3360. The strain was verified via Southern blot (Figure 4.14).

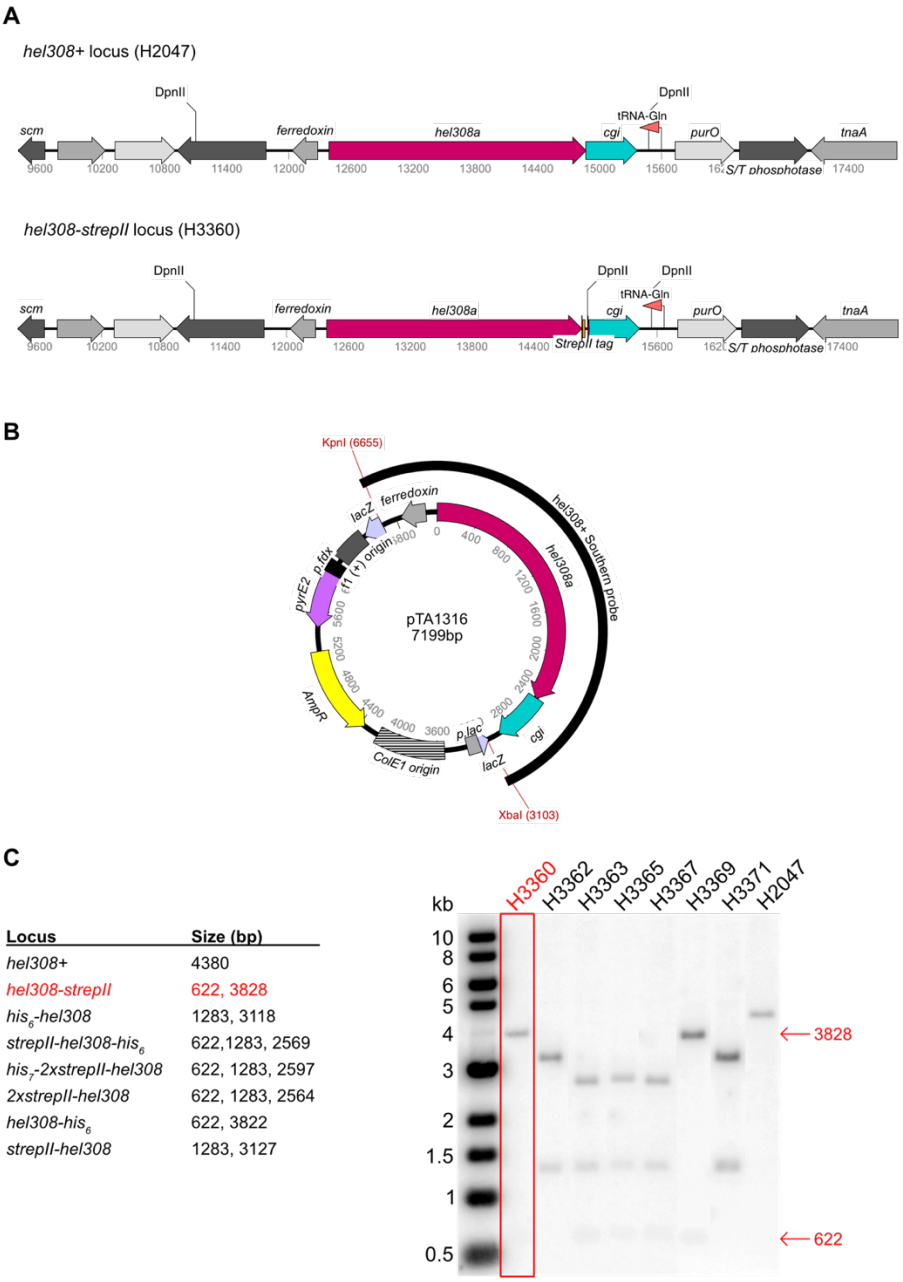


Figure 4.14 **Gene replacement of Δ*hel308::trpA*+ with *hel308-strepII* in H2131 (Δ*pyrE2, Nph-pitA,* Δ*mrr, cdc48d-ct,* Δ*trpA,* Δ*hel308::trpA*+).** A) Gene locus for *hel308*+ and *hel308-strepII*. The *Dpn*II sites used to digest genomic DNA are shown. B) A 3647bp *hel308*+ probe for southern hybridisation was made by digesting pTA1316 with *Kpn*I and *Xba*I. C) Predicted locus size and Southern blot, highlighted and boxed in red, showing H3360 is *hel308-strepII*+.

pTA1719 (*his*$_6$*-hel308* gene replacement construct)

The hel308 genomic clone, pTA1316, was digested with the restriction enzymes *Kpn*I and *Nsi*I to remove the 1565 bp fragment at the 5' end of *hel308.* pTA1662 was also digested with the restriction enzymes *Kpn*I and *Nsi*I to produce a 1586

bp fragment containing the *his$_6$-tagged* 5'-portion of *hel308*. The *5'his$_6$-hel308* fragment was ligated into the pTA1316 vector to generate *his$_6$-hel308* gene replacement construct, pTA1719 (Figure 4.15).

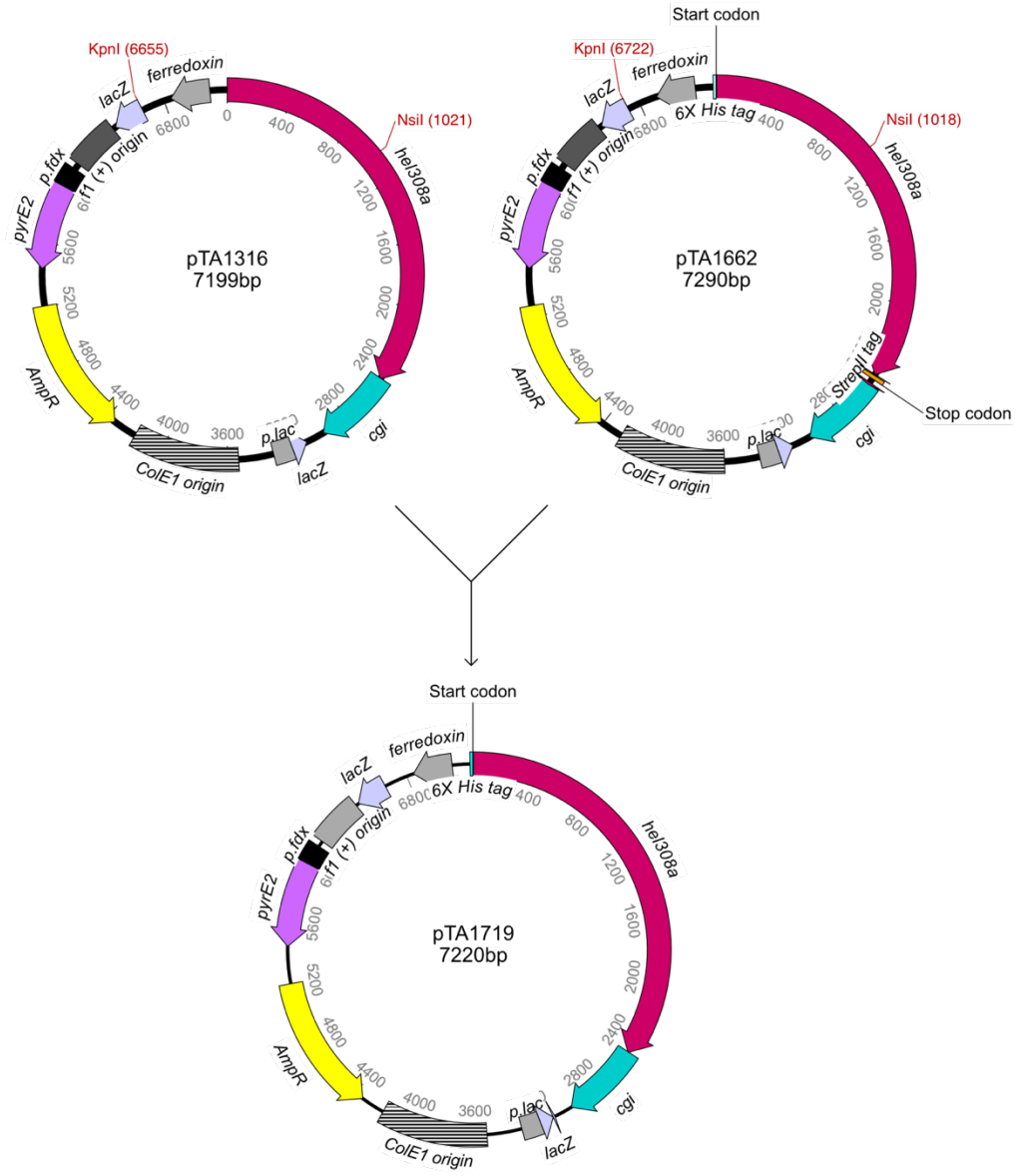


Figure 4.15: **Construction of N-terminally *his$_6$-tagged hel308* gene replacement construct, pTA1719.** 5' *his$_6$-hel308* was digested from pTA1662 using the *KpnI* and *Nsi*I restriction sites and ligated into pTA1316 at the *Nsi*I and *Xba*I restriction sites, to generate the *his$_6$-hel308* gene replacement construct , pTA1719.

An N-terminally *his$_6$-tagged hel308* was introduced into H2131 using the gene replacement construct pTA1719 to generate the strain H3362. The strain was confirmed via Southern blot (Figure 4.16).

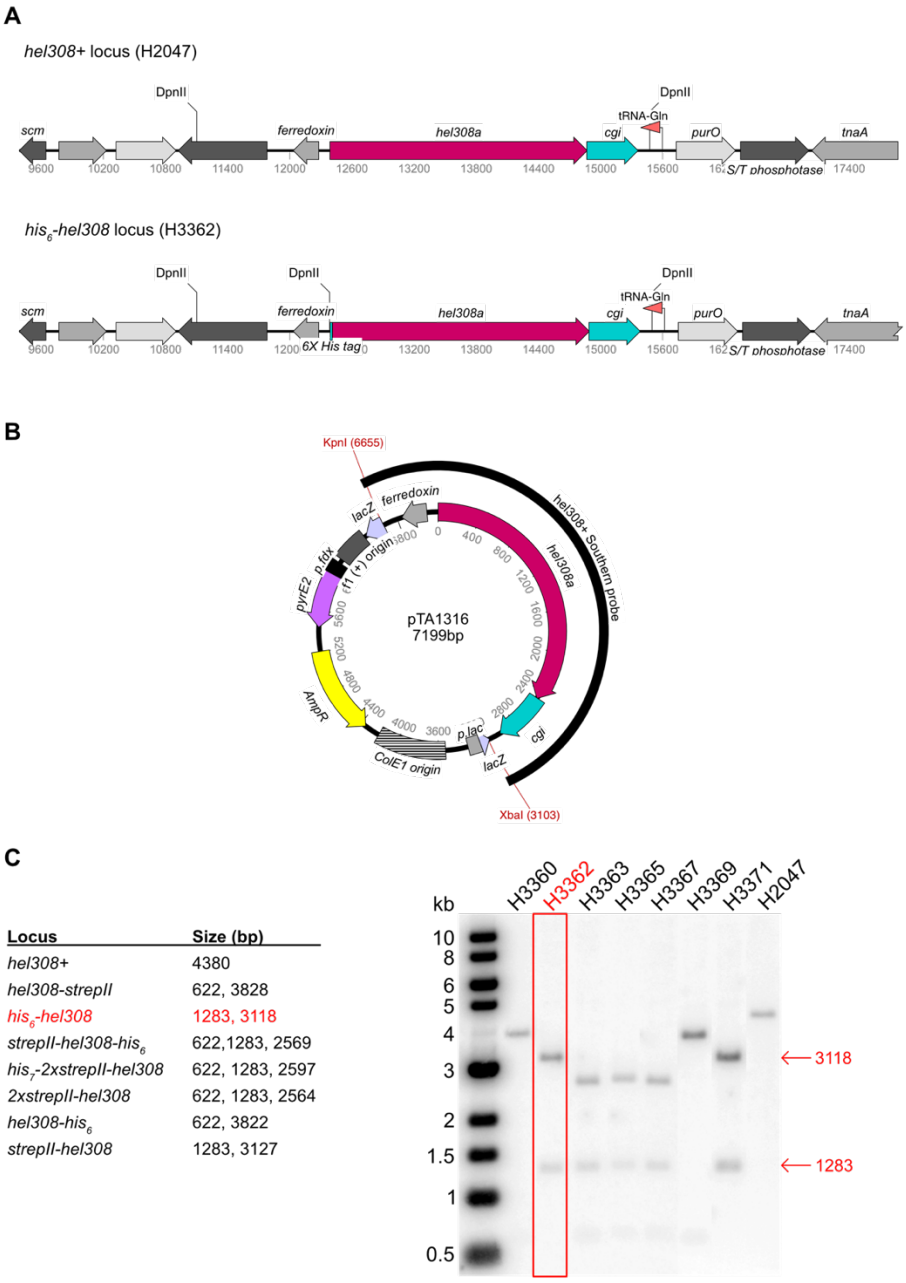


Figure 4.16 **Gene replacement of Δ*hel308::trpA*+ with *his$_6$-hel308* in H2131 (Δ*pyrE2, Nph-pitA,* Δ*mrr, cdc48d-ct,* Δ*trpA,* Δ*hel308::trpA*+).** A) Gene locus for *hel308*+ and *his$_6$-hel308*. The *Dpn*II sites used to digest genomic DNA are shown. B) A 3647bp *hel308*+ probe for southern hybridisation was made by digesting pTA1316 with *Kpn*I and *Xba*I. C) Predicted locus size and Southern blot, highlighted and boxed in red, showing H3362 is *his$_6$-hel308*+.

pTA1855 (*strepII-hel308-his$_6$* gene replacement construct)

The *strepII-hel308-his$_6$* episomal over-expression construct, pTA1425, was used as a template to PCR amplify the *strepII-hel308-his$_6$* CDS using the primers PtnaAFint and oriC-1trp-R. The product was digested with *Nde*I and *Bam*HI and

ligated into the *Nde*I and *Bam*HI sites in pTA1661 to create the *strepII-hel308-his*$_6$ gene replacement construct, pTA1855 (Figure 4.17).

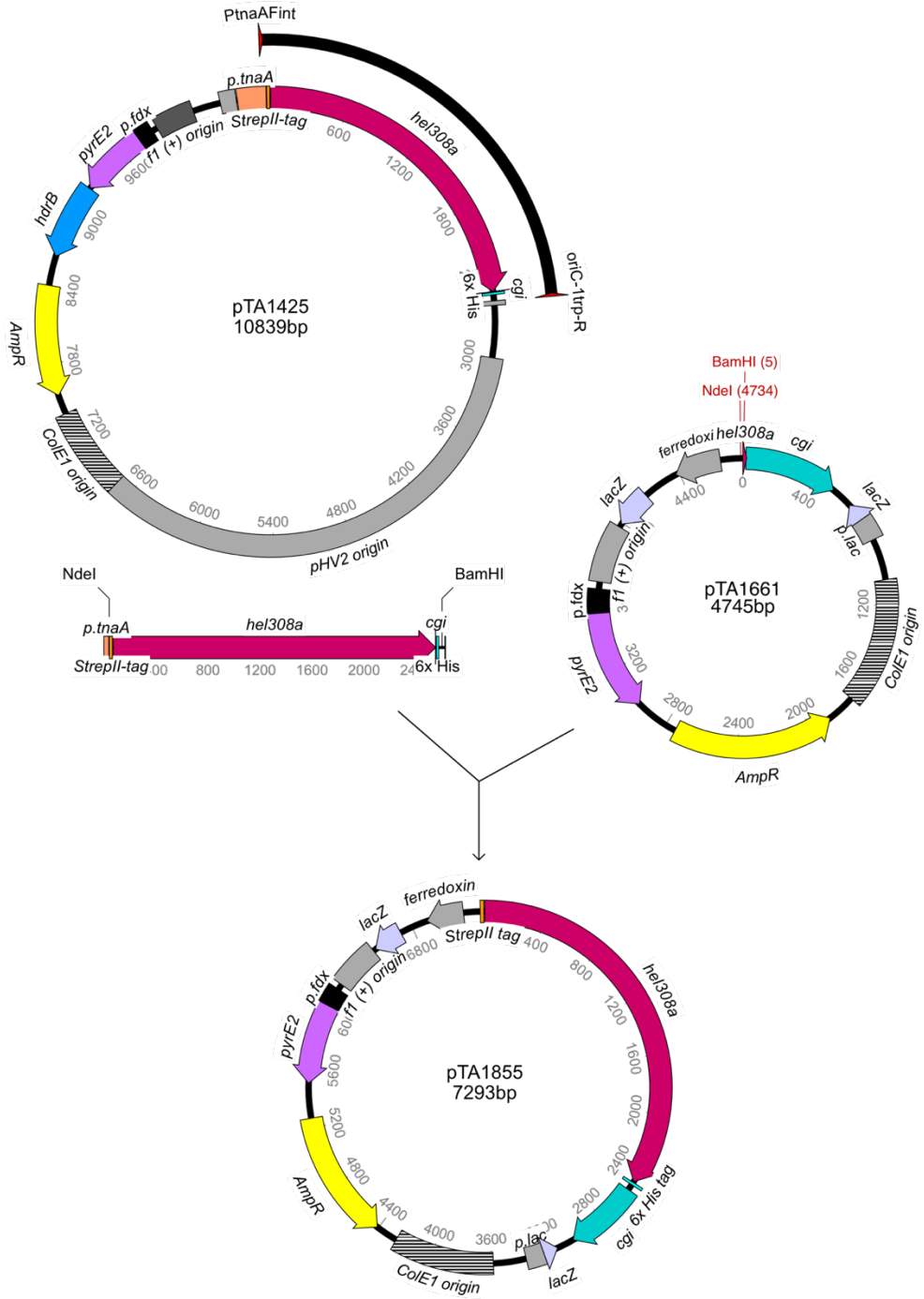


Figure 4.17: **Construction of N-terminally *strepII-tagged and* C-terminally *his*$_6$*-tagged hel308* gene replacement construct, pTA1855.** pTA1425 was used as a template to PCR amplify from PtnaAFint to oriC-1trp-R. The product was digested with *Nde*I and *Bam*HI and ligated into the *Nde*I and *Bam*HI sites in pTA1661 to generate the *strepII-hel308-his*$_6$ gene replacement construct, pTA1855.

An N-terminally *strepII-tagged* and C-terminally *his$_6$-tagged hel308* was introduced into H2131 using the gene replacement construct pTA1855 to generate the strain H3363. The strain was confirmed via Southern blot (Figure 4.18).

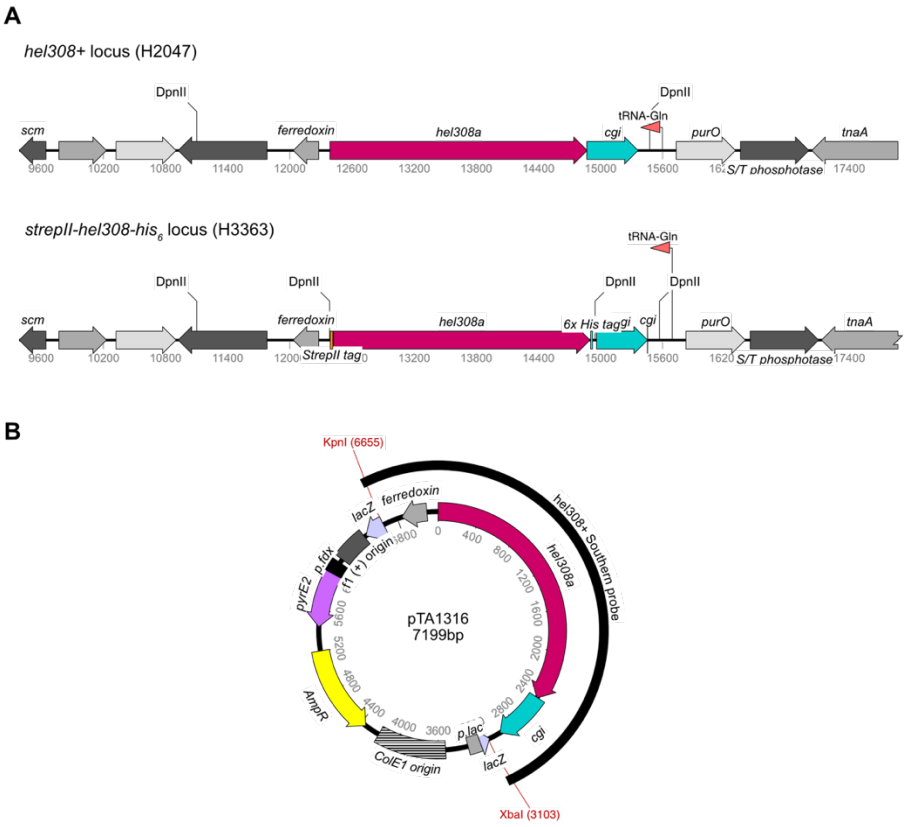


Figure 4.18 **Gene replacement of Δ*hel308::trpA*+ with *strepII-hel308-his$_6$* in H2131 (Δ*pyrE2, Nph-pitA,* Δ*mrr, cdc48d-ct,* Δ*trpA,* Δ*hel308::trpA*+).** A) Gene locus for *hel308+* and *strepII-hel308-his$_6$*. The *Dpn*II sites used to digest genomic DNA are shown. B) A 3647bp *hel308+* probe for southern hybridisation was made by digesting pTA1316 with *Kpn*I and *Xba*I. C) Predicted locus size and Southern blot, highlighted and boxed in red, showing H3363 is *strepII-hel308-his$_6$+*.

pTA1856 (*his$_7$-strepII-strepII-hel308* gene replacement construct)

The hel308 genomic clone, pTA1316, was used as a template to PCR amplify the *hel308* CDS using the primers hel308SphF and helQ(R)BamHI02 introducing *Sph*I and *Bam*HI sites respectively. The product was digested with *Sph*I and *Bam*HI and was ligated into the *Nsp*I and *Bam*HI sites in pTA1771 to create the intermediary plasmid, pTA1849. The *his$_7$-strepII-strepII-hel308* CDS was digested from pTA1849 with *Nde*I and *Bam*HI and ligated into the *Nde*I and *Bam*HI sites in pTA1661 to create the *his$_7$-strepII-strepII-hel308* gene replacement construct, pTA1856 (Figure 4.19).

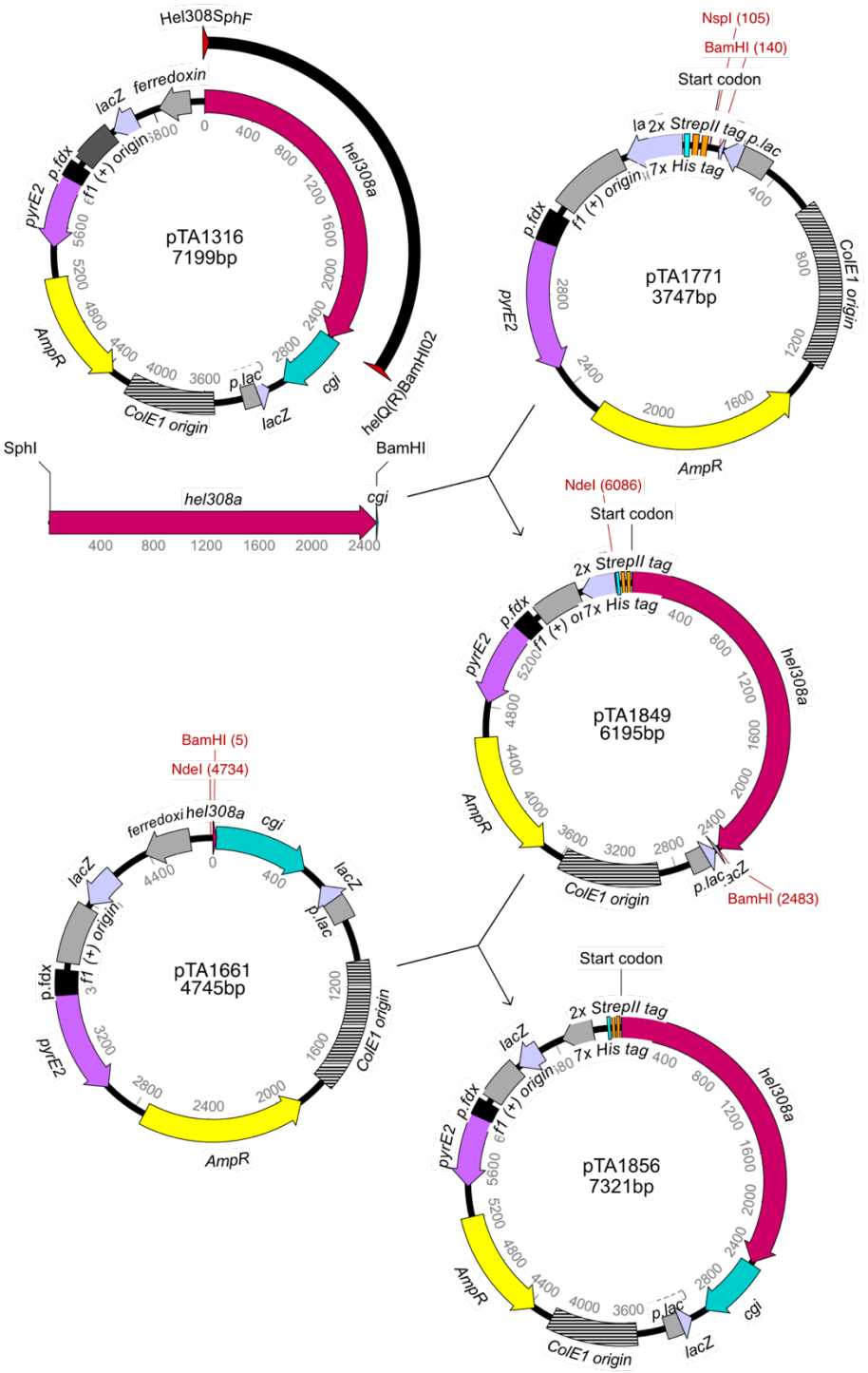


Figure 4.19: **Construction of N-terminally *his*$_7$*-strepII-strepII-tagged hel308* gene replacement construct, pTA1856.** pTA1316 was used as a template to PCR amplify from hel308SphF to helQ(R)BamHI02 introducing *Sph*I and *Bam*HI sites respectively. The product was digested with *Sph*I and *Bam*HI and was ligated into the *Nsp*I and *Bam*HI sites in pTA1771 to create the intermediary plasmid, pTA1849. pTA1849 was digested with *Nde*I and *Bam*HI and ligated into the *Nde*I and *Bam*HI sites in pTA1661 to create the *his*$_7$*-strepII-strepII-hel308* gene replacement construct, pTA1856.

An N-terminally *his*$_7$*-strepII-strepII-tagged hel308* was introduced into H2131 using the gene replacement construct pTA1856 to generate the strain H3365. The strain was verified via Southern blot (Figure 4.20).

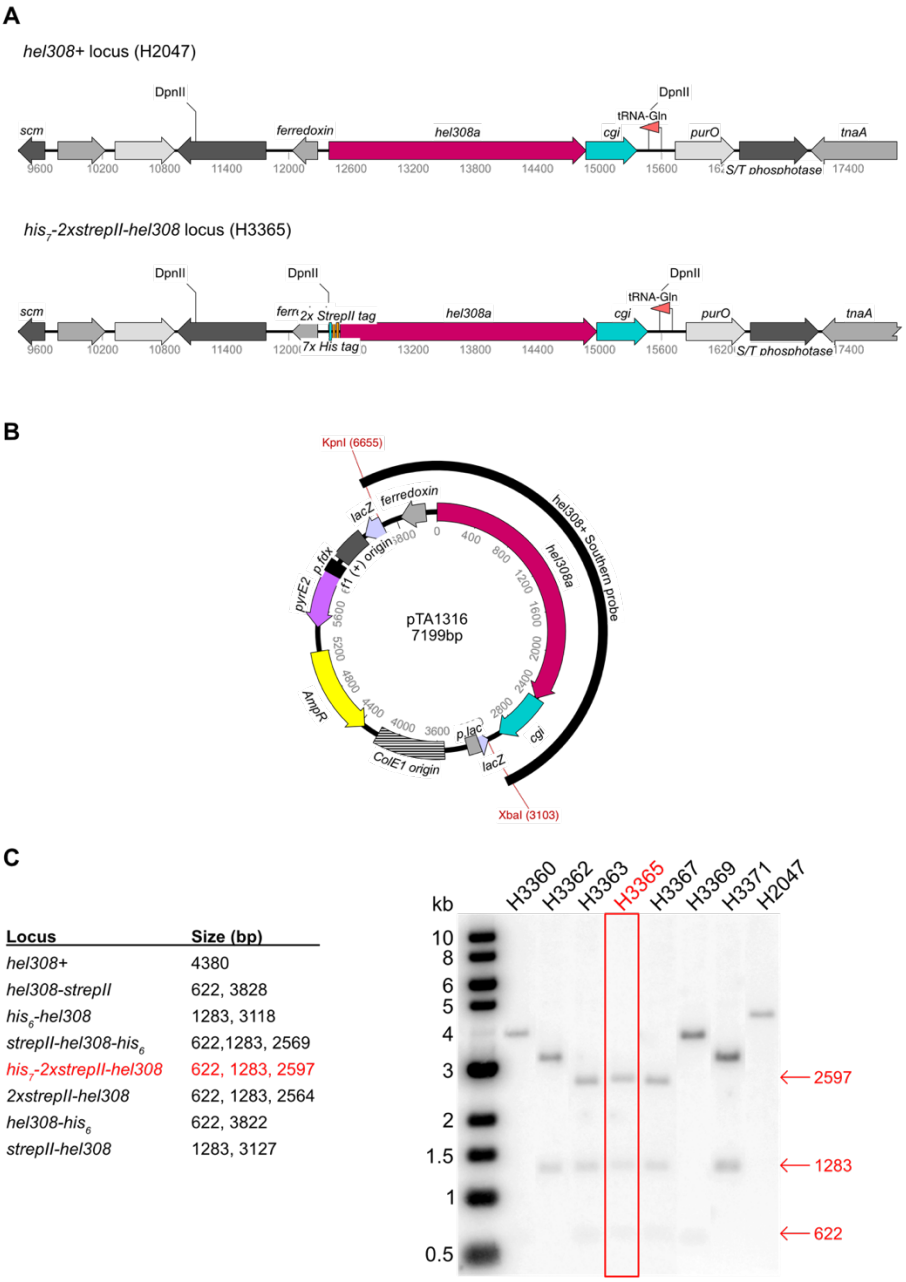


Figure 4.20 **Gene replacement of Δ*hel308::trpA*+ with *his*$_7$*-strepII-strepII-hel308* in H2131 (Δ*pyrE2, Nph-pitA,* Δ*mrr, cdc48d-ct,* Δ*trpA,* Δ*hel308::trpA*+).** A) Gene locus for *hel308*+ and *his*$_7$*-strepII-strepII-hel308*. The *Dpn*II sites used to digest genomic DNA are shown. B) A 3647bp *hel308*+ probe for southern hybridisation was made by digesting pTA1316 with *Kpn*I and *Xba*I. C) Predicted locus size and Southern blot, highlighted and boxed in red, showing H3365 is *his*$_7$*-strepII-strepII-hel308*+.

pTA1857 (*strepII-strepII-hel308* gene replacement construct)

The hel308 genomic clone, pTA1316, was used as a template to PCR amplify the *hel308* CDS using the primers hel308SphF and helQ(R)BamHI02 introducing *Sph*I and *Bam*HI sites respectively. The product was digested with *Sph*I and *Bam*HI and ligated into the *Nsp*I and *Bam*HI sites in pTA1797 to create the plasmid, pTA1850. The *strepII-strepII-hel308* CDS was digested from pTA1850 using *Nde*I and *Bam*HI and ligated into the *Nde*I and *Bam*HI sites in pTA1661 to create the *strepII-strepII-hel308* gene replacement construct, pTA1857 (Figure 4.21).

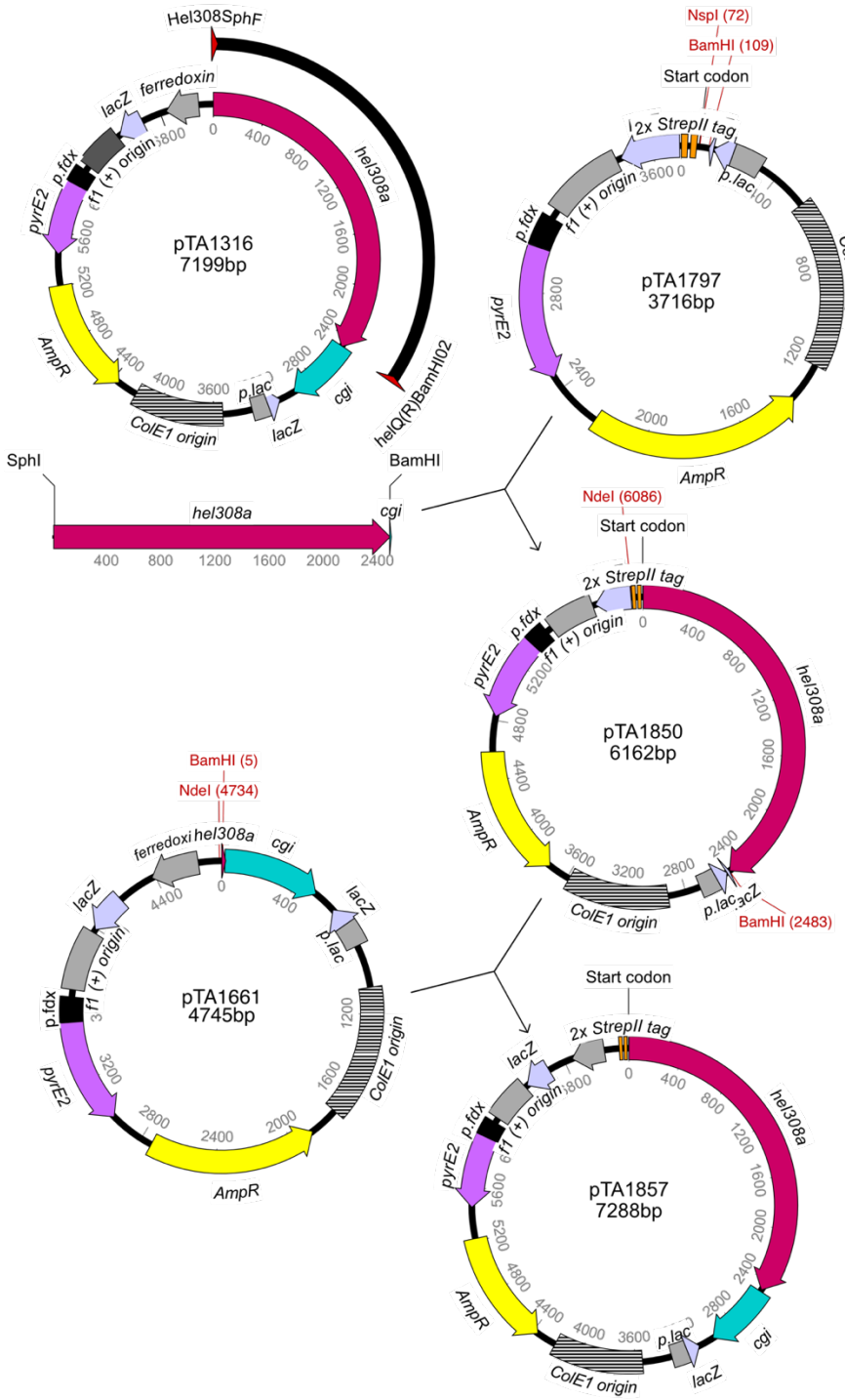


Figure 4.21: **Construction of N-terminally *strepII-strepII-tagged hel308* gene replacement construct, pTA1857.** pTA1316 was used as a template to PCR amplify from hel308SphF to helQ(R)BamHI02 introducing *Sph*I and *Bam*HI sites respectively. The product was digested with *Sph*I and *Bam*HI and was ligated into the *Nsp*I and *Bam*HI sites in pTA1797 to create the intermediate plasmid, pTA1850. pTA1850 was digested with *Nde*I and *Bam*HI and ligated into the *Nde*I and *Bam*HI sites in pTA1661 to create the *strepII-strepII-hel308* gene replacement construct, pTA1857.

An N-terminally *strepII-strepII-tagged hel308* was introduced into H2131 using the gene replacement construct pTA1857 to generate the strain H3367. The strain was verified via Southern blot (Figure 4.22).

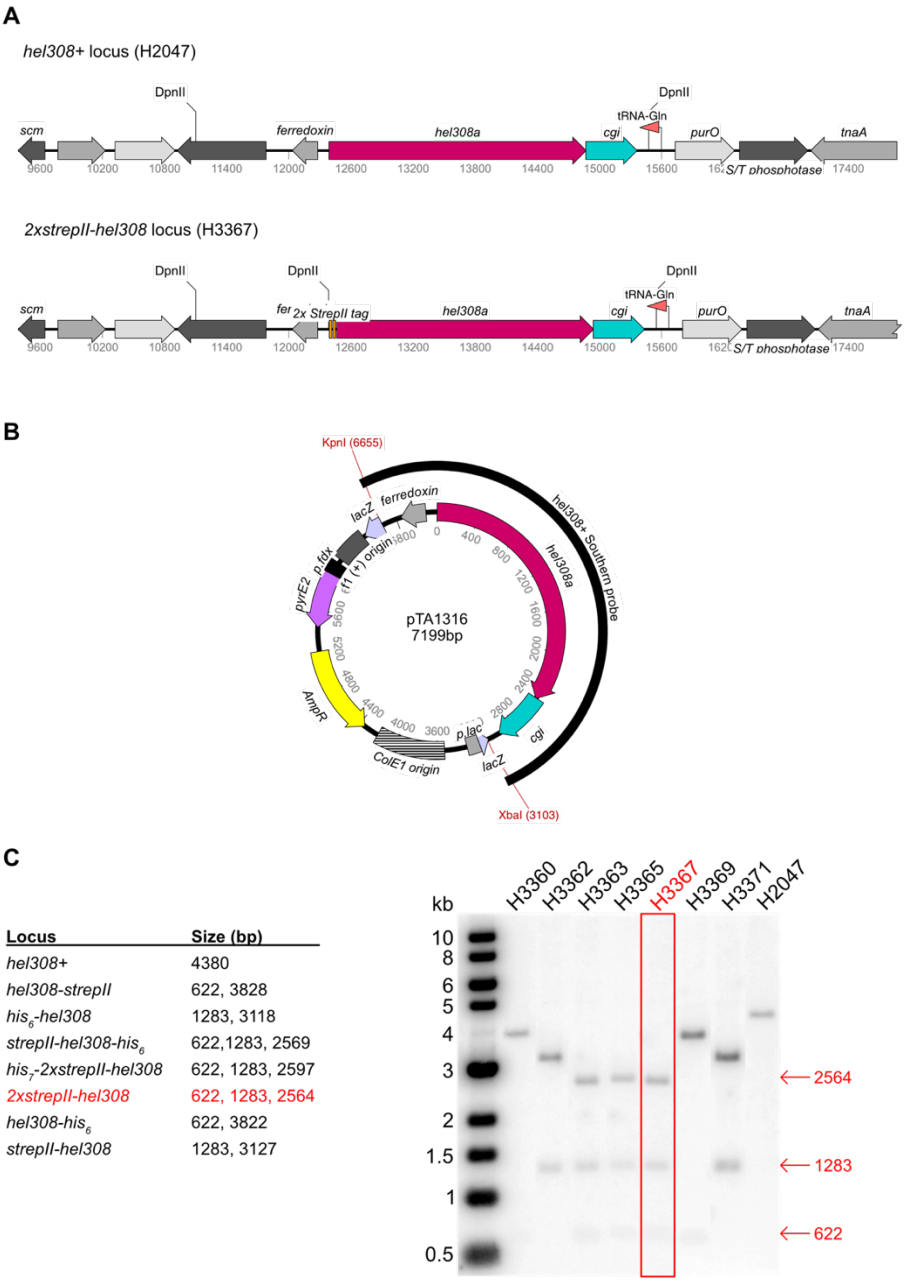


Figure 4.22 **Gene replacement of Δ*hel308::trpA*+ with *strepII-strepII-hel308* in H2131 (Δ*pyrE2, Nph-pitA*, Δ*mrr, cdc48d-ct*, Δ*trpA*, Δ*hel308::trpA*+).** A) Gene locus for *hel308*+ and *strepII-strepII-hel308*. The *Dpn*II sites used to digest genomic DNA are shown. B) A 3647bp *hel308*+ probe for southern hybridisation was made by digesting pTA1316 with *Kpn*I and *Xba*I. C) Predicted locus size and Southern blot, highlighted and boxed in red, showing H3367 is *strepII-strepII-hel308*+.

pTA1859 (*hel308-his$_6$* gene replacement construct)

The *strepII-hel308-his$_6$* gene replacement construct, pTA1855, was digested with the restriction enzymes *Kpn*I and *Nsi*I to remove the 1595 bp fragment containing the *strepII-tagged* 5'-portion of *hel308.* pTA1316 was also digested with the restriction enzymes *Kpn*I and *Nsi*I to produce a 1565 bp fragment containing the untagged 5'-portion of *hel308*. The untagged *5'hel308* fragment was ligated into the pTA1855 vector to generate the *hel308-his$_6$* gene replacement construct, pTA1859 (Figure 4.23).

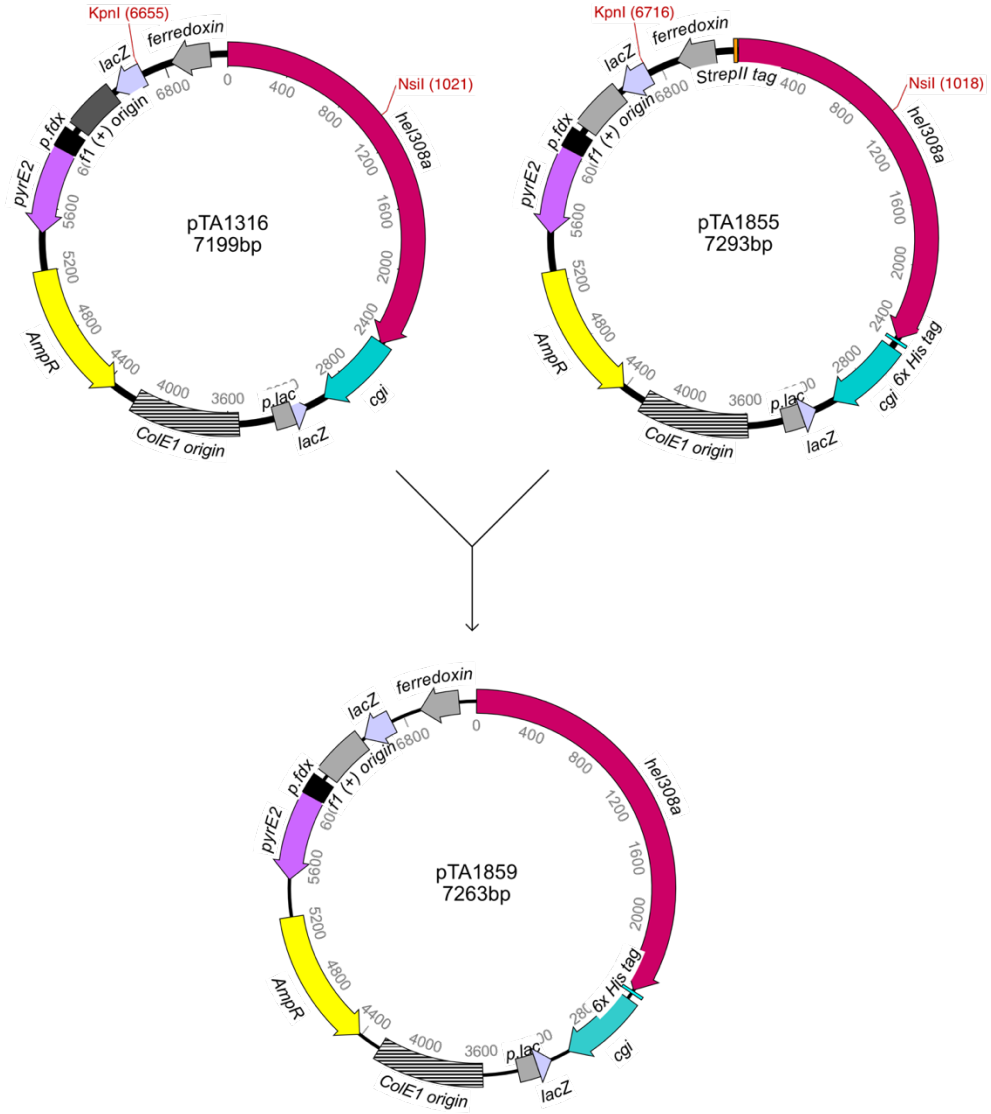


Figure 4.23: **Construction of C-terminally *his$_6$-tagged hel308* gene replacement construct, pTA1859.** pTA1316 was digested with *Kpn*I and *Nsi*I to produce a 1565 bp fragment containing the untagged 5'-portion of *hel308*, which was ligated into the *Kpn*I and *Nsi*I sites of pTA1855 to generate the *hel308-his$_6$* gene replacement construct, pTA1859.

A C-terminally *his$_6$-tagged hel308* was introduced into H2131 using the gene replacement construct pTA1859 to generate the strain H3369. The strain was verified via Southern blot (Figure 4.24).

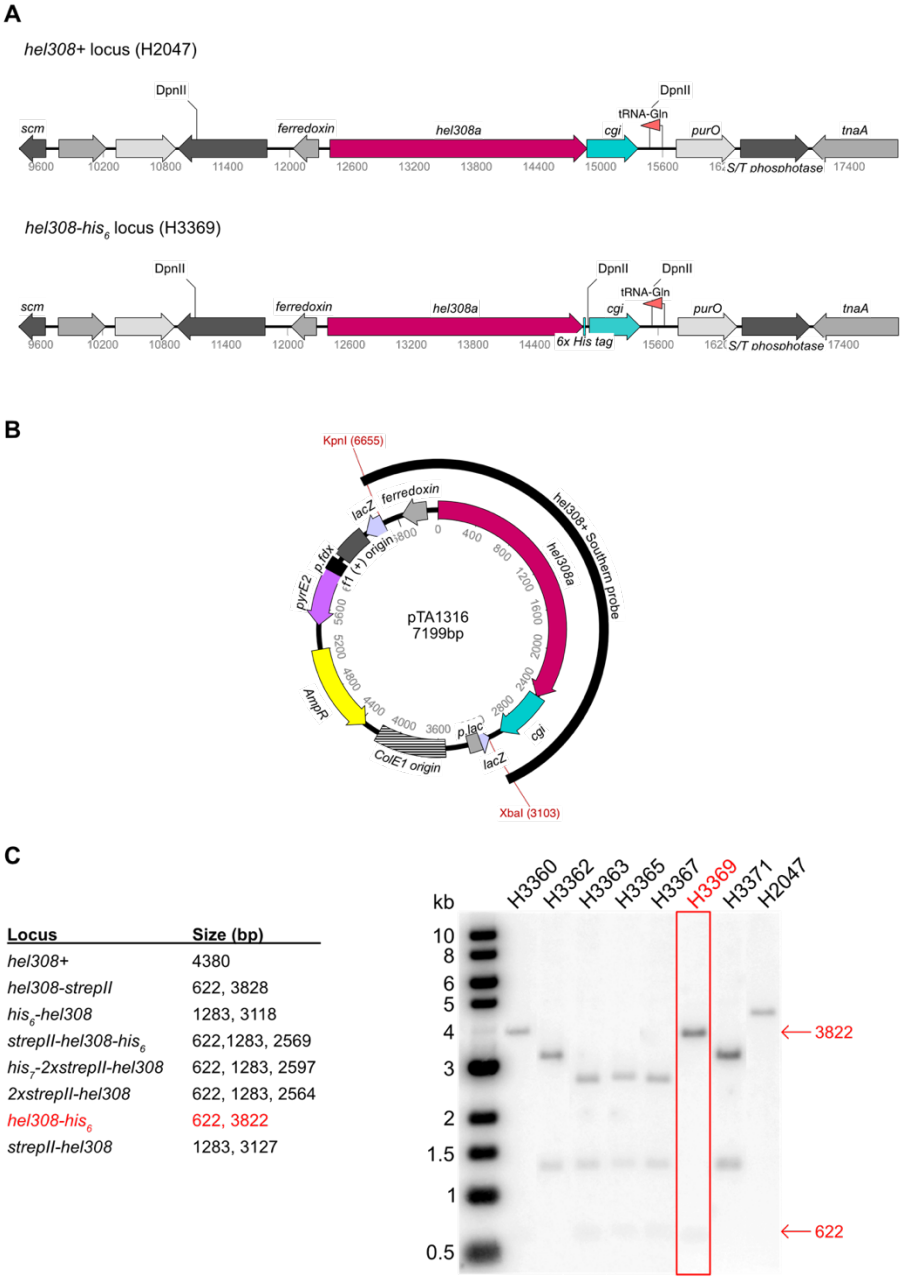


Figure 4.24 **Gene replacement of Δ*hel308::trpA*+ with *hel308- his$_6$* in H2131 (Δ*pyrE2, Nph-pitA,* Δ*mrr, cdc48d-ct,* Δ*trpA,* Δ*hel308::trpA*+).** A) Gene locus for *hel308*+ and *hel308-his$_6$*. The *Dpn*II sites used to digest genomic DNA are shown. B) A 3647bp *hel308*+ probe for southern hybridisation was made by digesting pTA1316 with *Kpn*I and *Xba*I. C) Predicted locus size and Southern blot, highlighted and boxed in red, showing H3369 is *hel308-his$_6$*+.

pTA1860 (*strepII-hel308* gene replacement construct)

The *strepII-hel308-his*$_6$ gene replacement construct, pTA1855, was digested with the restriction enzymes *Nsi*I and *Xba*I to remove the 2146 bp fragment containing the *his*$_6$*-tagged* 3'-portion of *hel308*. pTA1316 was also digested with the restriction enzymes *Nsi*I and *Xba*I to produce a 2082 bp fragment containing the untagged 3'-portion of *hel308*. The untagged *3'hel308* fragment was ligated into the pTA1855 vector to generate the *strepII-hel308* gene replacement construct, pTA1860 (Figure 4.25).

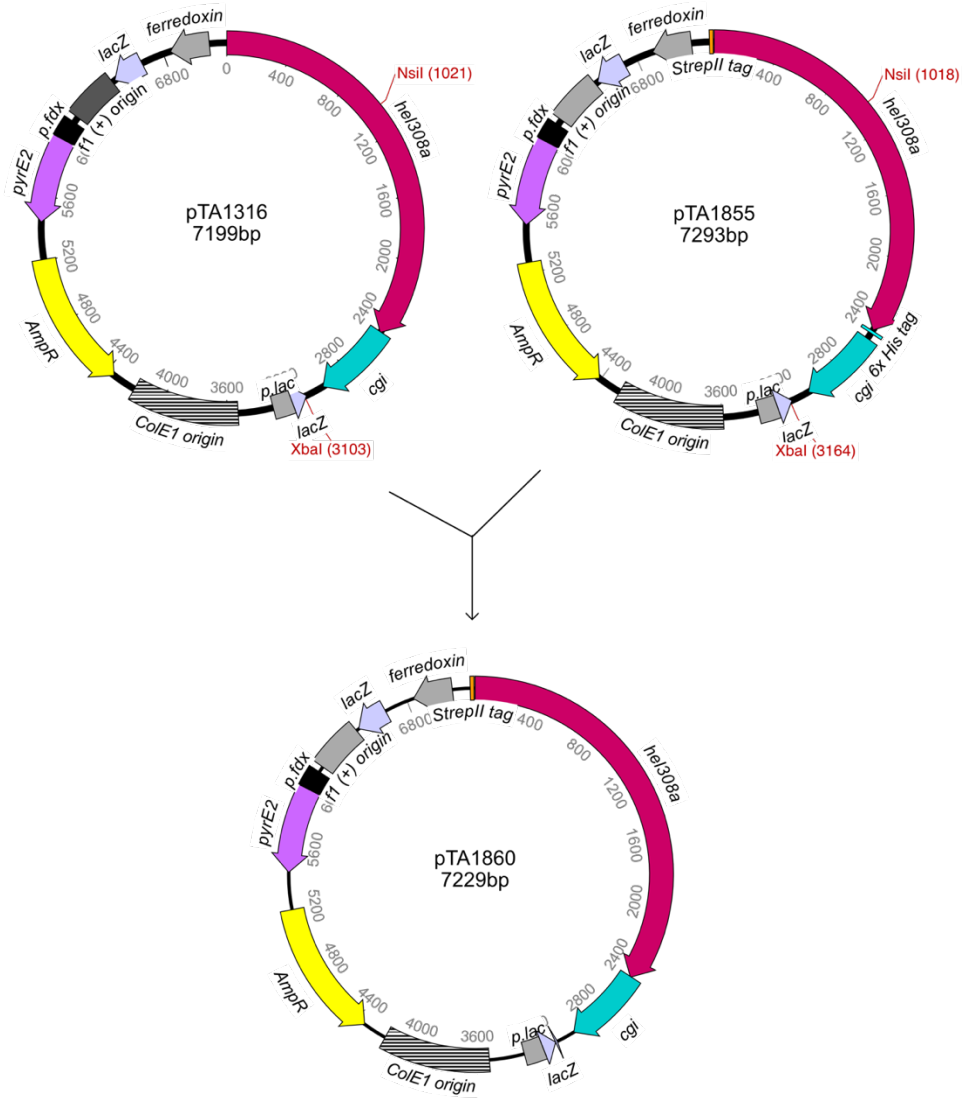


Figure 4.25: **Construction of N-terminally *strepII-tagged hel308* gene replacement construct, pTA1860.** pTA1316 was digested with *Nsi*I and *Xba*I to produce a 2082 bp fragment containing the untagged 3'-portion of *hel308*, which was ligated into the *Nsi*I and *Xba*I sites of pTA1855 to generate the *strepII-hel308* gene replacement construct, pTA1860.

An N-terminally *strepII-tagged hel308* was introduced into H2131 using the gene replacement construct pTA1860 to generate the strain H3371. The strain was verified via Southern blot (Figure 4.26).

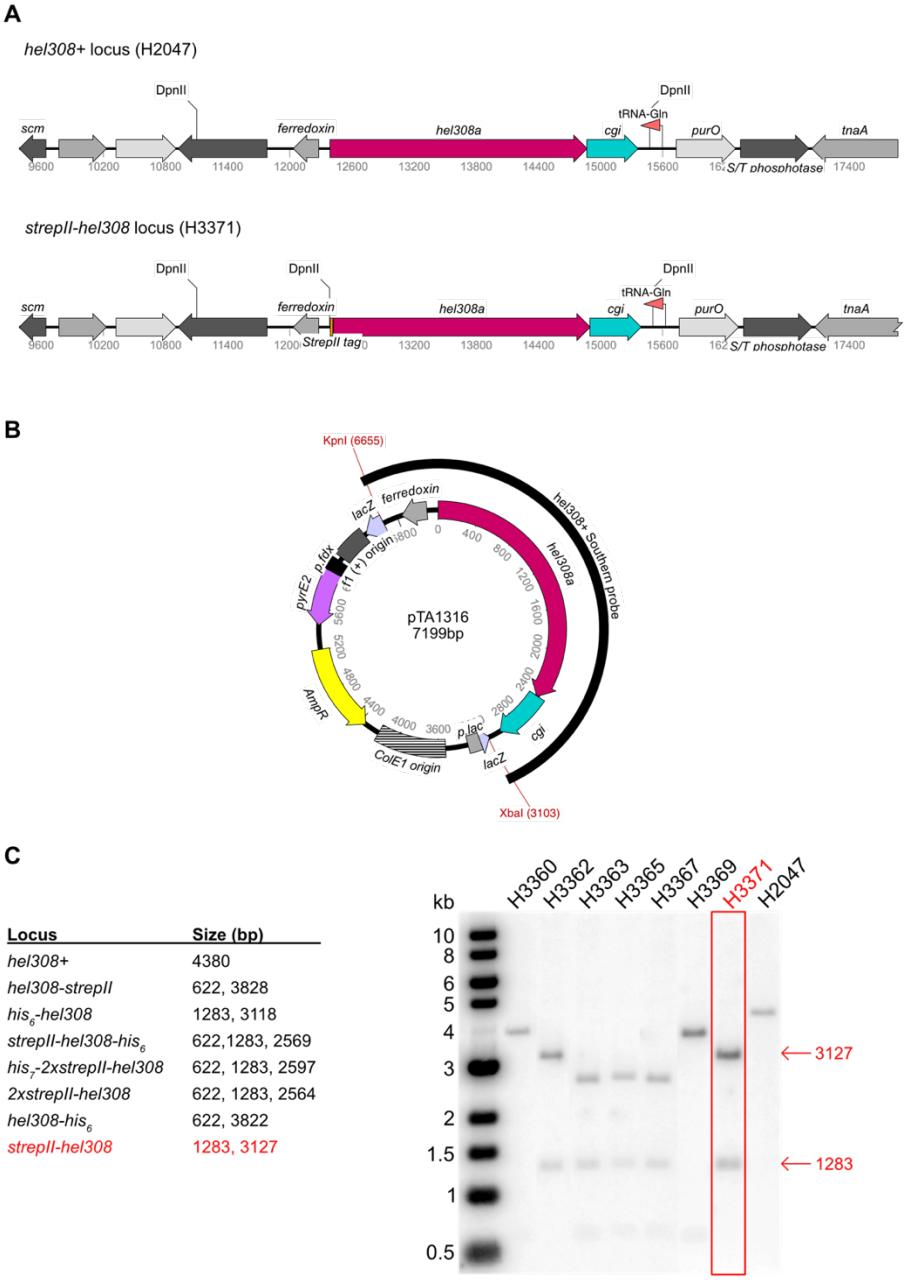


Figure 4.26 **Gene replacement of Δ*hel308::trpA+* with *strepII-hel308* in H2131 (Δ*pyrE2, Nph-pitA, Δmrr, cdc48d-ct, ΔtrpA, Δhel308::trpA+*).** A) Gene locus for *hel308+* and *strepII-hel308*. The *Dpn*II sites used to digest genomic DNA are shown. B) A 3647bp *hel308+* probe for southern hybridisation was made by digesting pTA1316 with *Kpn*I and *Xba*I. C) Predicted locus size and Southern blot, highlighted and boxed in red, showing H3371 is *strepII-hel308+*.

4.4.2 Characterisation of Chromosomally-Tagged Expression Strains

Growth Rate

The standard growth rates, in liquid media, were quantified for the differently tagged Hel308 strains. Strains were grown over two consecutive overnights to form a vigorously growing culture. The exponential phase of cell growth was calculated using an Epoch2 Microplate Spectrophotometer (BioTek) by measuring the A_{600} of the culture at 15-minute intervals over a 72-hour time course. A growth curve was plotted and the generation time during exponential phase for each strain was calculated (Figure 4.27).

In this experiment, the wild-type strain H2047 had a generation time of 2.25h. The strains H3360 (*hel308-strepII*), H3363 (*strepII-hel308-his*$_6$), and H3367 (*strepII-strepII-hel308*) all had a generation time greater than 5h. H3367 had a generation time of 5.25h and both H3360 and H3363 had a generation time of 6.75h, though not as slow as H2131 (Δ*hel308::trpA+*) of 8.75h, the slow generation times still suggest that these tagging strategies interfere with Hel308 function, whether it be by preventing correct folding of the protein or in some other way. For this reason, these strains were deemed inappropriate to study Hel308 protein interacting partners and were no longer used in this study.

The strains H3362 (*his*$_6$*-hel308*), H3365 (*his*$_7$*-strepII-strepII-hel308*), H3369 (*hel308-his*$_6$), and H3371 (*strepII-hel308*) all had a generation time less than or equal to 3h, which is more similar to the wild-type strain, H2047, with a generation time of 2.25h. H3362, H3369, and H3371 all have marginally slower generation times compared with H2047, with generation times of 3h, 2.75h, and 2.5h respectively. In contrast, H3365 has a slightly faster generation time of 2h.

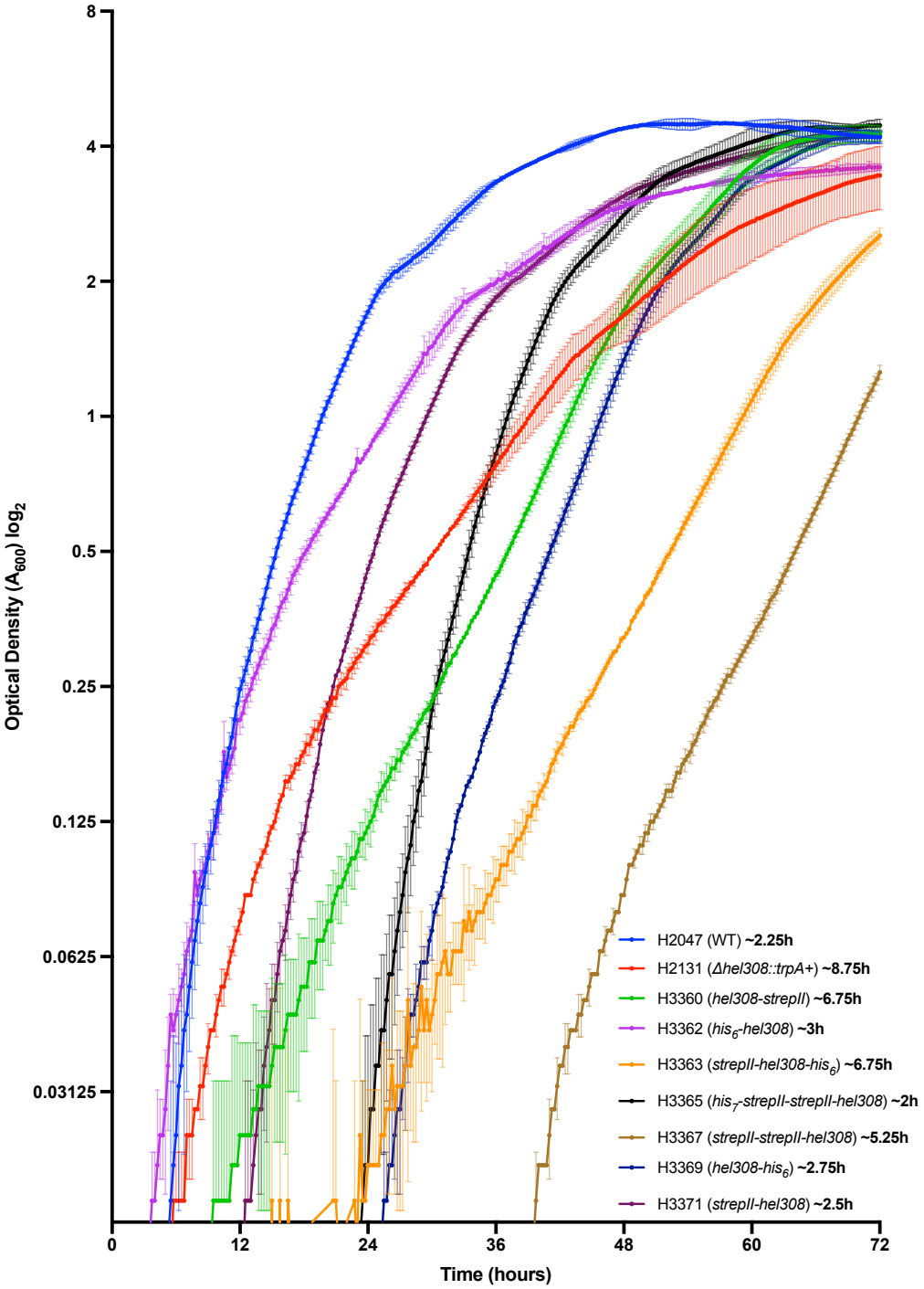


Figure 4.27: **Exponential growth rate of tagged Hel308 strains.** H2047 (*hel308+*), H2131 (Δ*hel308*), H3360 (*hel308-strepII*), H3362 (*his$_6$-hel308*), H3363 (*strepII-hel3080-his$_6$*), H3365 (*his$_7$-strepII-strepII-hel308*), H3367 (*strepII-strepII-hel308*), H3369 (*hel308-his$_6$*), and H3371(*strepII-hel308*) were grown in liquid culture for two consecutive overnights. The strains were then diluted and grown for 72 hours. Growth was measured by A_{600}. Generation time is indicated next to each strain. All strains were incubated on the same 96-well plate and measured simultaneously using an Epoch2 Microplate Spectrophotometer (BioTek). Data points are a mean of two technical repeats within two biological repeats. Standard deviation is shown.

Survival Following Treatment with DNA-damaging Agents

To determine which of the strains H3362 (*his$_6$-hel308*), H3365 (*his$_7$-strepII-strepII-hel308*), H3369 (*hel308-his$_6$*), and H3371 (*strepII-hel308*) behaved most like the wildtype strain, suggesting the tagged Hel308 is functioning correctly, these strains were characterised further by analyses of the survival frequency following treatment with UV irradiation or MMC (Figure 4.28).

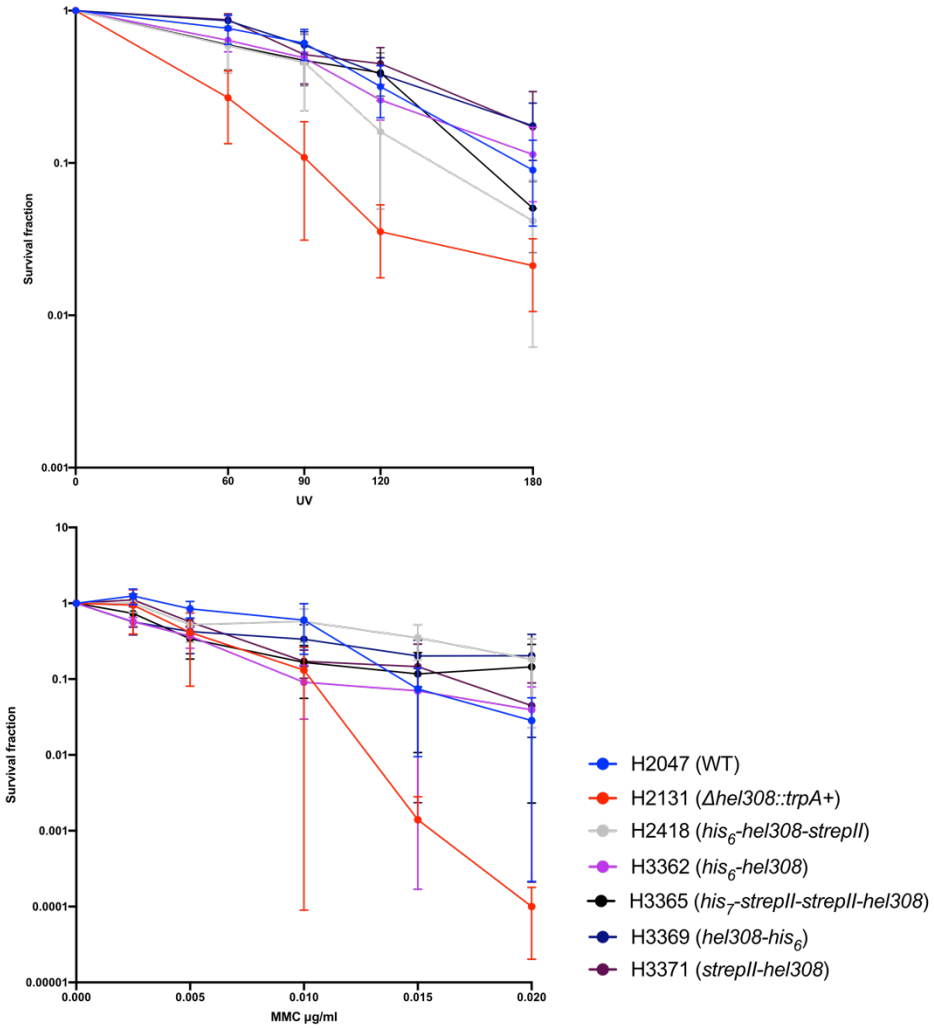


Figure 4.28: **Survival frequency of tagged Hel308 strains following treatment with DNA damaging agents.** No observed difference in survival fraction of any tagged hel308 strain compared to wildtype H2047 (*hel308+*) after UV irradiation or treatment with MMC. Survival fraction is calculated relative to un-treated control. Each data point is generated as the mean of at least 3 independent trials. Standard deviation is shown.

None of the four strains with near-wild type growth rates, H3362 (*his$_6$-hel308*), H3365 (*his$_7$-strepII-strepII-hel308*), H3369 (*hel308-his$_6$*), and H3371 (*strepII-hel308*), were sensitive to either of the DNA damaging agents, with survival fractions identical to H2047 in both assays.

As all of the four strains behave like wildtype, the decision as to which strain to use to study the protein interacting partners of Hel308 came down to which tag would result in the cleanest purification of Hel308 and would also be the most versatile. Therefore, H3365 (*his$_7$-strepII-strepII-hel308*) was chosen as the tag allowed purification via both nickel-affinity chromatography and application to a StrepTactin column.

4.4.3 Protein purification of natively expressed proteins

Metal affinity Chromatography

Proteins containing a His$_x$-tag were purified from cell lysate using gravity columns packed with Ni^{2+} charged Sepharose beads. Ni^{2+} charged beads are preferential to Co^{2+} charged beads, due to a higher protein yield in conditions of >1 M NaCl (Allers *et al.*, 2010). Increasing concentrations of imidazole, 100-500 mM, were used to elute the tagged protein. If subsequent purification via StrepTactin chromatography was to be carried out, the purified protein sample was first diafiltrated using appropriate molecular weight cut-off (MWCO) Vivaspin 20 ultrafiltration spin columns to remove residual imidazole and to concentrate the protein sample.

StrepTactin Chromatography

Proteins were purified from the cell lysate using gravity columns packed with StrepTactin Sepharose. Protein was eluted using 5 mM D-desthiobiotin. Protein elutions were precipitated using trichloroacetic acid (TCA) to assist visualisation of protein bands on SDS-PAGE gel and to ensure a sufficient concentration of proteins for mass spectrometry.

Hel308 purification strategy

In order to determine the most efficient method of purification, the natively-expressed N-terminally *his$_7$-strepII-strepII-tagged hel308* from the strain H3365 was purified via metal affinity chromatography only, StrepTactin chromatography only, and metal affinity chromatography followed by StrepTactin chromatography (Figure 4.29).

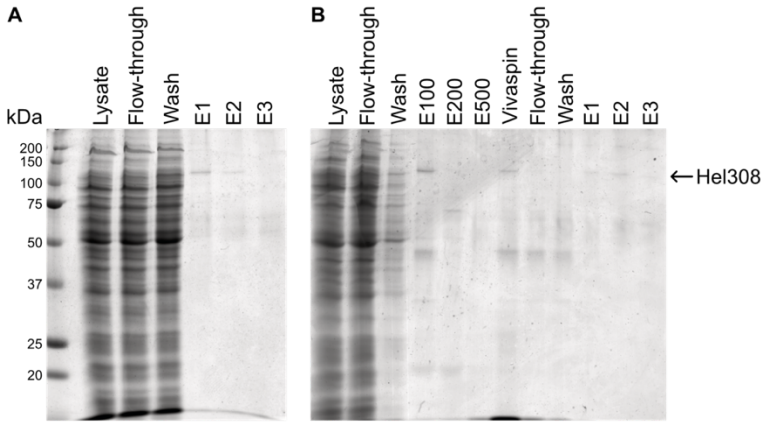


Figure 4.29: **Purification of N-terminally *his$_7$-strepII-strepII* tagged Hel308 via A)** StrepTactin affinity chromatography only, using a 3-step elution with 0.8 ml, 1.4 ml, and 0.8 ml of elution buffer containing 5 mM D-desthiobiotin (E refers to elution number). **B)** nickel affinity chromatography only (Lysate-E500) using a step elution with buffer containing imidazole (E refers to imidazole concentration in mM), and nickel followed by StrepTactin affinity chromatography (Vivaspin-E3) using a 3 step elution with 0.4 ml, 0.7 ml, and 0.4 ml of elution buffer containing 5 mM D-desthiobiotin. A His_x tagged protein (~100 kDa) was detected via western blot of (A) using a specific anti-His_6 antibody (blot not shown). This is presumed to be His_7-StrepII-StrepII-tagged Hel308 (indicated).

As anticipated, purification of *his$_7$-strepII-strepII* tagged *hel308* via nickel affinity chromatography only yields many histidine-rich protein contaminants in the elutions E100-E500 (Figure 4.29B, all bands that are not indicated as Hel308). Application to a StrepTactin column, following nickel affinity chromatography, removes a large number of these contaminating bands (Figure 4.29B, E1-E3); however, recovery of Hel308 is very poor. Purification of *his$_7$-strepII-strepII* tagged *hel308* via StrepTactin affinity chromatography only has few contaminants in elutions E1-E3, the same contaminants are also present in elutions E1-E3 of the sequential purification strategy (Figure 4.29A, all bands that are not indicated as Hel308) and a greater recovery of Hel308. Therefore, purification of His_7-StrepII-StrepII-Hel308 is most successful via StrepTactin chromatography alone.

4.4.4 Protein:Protein Interactions *in vivo*

To understand protein function and cellular processes, the study of protein:protein interactions *in vivo* is an essential tool. This has been carried out in a variety of archaeal species.

In vivo immunoprecipitation analyses has been carried out in *P. furiosus*. The proliferating cell nuclear antigen (PCNA) was found to co-precipitate with Hjm, the Hel308 homologue, via an interaction in the C-terminal domain of Hjm.

PCNA is the DNA sliding clamp which prevents DNA dissociation by means of anchoring the DNA polymerase to the template DNA. Moreover, PCNA was shown to stimulate the helicase activity of Hjm at fork-structured DNA (Fujikane, Shinagawa and Ishino, 2006).

Gel filtration, affinity pulldown, and yeast two-hybrid analyses of the *Sulfolobus tokodaii* Hel308 homologue, Hjm, showed a physical interaction with Hjc *in vitro* (Li *et al.*, 2008). The Holliday junction resolvase Hjc, binds specifically to Holliday junctions and symmetrically cleaves two opposing strands generating two recombinant duplexes, which is thought to restart stalled replication forks (Lestini, Duan and Allers, 2010)

In vitro analysis of immobilised MBP-tagged Hel308 from *M. thermautotrophicus* revealed an interaction with purified His_6-tagged replication protein A (RPA) after incubation overnight at room temperature (Woodman, Brammer and Bolt, 2011). RPA is the archaeal and eukaryotic single stranded binding (SSB) protein that binds unwound single stranded DNA to protect it from degradation and prevent secondary structure formation.

Outside of the Archaea, HelQ is the mammalian Hel308 homologue and has been shown to associate *in vivo* with the RAD51 paralogues RAD51B/C/D and XRCC2, as well as with the DNA damage responsive kinase ATR (Takata *et al.*, 2013).

The study of *in vivo* protein: protein interactions in *H. volcanii* was therefore performed using natively expressed Hel308, in order to gain further insight into the cellular role of Hel308.

Preliminary Screening for Interacting partners of Hel308

Hel308 was purified from the strain H3365 containing a chromosomal N-terminally *his_7-strepII-strepII-tagged hel308*. Strains deleted for *hel308* have a marked survival deficiency following treatment with the DNA crosslinking agent MMC; therefore, Hel308 is thought to act in the repair of DNA crosslinks (Adelman *et al.*, 2013; Takata *et al.*, 2013). First, an appropriate dose of MMC for acute treatment of cells was determined. The H3365 strain was incubated with various concentrations of MMC for 3 hours before spotting on non-selective media. Once grown, colonies were counted and survival fractions were calculated relative to an untreated control (Figure 4.30).

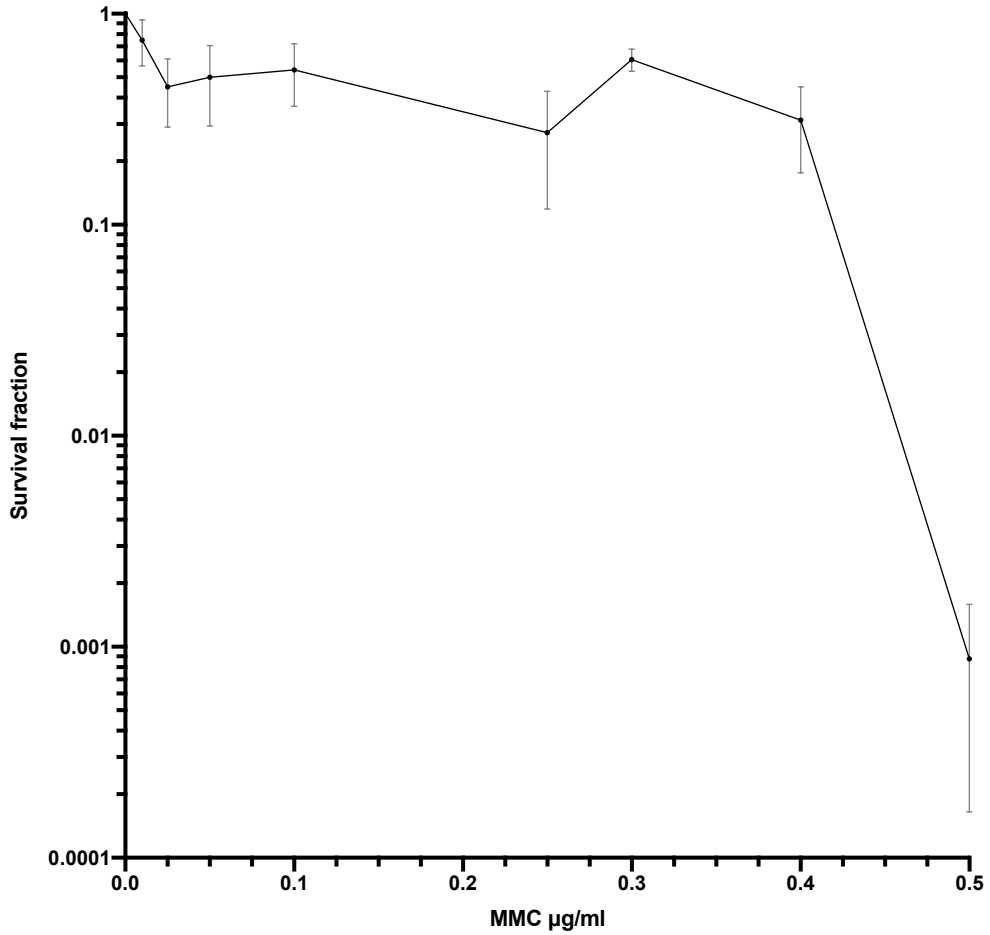


Figure 4.30: **Survival fraction of H3365 (*his7-strepII-strepII-hel308*) following acute treatment with MMC.** The H3365 strain was incubated with MMC for 3 hours in liquid culture before spotting on non-selective media. There is ~50% cell survival up to a concentration of 0.3 µg/ml MMC before a sharp decline in cell survival to <0.1% at 0.5 µg/ml MMC. Survival fraction is calculated relative to un-treated control. Each data point is generated as the mean of at least 5 independent trials. Standard deviation is shown.

The H3365 strain was then incubated with 0 (dH_2O only), 0.3, or 0.5 µg/ml MMC for 3 h prior to cell harvesting for StrepTactin chromatography only, giving cell survival fraction of 1, 0.6, and 0.001 respectively. These doses were chosen to investigate the difference in response to no DNA damage stress, low-level DNA damage stress, and extreme DNA damage stress. The result of the co-purification assay and TCA precipitation is shown below (Figure 4.31 A and B). Western blotting was used to verify His_7-StrepII-StrepII-Hel308 purification independently of the Streptactin column by probing for the His-tag using a specific anti-His_6 antibody (Figure 4.31 C, see Material and Methods for details).

A His_x tagged protein (~100 kDa) was detected via western blot using a specific anti-His_6 antibody (Figure 4.31C). This is presumed to be His_7-StrepII-StrepII-tagged Hel308. Bands (boxed) a, b, and e correspond to the bands presumed to be Hel308 (Figure 4.31B). These bands were cut out and formally identified via mass spectrometry. Boxed band d corresponds to a ~140 kDa protein (Figure 4.31B), which was detected via western blot using a specific anti-His_6 antibody, but only in the first elution of the *his_7-strepII-strepII-hel308* strain after incubation with 0.5 µg/ml MMC (Figure 4.31C). Band d was cut out and formally identified via mass spectrometry. The recovery of the protein in bands (boxed) c and f (~50 kDa) appeared to increase with regards to the recovery of Hel308 (Figure 4.31B). These bands were cut out and formally identified via mass spectrometry.

The complete list of identified proteins can be found in Appendix 1. Identified proteins, which either have a role either in DNA repair or replication, or are yet to be characterised, were included in the summary table (Table 4.4). These data are very preliminary due to the lack of appropriate controls and repeats (N=1).

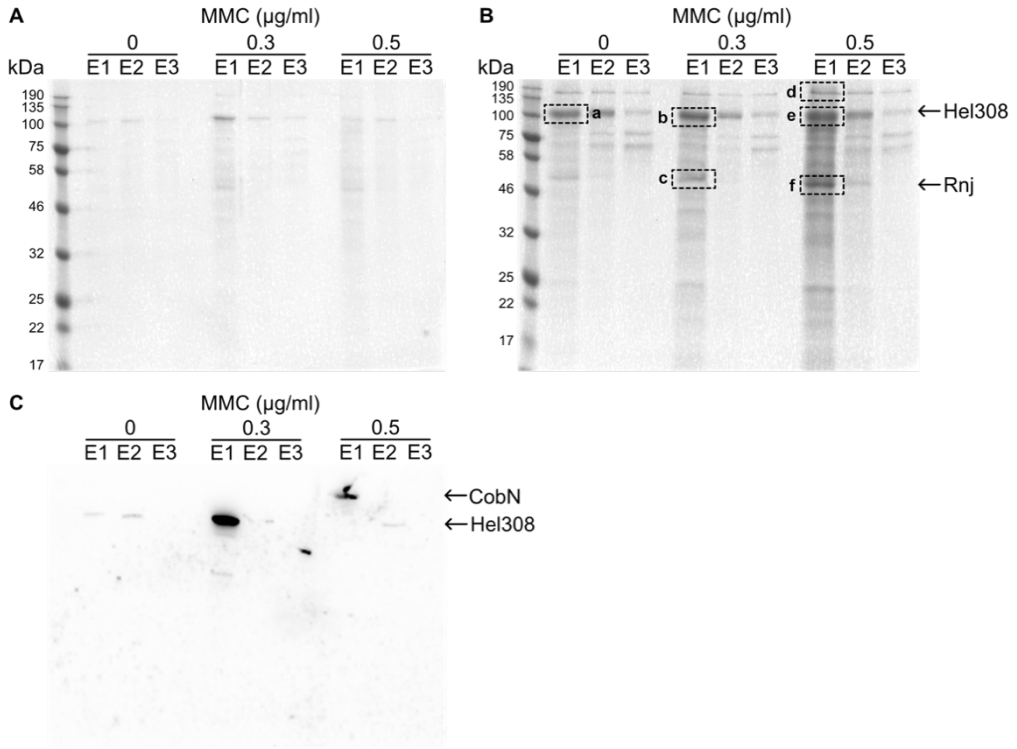


Figure 4.31: **Purification of N-terminally *his$_7$-strepII-strepII* tagged Hel308 following acute treatment with MMC.** A) The H3365 strain was incubated with 0 (dH_2O only), 0.3, or 0.5 µg/ml MMC for 3 hours prior to cell harvesting for StrepTactin chromatography using a 3 step elution with 0.8 ml, 1.4 ml, and 0.8 ml of elution buffer containing 5 mM D-desthiobiotin (E refers to elution number). B) Elutions were precipitated using TCA to aid visualisation. C) A His_x tagged protein (~100 kDa) was detected via western blot using a specific anti-His_6 antibody. This is presumed to be His_7-StrepII-StrepII-tagged Hel308 (indicated). A ~140 kDa protein was detected only in the first elution of H3365 incubated with 0.5 µg/ml MMC. Boxed bands were identified via mass spectrometry.

Table 4.4: **Summary of key proteins identified by Mass spectrometry.**

Band	**a**	**b**	**c**	**d**	**e**	**f**
~MW (kDa)	100	100	50	150	100	50
Proteins identified	Hel308a	Hel308a	Hel308a	Hel308a	Hel308a	Hel308a
	Cdc48a	RecJ4	CetZ1	CobN	RecJ4	RnJ
	RecJ4	Cdc48a	FtsZ1	NrdA	Cdc48a	FtsZ1
	Top6B	Top6B	RnJ	RpoA1	Top6B	RpoA2
	TopA	NrdA	RpoA2	Cdc48a	TopA	Top6A
	NrdA	TopA	Top6A	Rad50	RpoA1	Cdc48d
	RpoA1	RpoA1	Cdc48d	RecJ4	MutLa/b	Orc8
	GyrA	MutLa/b	RadA	Top6B	Lhr2	Orc5
	MutLa/b	GyrA	Orc8	UvrA	GyrA	Cdc48a
		UvrA		Cdc48b	RecJ3	RadA
				GyrA	RpoB2	RecJ4
				TopA	UvrA	PriS
				RecJ3	CobN	Orc15
				RpoB2	RadA	Cas7
				RnJ	MCM	
				RadA		

See Appendix 1 for complete table of all proteins identified by Mass spectrometry.

Cdc48 proteins

The cell division control protein 48 (CDC48) family are AAA+ class ATPases. Cdc48 is a central component in many ubiquitin-mediated pathways (Baek *et al.*, 2013). There are four Cdc48 homologues in *H. volcanii*, designated *cdc48a, cdc48b, cdc48c,* and *cdc48d.* Cdc48d was identified as a histidine-rich protein and a common contaminant of nickel affinity chromatography; therefore, a truncated version of Cdc48d is often required (Allers *et al.*, 2010). Whilst a truncated Cdc48d is tolerated by *H. volcanii*, a full deletion is not viable. Cdc48a is found in the RadB neighbourhood and is thought to play a key role in regulating the restart of replication by targeting the sampylated replisome for destruction by the proteasome (Forouzan *et al.*, 2012) and allowing fork restart via homologous recombination (Figure 4.32). Cdc48 has been shown to co-purify with the modifier protein SAMP1. Deletion of *cdc48* results in slow growth but increased resistance to DNA damage. This phenotype is also seen when wild-type cells are treated with DBeQ, a specific reversible inhibitor of Cdc48, whereas strains deleted for *cdc48* are not affected by DBeQ (Miranda *et al.*, 2011). Based on previous studies with Δ*mre11-rad50* mutants (Cao, Hepowit and Maupin-Furlow, 2015), it is proposed that increased resistance to DNA damage is due to uncontrolled recombination in the absence of Cdc48, suggesting that sampylation is used by Cdc48 to target proteins for remodelling and thereby modulate the activity of recombination.

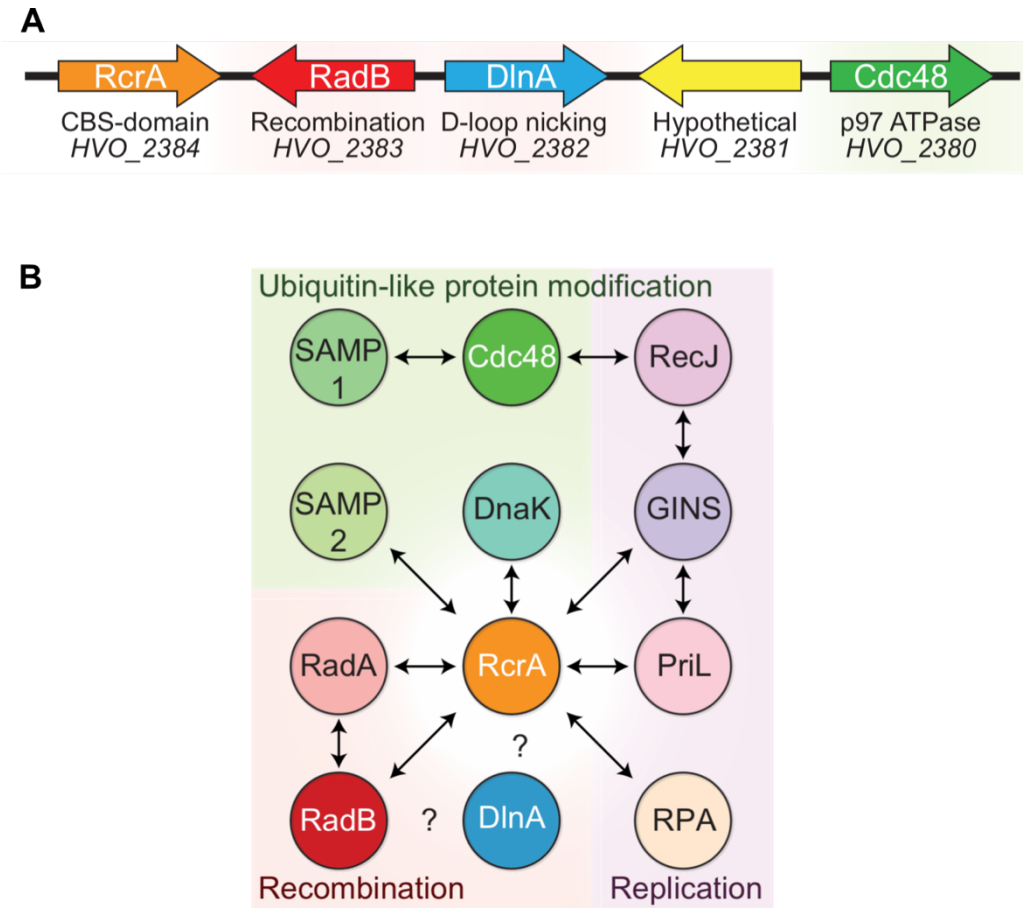


Figure 4.32: **The RadB gene neighbourhood.** A) The RadB gene neighbourhood *HVO_2380*-2384 is highly conserved throughout archaea. B) The proposed network of DNA replication and recombination proteins associated with ubiquitin-like SAMPylation. Figure adapted from (Wardell, 2013. Personal correspondence, Thorsten Allers).

Topoisomerases

Type I

TopA (HVO_0681)

TopA is a type IA topoisomerase, which releases supercoiling and torsional tension by cleaving one strand of DNA and passing the cleaved strand around the unbroken strand (Changela, DiGate and Mondragón, 2001).

Type II

Type II topoisomerases operate by cleaving both strands of DNA and passing a second segment of DNA through the break in an ATP-dependent manner, this is known as the strand-passage mechanism (Stuchinskaya *et al.*, 2009).

Top6A and Top6B (HVO_1570 and HVO_1571)

Topo VI is a type IIB topoisomerase, which is responsible for supporting archaeal and plant viability, and has a distant homologue Spo11 that acts in meiotic recombination. Topo VI is a heterotetramer comprising two different polypeptide subunits, Top6A and Top6B. Top6B recognises DNA crossovers generated by supercoiled DNA and stimulates subunit dimerisation via binding of ATP. The Top6B homodimer then extensively bends the DNA to support duplex cleaving by Top6A (Wendorff and Berger, 2018).

GyrA (HVO_1573)

DNA gyrase is a topoisomerase II found primarily in bacteria and some archaea. DNA gyrase is a heterotetramer comprising two different polypeptide subunits, GyrA and GyrB. GyrA is thought to bind DNA via the C-terminal domain in order to mediate a positive super-helical wrap about the protein (Corbett, Shultzaberger and Berger, 2004). Interestingly, the gyrase subunits are adjacent to the Topo VI subunits on the *H. volcanii* main chromosome, and are divergently transcribed (Figure 4.33).

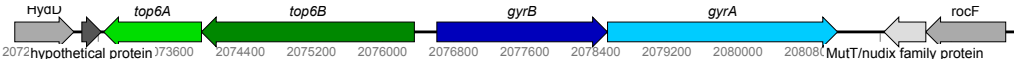


Figure 4.33: **Divergent expression of the type II topoisomerases Topo6 and Gyrase.** The subunits of Topo6, *top6A* (HVO_1570) and *top6B* (HVO_1571), are shown in green. The subunits of Gyrase, *gyrA* (HVO_1573) and *gyrB* (HVO_1572), are shown in blue.

RpoA1 and RpoA2 (HVO_0349 and HVO_0350)

In archaea, RpoA is the largest subunit of the DNA-directed RNA polymerase. RpoA comprises two polypeptides, A' and A" subunits, which are encoded by separate genes within an operon (Langer *et al.*, 1995; Jun *et al.*, 2011).

RecJ3 and RecJ4 (HVO_1018 and HVO_2889)

Bioinformatic and structural studies have revealed that human Cdc45 and archaeal RecJ are evolutionarily related to bacterial RecJ (Makarova, Koonin and Kelman, 2012; Pellegrini, 2017). The RecJ family of proteins are typically 5'-3' exonucleases and have been implicated in several DNA repair pathways including homologous recombination, mismatch repair, and base excision repair. It is also thought that RecJ is important in the rescue of stalled replication forks (Nagata *et al.*, 2017). *Haloferax volcanii* has four RecJ homologues, designated

RecJ1 (HVO_0073), RecJ2 (HVO_1147), RecJ3 (HVO_1018), and RecJ4 (HVO_ 2889).

MCM (HVO_0220)

MCM is the replicative helicase, which forms part of the CMG complex along with Cdc45 and GINS (Makarova, Koonin and Kelman, 2012). Archaeal replication resembles its eukaryotic counterpart, for which the CMG complex is essential. Whilst *H. volcanii* MCM and GINS homologues are known, a Cdc45 homologue has yet to be identified (See Chapter 6 for further details).

MutLa/b (HVO_1939/0551)

Due to limited peptide coverage, mass spectrometry analysis was unable to distinguish between the MutL proteins. MutL in an ATP dependent endonuclease responsible for nicking the DNA backbone at a site of replication error in order to excise the miss-incorporated dNTPs during the mismatch repair pathway (Pluciennik *et al.*, 2010).

UvrA (HVO_0393)

Homologues of the bacterial UvrABC system are found in *H. volcanii* and are thought to act in nucleotide excision repair (Crowley *et al.*, 2006; Lestini, Duan and Allers, 2010). DNA damage distorts the DNA helix, which is detected by the UvrA2B heterotrimer. Interaction of the UvrA2B with the lesion causes local unwinding of the DNA duplex (Zou and Van Houten, 1999).

CetZ1 and FtsZ1 (HVO_2204 and HVO_0717)

CetZ1 and FtsZ1 are cell division GTPase proteins belonging to the FtsZ/tubulin superfamily. These genes are abundant in the haloarchaea. FtsZ1 forms a ring at the cell mid-region, indicating the future site of cell division. Inactivation of FtsZ1 renders the cell unable to divide.

Conversely, inactivation of CetZ1 does not affect cell division. CetZ1 forms dynamic cytoskeletal structures in order to remodel the cell envelope from an irregular plate-shaped cell into a rod-shaped cell type that is essential for normal swimming motility (Duggin *et al.*, 2015).

Rnj (HVO_2724)

Ribonuclease J is a bacterial RNase J-like ribonuclease. Bacterial RNase J is a member of the β-CASP family of ribonucleases involved in mRNA processing

and degradation. Archaea have two distinct β-CASP orthologous groups, archaeal RNase J (aRNase J) and archaeal cleavage and polyadenylation specificity factor (aCPSF), which is more closely related to the eukaryotic CPSF-73 (Phung *et al.*, 2013).

RadA (HVO_0104)

Homologous recombination involves the exchange of nucleotide sequence between near-identical DNA sequences for which RadA, the archaeal recombinase, is essential. RadA forms a nucleoprotein filament on ssDNA with the assistance of RadB. The RadA nucleoprotein filament then binds dsDNA molecules in order to search for a region of homology (Wardell, 2013). Once found, RadA catalyses strand invasion and D-loop formation (Kil, Baitin and Lanzov, 2000) (Figure 3.4).

Orc Proteins

Orc5 (HVO_1725)

Orc5 is one of several Cdc6/Orc1 homologues present in *H. volcanii.* The main chromosome of the wild isolate, DS2, has three origins of replication, *oriC1-3*, each with a distinct associated Orc protein. Orc5 is associated with the *oriC2* replication origin (Norais *et al.*, 2007; Hawkins *et al.*, 2013).

Orc8 (HVO_C0057)

Similarly, Orc8 is one of several Cdc6/Orc1 homologues present in *H. volcanii.* However unlike Orc5, Orc8 is located on the mini-chromosome pHV1 and is not origin associated (Norais *et al.*, 2007).

Orc15 (HVO_1537)

Orc15 and Orc16, distant Cdc6/Orc1 homologues, have been implicated in glycerol metabolism. The *orc15* and *orc16* genes are encoded within distinct glycerol metabolic operons. These operons are highly conserved within halophilic archaea (Personal correspondence, Julie Maupin-Furlow).

CobN and NrdA (HVO_B0050 and HVO_2452)

Cobaltochelatase consists of three subunits CobN, CobS and CobT (Cameron *et al.*, 1991; Debussche *et al.*, 1992). Cobaltochelatases catalyse the insertion of Co^{2+} into sirohydrochlorin during the anaerobic biosynthesis of cobalamin (vitamin B_{12}) (Yin *et al.*, 2006).

NrdA is a Class II ribonucleotide reductase (RNR), responsible for the formation of deoxyribonucleotides from ribonucleotides. During this process, NrdA produces a 5'-deoxyadenosyl radical by homolytic cleavage of the C-Co bond in adenosylcobalamin (Hagmann, 1997; Stubbe and Riggs-Gelasco, 1998). The cleavage of this bond is not inhibited by hydroxyurea.

Lhr2 (HVO_1333)

Lhr2 is an SF2 helicase. The SF2 is one of the largest super family of helicases, members of which participate in almost all steps of DNA and RNA metabolism. Phylogenetic analysis of the distribution of SF2 helicase reveals that Lhr is the most abundant helicase in archaea. (Chamieh, Ibrahim and Kozah, 2015). Lhr is a 3′–5′ translocase able to unwind both RNA:DNA hybrids and DNA:DNA duplex (Fairman-Williams, Guenther and Jankowsky, 2010; Ordonez and Shuman, 2013). There are two Lhr proteins encoded in *H. volcanii*, Lhr2, which is counterintuitively found in the Lhr-1 clade, and Lhr1, which is similarly found in the Lhr-2 clade (Figure 4.34, adapted from Chamieh, Ibrahim and Kozah, 2015).

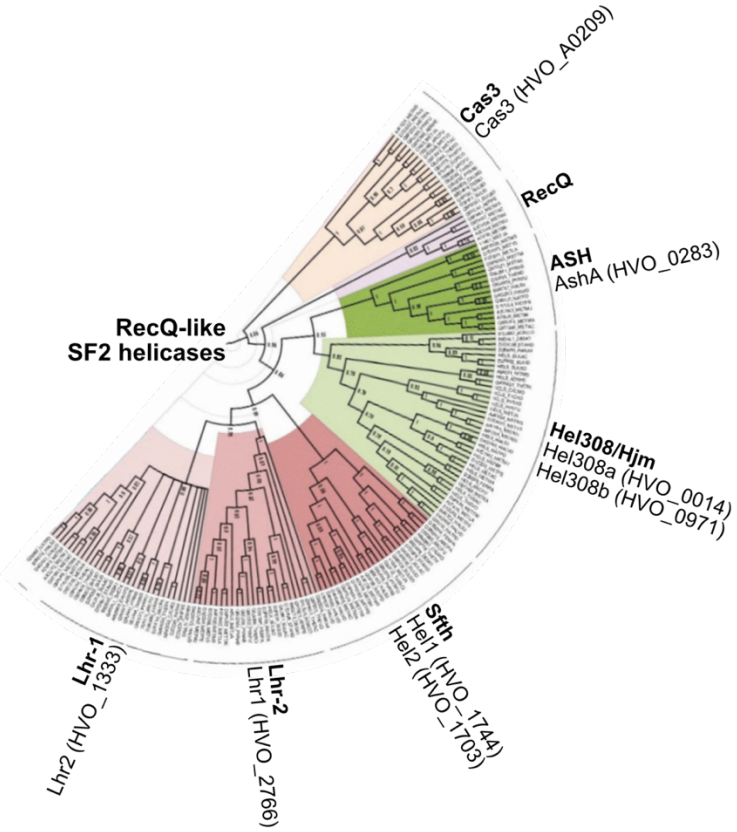


Figure 4.34: **Distribution of Archaeal RecQ-like SF2 helicases.** Phylogenetic distribution of the archaeal RecQ-like helicases. *Haloferax volcanii* genes are annotated next to the clade in which they are found. Figure adapted from Chamieh, Ibrahim and Kozah, 2015.

PriS (HVO_2697)

PriS is the small catalytic subunit of the eukaryotic-type DNA primase. DNA replication is crucial for cell survival in all organisms. DNA primases synthesise RNA primers on both the leading and lagging strands in order to initiate DNA synthesis (Le Breton *et al.*, 2007).

Cas7 (HVO_A0207)

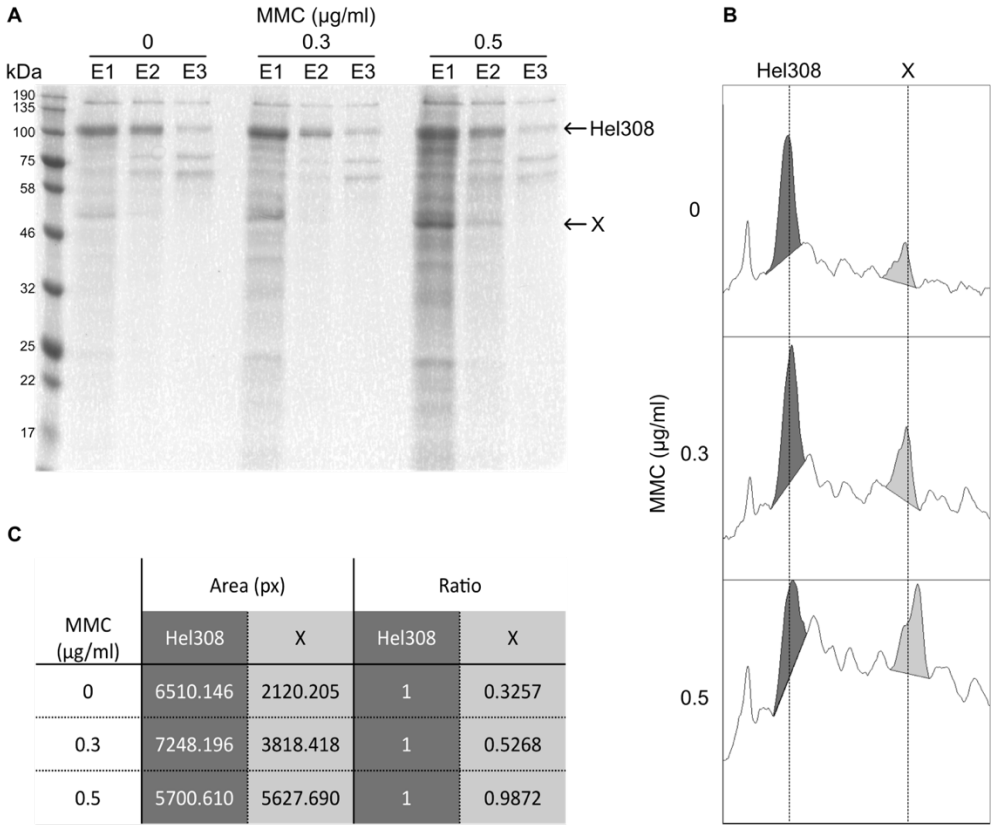
The presence of a CRISPR-Cas adaptive immune system is preserved throughout the archaea. *H. volcanii* possesses a single type I-B CRISPR-Cas system. Cas7 forms Cascade-like complex along with the Cas proteins 5 and 6, which is important for the stability and maintenance of the CRISPR-derived RNAs (crRNAs) (Brendel *et al.*, 2014; Maier, Dyall-Smith and Marchfelder, 2015; Maier *et al.*, 2019).

Rad50 (HVO_0854)

Rad50 associates with Mre11 in order to target and repair DNA double-strand breaks (DSBs). Mre11 and Rad50 proteins are found in all domains of life. Compared to wild-type cells, *mre11 rad50* mutants are more resistant to DNA damaging agents; however, they recover more slowly. Mre11-Rad50 appears to delay the repair of DSBs via homologous recombination, allowing microhomology-mediated end-joining to act as the primary repair pathway (Delmas *et al.*, 2009).

The possible significance of these interactions will be discussed in further detail in Section 4.5: Discussion.

Precipitation of the elutions indicated an increase in recovery of the Hel308 interacting protein RnJ (~50kDa) with increased concentrations of MMC, compared with the recovery of Hel308a. The intensity of the protein bands (peaks) relative to the background (baseline) and the area under the peak was analysed using ImageJ software. Normalising the data to the intensity of the Hel308 peak indicates that RnJ is upregulated with increasing concentrations of MMC (Figure 4.35).



C

MMC (µg/ml)	Area (px)		Ratio	
	Hel308	X	Hel308	X
0	6510.146	2120.205	1	0.3257
0.3	7248.196	3818.418	1	0.5268
0.5	5700.610	5627.690	1	0.9872

Figure 4.35: **Increase in recovery of the Hel308 interacting protein X with increase in concentration of MMC.** A) The H3365 strain was incubated with 0 (dH_2O), 0.3, or 0.5 µg/ml MMC for 3 hours prior to cell harvesting for StrepTactin chromatography using a 3 step elution with 0.8 ml, 1.4 ml, and 0.8 ml of elution buffer containing 5 mM D-desthiobiotin (E refers to elution number). Elutions were precipitated using TCA to aid visualisation. B) ImageJ was used to analyse the intensity of the protein bands (peaks) relative to the background (baseline) and calculate the area under the peak. C) Normalising the data to the intensity of the Hel308 peak indicates that protein X is upregulated with increasing concentrations of MMC.

4.4.5 Development of episomal overexpression plasmid constructs

At overexpressed levels, Hel308 is toxic to *H. volcanii* (Gamble-Milner, 2016). However, native expression does not yield sufficient concentrations of Hel308 for further biochemical assays. For this reason, episomal over-expression constructs incorporating the lower-activity (and therefore less leaky) *p.tnaM3* promoter and the improved tandem gene tag were constructed. In addition to wildtype *hel308*, an overexpression construct for the ATPase-null mutant, *hel308-D145N*, was created under the assumption that the Hel308-D145N protein would not be toxic to *H. volcanii*, since the point mutation in the Walker B motif renders the protein non-functional. For completion, the hyper-recombinant point mutant, *hel308-F316A*, was also analysed in this manner.

pTA2118 (*p.tnaM3-his$_7$-strepII-strepII*-tagged gene episomal over-expression construct)

The generic gene tagging construct, pTA1451, was digested with *Kpn*I and *Nde*I to liberate the fragment containing the *p.tnaM3* promoter. The promoter was ligated into the *Kpn*I and *Nde*I sites of pTA1771 to create the plasmid pTA2096. The *p.tnaM3-his$_7$-strepII-strepII*-tag CDS was digested from pTA2096 using *Bst*BI and *Bam*HI, and was ligated into the *Bst*BI and *Bam*HI of pTA1228 (see 4.1.2 Protein purification, Plasmid constructs) to create the generic *p.tnaM3-his$_7$-strepII-strepII*-tagged gene episomal over-expression construct, pTA2118 (Figure 4.36).

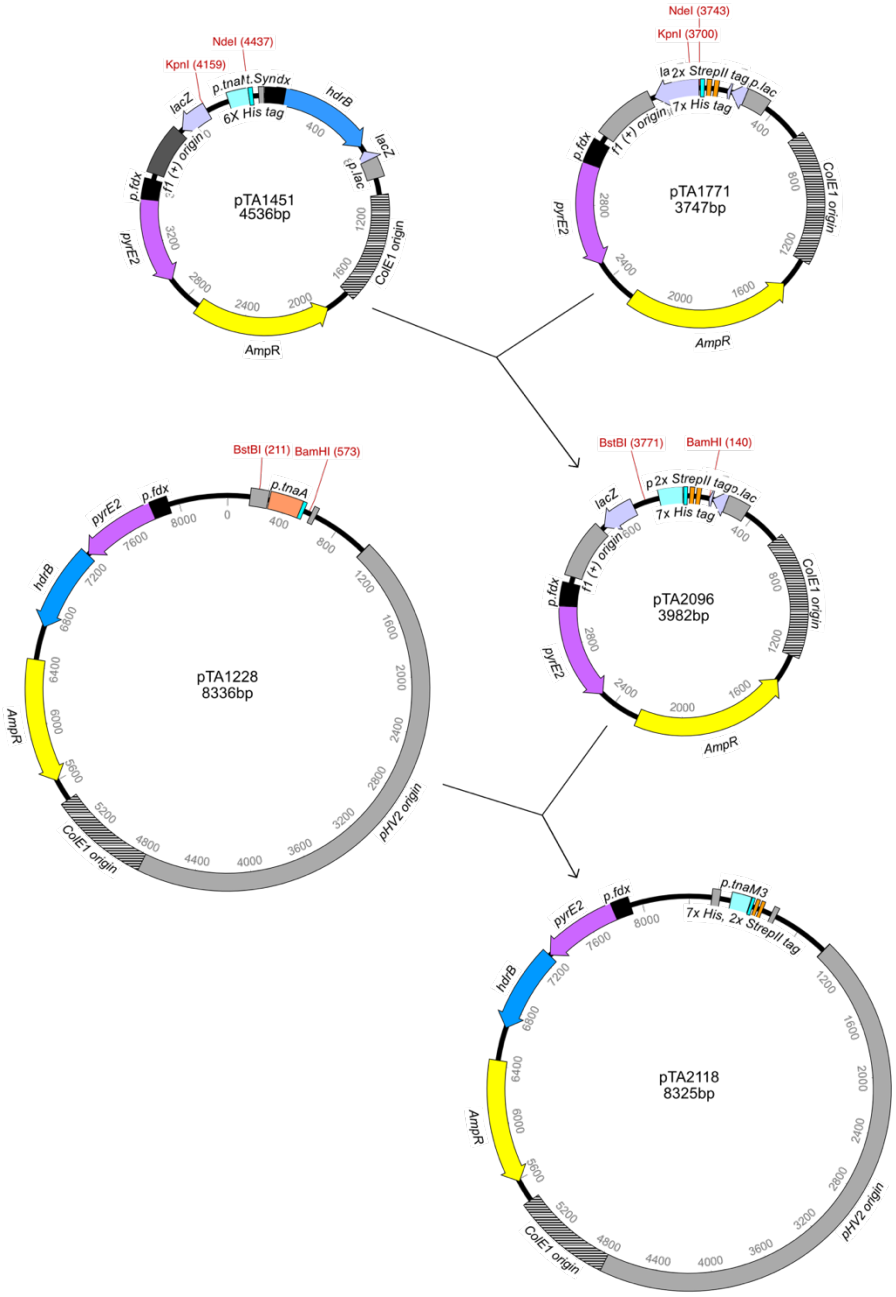


Figure 4.36: **Construction of *p.tnaM3-his*$_7$*-strepII-strepII*-tagged gene episomal over-expression construct, pTA2118.** pTA1451 was digested with *Kpn*I and *Nde*I to liberate the *p.tnaM3* promoter. The promoter was ligated into the *Kpn*I and *Nde*I sites of pTA1771 to create the plasmid pTA2096. pTA2096 was digested with *Bst*BI and *Bam*HI and ligated into the *Bst*BI and *Bam*HI of pTA1228 to create the generic *p.tnaM3-his*$_7$*-strepII-strepII*-tagged gene episomal over-expression construct, pTA2118.

pTA2121 (*p.tnaM3-his$_7$-strepII-strepII-hel308* overexpression construct)

The *his$_7$-strepII-strepII-hel308* gene replacement construct, pTA1856, was digested with *Nde*I and *Pci*I to generate a 3648 bp fragment containing the *his$_7$-strepII-strepII-hel308* CDS. The fragment was ligated into the *Nde*I and *Pci*I sites of p2118 to create the *p.tnaM3-his$_7$-strepII-strepII-hel308* overexpression construct, pTA2121 (Figure 4.37).

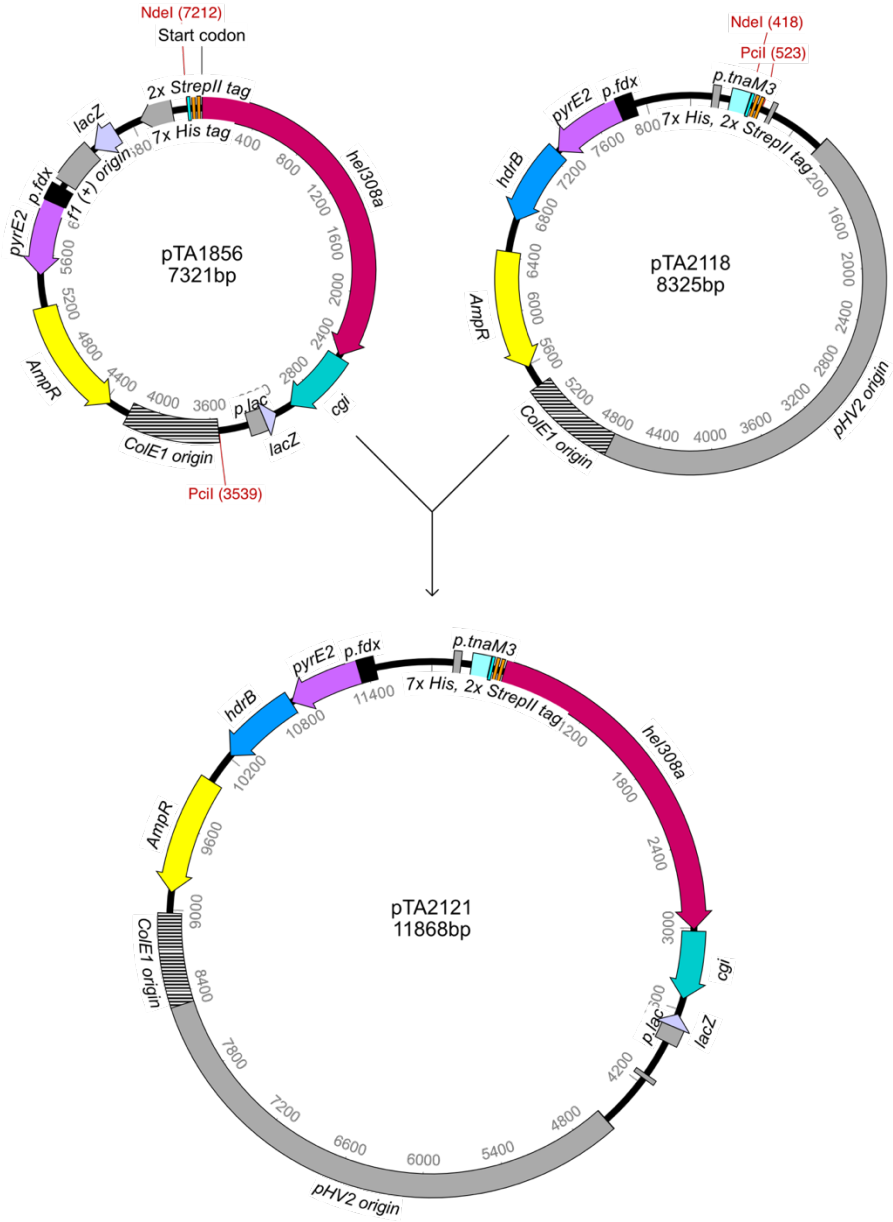


Figure 4.37: **Construction of *p.tnaM3-his$_7$-strepII-strepII-hel308* overexpression construct, pTA2121.** pTA1856 was digested with *Nde*I and *Pci*I to generate a 3648 bp fragment containing the *his$_7$-strepII-strepII-hel308* CDS, which was ligated into the *Nde*I and *Pci*I sites of p2118 to create the *p.tnaM3-his$_7$-strepII-strepII-hel308* overexpression construct, pTA2121.

The *p.tnaM3-his$_7$-strepII-strepII-hel308* overexpression construct, pTA2121, was introduced to H4181 (Δ*pyrE2, bgaHa-Bb, leuB-Ag1,* Δ*hdrB,* Δ*hel308*) to generate the strain H4195, strain was verified by *hdrB* phenotype. To prevent recombination between the episome and the chromosomal *hel308*, the expression strain first needed to be deleted for *hel308*. To maintain selection for the episome, the expression strain also needed to be deleted for *hdrB* before the introduction of the episome. In the interest of time, H4181 was used for the transformation as there were no existing Δ*hdrB* Δ*hel308* strains in a H2047 background and because purification was carried out by StrepTactin affinity only, therefore the presence of the major His-contaminants should have little effect on purity of the protein.

pTA2122 (*p.tnaM3 - his$_7$ – strepII – strepII - hel308 - D145N* overexpression construct)

The generic gene tagging construct, pTA1451, was digested with *Kpn*I and *Nde*I to liberate the fragment containing the *p.tnaM3* promoter. The promoter was ligated into the *Kpn*I and *Nde*I sites of pTA1856 to create the plasmid pTA2081. The central *hel308-D145N* fragment from pTA1335 (see Chapter 3: Separation of function in Hel308 for more detail) was digested using restriction enzymes *Bsr*GI and *Nsi*I, and was ligated into the *Bsr*GI and *Nsi*I sites in pTA2081 to create the plasmid pTA2112. pTA2112 was digested with *Nde*I and *Pci*I to generate a 3648 bp fragment containing the *his$_7$-strepII-strepII-hel308-D145N* CDS. The fragment was ligated into the *Nde*I and *Pci*I sites of p2118 to create the *p.tnaM3-his$_7$-strepII-strepII-hel308-D145N* overexpression construct, pTA2122 (Figure 4.38).

The *p.tnaM3-his$_7$-strepII-strepII-hel308-D145N* overexpression construct, pTA2122, was introduced to H4181 (Δ*pyrE2, bgaHa-Bb, leuB-Ag1,* Δ*hdrB,* Δ*hel308*) to generate the strain H4196, strain was verified by *hdrB* phenotype.

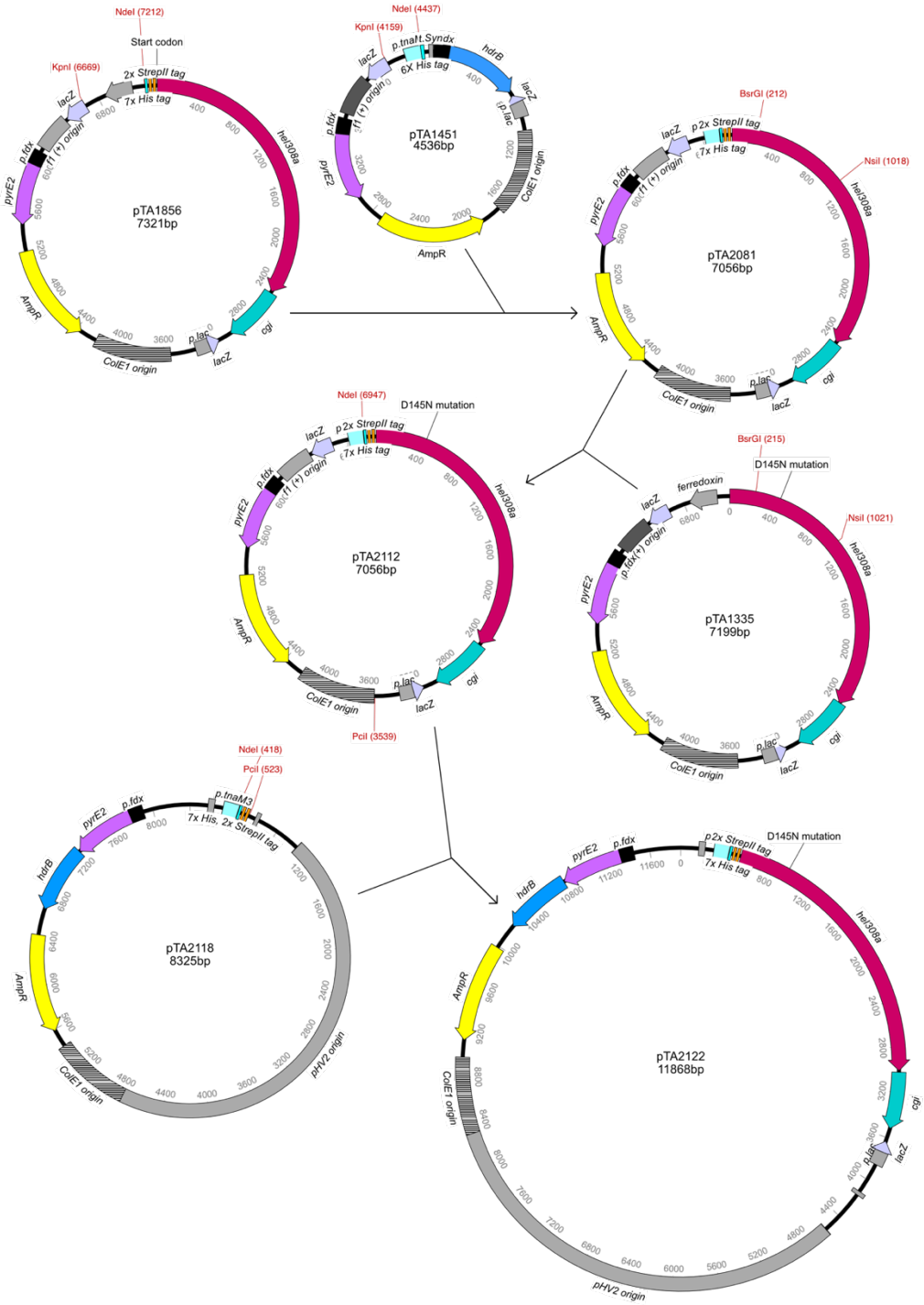


Figure 4.38: **Construction of *p.tnaM3-his$_7$-strepII-strepII-hel308-D145N* overexpression construct, pTA2122.** The *p.tnaM3* was digested from pTA1451 using *Kpn*I and *Nde*I and was ligated into the *Kpn*I and *Nde*I sites of pTA1856 to create the plasmid pTA2081. pTA1335 was digested with *Bsr*GI and *Nsi*I, and the central *hel308-D145N* fragment was ligated into the *Bsr*GI and *Nsi*I sites in pTA2081 to create the plasmid pTA2112. pTA2112 was digested with *Nde*I and *Pci*I to generate a 3648 bp fragment, which was ligated into the *Nde*I and *Pci*I sites of p2118 to create the *p.tnaM3-his$_7$-strepII-strepII-hel308-D145N* overexpression construct, pTA2122.

pTA2201 (*p.tnaM3 – his$_7$ – strepII – strepII - hel308 - F316A* overexpression construct)

The central *hel308-F316A* fragment from pTA1642 (see Chapter3: Separation of function in Hel308 for more detail) was digested using restriction enzymes *Bsr*GI and *Nsi*I, and was ligated into the *Bsr*GI and *Nsi*I sites in pTA2121 to create the *p.tnaM3-his$_7$-strepII-strepII-hel308-F316A* overexpression construct, pTA2201 (Figure 4.39).

The *p.tnaM3- his$_7$-strepII-strepII-hel308-F316A* overexpression construct, pTA2201, could not be generated in *E. coli*, the preferred organism for plasmid construction due to speed of growth and higher copy number than *H. volcanii*. Instead pTA2201 was generated directly in H1209 (Δ*pyrE2,* Δ*hdrB, Nph-pitA,* Δ*mrr*) to produce the strain H4400. The strain was verified by extraction and sequencing of pTA2201 from H4400. Introduction of pTA2201 to H4181 (Δ*pyrE2, bgaHa-Bb, leuB-Ag1,* Δ*hdrB,* Δ*hel308*) was attempted. However, upon checking via sequencing of the plasmid DNA, it was discovered that the episome had spontaneously lost ~1000bp of the *hel308* CDS.

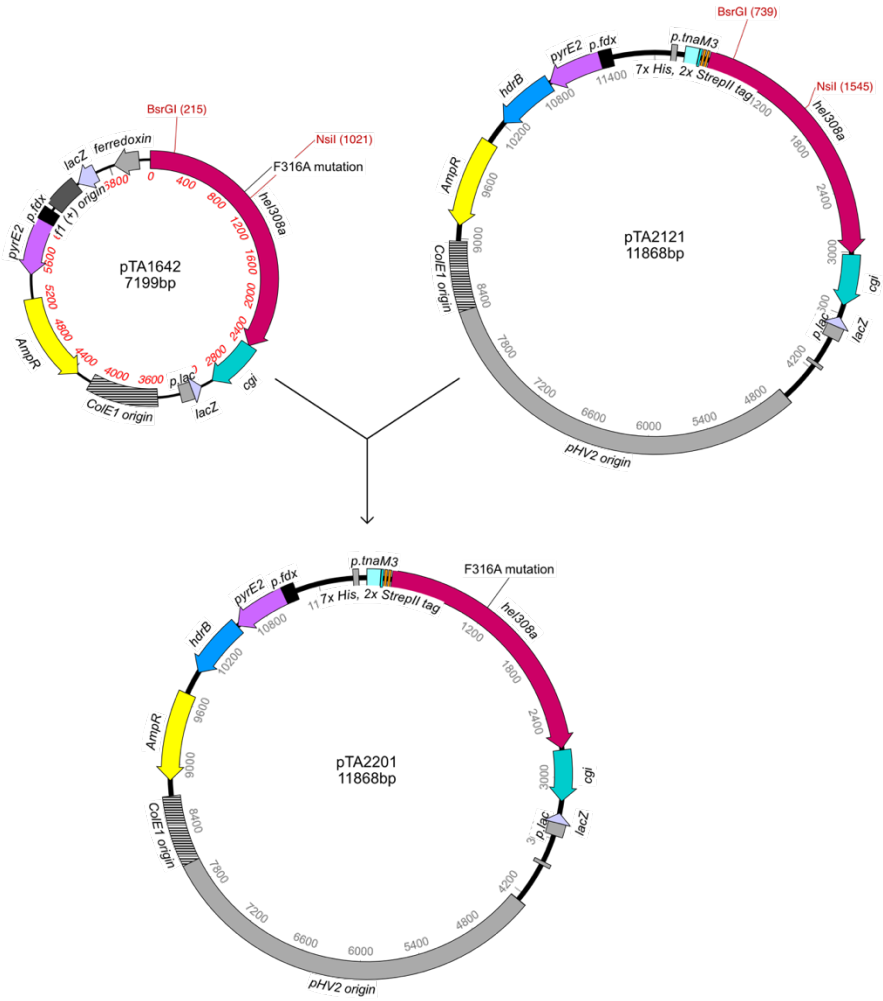


Figure 4.39: **Construction of *p.tnaM3-his*$_7$*-strepII-strepII-hel308-F316A* overexpression construct, pTA2122.** The central *hel308-F316A* fragment from pTA1642 was digested using *Bsr*GI and *Nsi*I and was ligated into the *Bsr*GI and *Nsi*I sites in pTA2121 to create the *p.tnaM3-his*$_7$*-strepII-strepII-hel308-F316A* overexpression construct, pTA2201.

4.4.6 Protein overexpression and purification

Induction of expression

Protein induction is under the control of the tryptophan inducible promoter *p.tnaM3*. High concentrations of tryptophan, >1mM, affect the growth of *H. volcanii* (Allers *et al.*, 2010) therefore, the cultures were first grown to an A_{650} = ~0.8 before induction of protein expression with 6 mM tryptophan overnight.

Protein purification via StrepTactin Chromatography

Proteins were purified from the cell lysate using gravity columns packed with StrepTactin Sepharose. Protein was eluted using 5 mM D-desthiobiotin.

Purification of N-terminally His_7-StrepII-StrepII-tagged Hel308

Episomally expressed *p.tnaM3-his$_7$-strepII-strepII-hel308* was purified via StrepTactin chromatography only, from the strains H4195 (*wild-type hel308*), H4196 (*D145N*), and H4400 (*F316A*) following induction of the *p.tnaM3* promoter using 6mM tryptophan. Western blotting was used to identify purified His_7-StrepII-StrepII-Hel308 by probing for the His-tag (see Material and Methods) (Figure 4.40 B). The ~30 kDa protein, was highly overexpressed but was not detectable via probing with a specific anti-His_6 antibody and is therefore unlikely to be a truncated Hel308 protein. Mass spectrometry was used to analyse all indicated bands, the 30 kDa protein was identified as HVO_2382 DlnA (Figure 4.40 A).

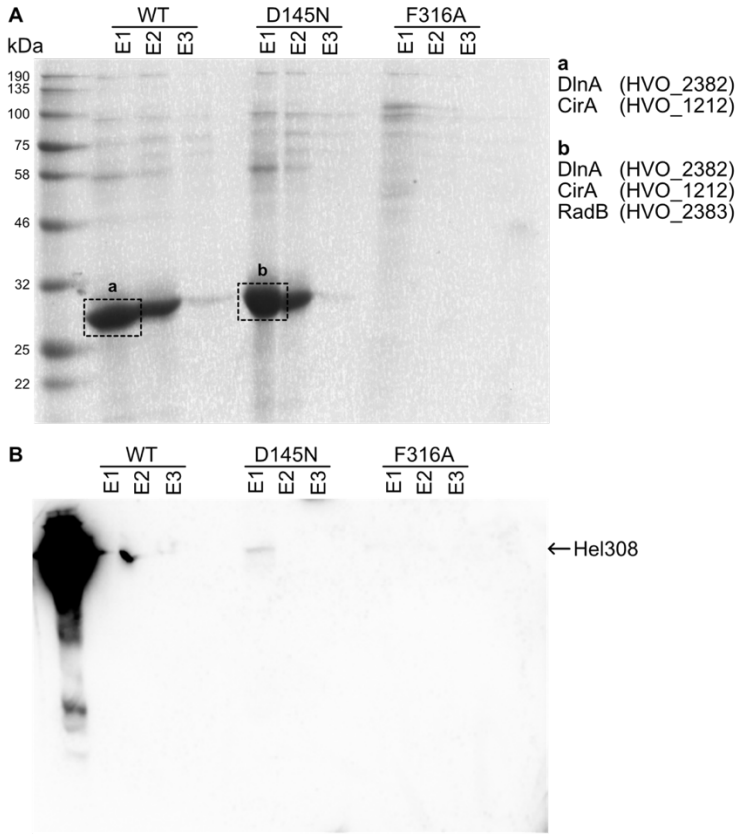


Figure 4.40: **Overexpression and purification of *p.tnaM3-his$_7$-strepII-strepII-hel308-* point mutants** A) expression of *p.tnaM3-his$_7$-strepII-strepII-hel308-* was induced by addition of 6mM tryptophan to cultures of the strains H4195 (*wild-type*), H4196 (*D145N*), and H4400 (*F316A*). *p.tnaM3-his$_7$-strepII-strepII-hel308-* was purified via StrepTactin chromatography using a 3 step elution with 0.8 ml, 1.4 ml, and 0.8 ml of elution buffer containing 5 mM D-desthiobiotin (E refers to elution number). B) A His_x tagged protein (~100 kDa) was detected via western blot using a specific anti-His_6 antibody. This is presumed to be His_7-StrepII-StrepII-tagged Hel308 (indicated). A ~30 kDa protein (indicated as Z) was highly overexpressed but was not detectable via probing with a specific anti-His_6 antibody and is therefore unlikely to be a truncated Hel308 protein. Bands were identified via mass spectrometry.

Table 4.5: **Proteins identified by mass spectrometry**

Band a (~30 kDa)

Protein accesion	Identified proteins	HVO number	Predicted MW (Da)	MASCOT score	No. of peptides	Peptide sequences
D4GWN0	**DlnA**	2382	26310	1743	426(282)	23(19)
D4GTX4	rpl15	2542	17828	392	6(5)	6(5)
D4GWY5	**KaiC-type CirA**	1212	27563	338	6(5)	6(5)
D4GYI9	DhaL	1545	24838	332	7(5)	6(4)
Q9V2V6/ Q9V2V5	Protease subunit α1/2	1091/ 2923	27613/ 26727	290	6(3)	6(3)
D4GYB7	UPF0148 family protein	3008	21861	270	4(3)	4(3)
D4GVF6	nuoB	0979	26139	267	4(3)	4(3)
D4GQG3	TAT pathway signal sequence domain protein	A0133	30836	204	4(2)	3(2)
D4GZB9	pcnA	0175	26672	204	3(2)	3(2)
D4GU21	fba2	0790	27843	192	4(3)	4(3)
D4GXC5	OsmC-like protein superfamily protein	1289	14220	183	2(2)	2(2)
D4GPB5	lclR family transcription regulator	B01114	28428	172	4(2)	4(2)
D4GY62	upp	2981	24502	147	4(2)	3(2)
D4GX47	DSBA-like	1245	24200	145	3(3)	2(2)
D4GUQ7	lclR family transcription regulator	2110	28334	141	3(1)	3(1)
D4GV36	HpcH/HpaI	2665	28323	122	3(0)	3(0)

Band b (~30 kDa)

Protein accesion	Identified proteins	HVO number	Predicted MW (Da)	MASCOT score	No. of peptides	Peptide sequences
D4GWN0	**DlnA**	2382	26310	1672	425(297)	23(17)
Q9V2V6/ Q9V2V5	Protease subunit α1/2	1091/ 2923	27613/ 26727	566	14(8)	10(6)
D4GTX4	rpl15	2542	17828	505	12(9)	9(6)
D4GWY5	**KaiC-type CirA**	1212	27563	360	6(5)	6(5)
D4GYI9	DhaL	1545	24838	346	7(4)	6(4)
D4GU21	fba2	0790	27843	342	7(4)	7(4)
D4GV36	HpcH/HpaI	2665	28323	287	6(3)	6(3)
D4GQG3	TAT pathway signal sequence domain protein	A0133	30836	273	6(4)	5(3)
D4GYG5	aglP	1522	26263	258	6(3)	6(3)
D4GY62	upp	2981	24502	254	7(3)	6(3)
D4GPB5	lclR family transcription regulator	B01114	28428	244	4(3)	4(3)

D4GUQ7	lclR family transcription regulator	2110	28334	232	6(1)	6(1)
D4H028	DUF124 family protein	0402	23995	219	3(3)	3(3)
D4GVF6	nuoB	0979	26139	206	4(2)	4(2)
D4GZ20	Hypothetical	1591	28766	196	5(2)	5(2)
D4GYB7	UPF0148 family protein	3008	21861	190	3(3)	3(3)
D4GXC5	OsmC-like protein superfamily protein	1289	14383	188	2(2)	2(2)
D4GSN5	ABC-type transport system ATP-binding protein	2430	26639	181	3(3)	3(3)
D4GWN1	**RadB**	2383	24600	153	4(2)	3(2)
D4GX47	DSBA-like	1245	24200	151	3(3)	2(2)

Protein accession: the UniProt database entry code, e.g. D4GWN0**.**
Predicted MW (Da): The molecular weight in Daltons (Da) of the protein sequence listed as identified by MASCOT.
MASCOT score: the score associated with protein identification.

$$Ions\ score = -10log_{10}(P)$$

P = probability that the observed match is a random event
Individual ions scores > 45 indicate identity or extensive homology ($p<0.05$). Protein scores are derived from ions scores as a non-probabilistic basis for ranking protein hits.
Number of peptides: number of peptides associated with protein identification by MASCOT.
Peptide sequences: the number of distinct peptide sequences associated with the protein identified by MASCOT.

DlnA (HVO_2382)

DlnA is found in the highly conserved RadB neighbourhood. DlnA belongs to the YqgF family of endonucleases. In *E. coli*, YqgF is thought to function as an alternative to RuvC (Aravind, Makarova and Koonin, 2000), which acts to resolve Holliday junctions (Rocha, Cornet and Michel, 2005). However, the role of DlnA is yet to be determined.

CirA (HVO_1212)

There are four cyanobacterial-like circadian rhythm gene homologs in *H. volcanii* (Hartman *et al.*, 2010), designated *cirA, cirB, cirC,* and *cirD.* These genes are member of the RecA-superfamily ATPases (Haldenby, White and

Allers, 2009) and resemble the KaiC-family of cyanobacterial clock proteins, which act to regulate rhythmic gene expression and to control the timing of cell division. The four *H. volcanii cir* genes are distantly located from one another around the chromosome, with *cirA* located among a group of flagellar genes. The *cirA* gene is divergently transcribed from its flanking genes and appears to be essential (Maniscalco *et al.*, 2014).

RadB (HVO_2383)

RadB, a paralogue of RadA, is required for efficient homologous recombination. Homologous recombination involves the exchange of nucleotide sequence between near-identical DNA sequences for which RadA, the archaeal recombinase, is essential. RadB assists in the formation of the RadA nucleoprotein filament on ssDNA (Wardell *et al.*, 2017) (Figure 3.4).

The possible significance of these interactions will be discussed in further detail in Section 4.5: Discussion.

4.5 Discussion

4.5.1 Hel308 interacting partners *in vivo*

Hel308 may interact with the CMG complex and replication factors in *H. volcanii*

Archaeal replication resembles its eukaryotic counterpart, for which the CMG complex is essential. The eukaryotic CMG complex comprises Cdc45, MCM, and GINS (Makarova, Koonin and Kelman, 2012). Whilst *H. volcanii* MCM and GINS homologues are known, a Cdc45 homologue has yet to be identified. Bioinformatic and structural studies have revealed that human Cdc45 and archaeal RecJ are evolutionarily related to bacterial RecJ (Makarova, Koonin and Kelman, 2012; Pellegrini, 2017) and therefore, have led to the proposal that archaeal RecJ is the Cdc45 homologue. Both the RecJs 3 and 4 and MCM, were identified in band e (Figure 4.31 B), suggesting that either RecJ3 or RecJ4 could be the Cdc45 homologue. Due to sequence and structural similarities between RecJ3 and RecJ4, it is probable that RecJ3 has a similar role to RecJ4 or that either protein is redundant (See chapter 6 for further details). Furthermore, RecJ4 co-purified with the replication initiation protein Orc1 (Darya Ausannikava, University of Nottingham, personal communication), providing further evidence that RecJ4 is the Cdc45 homologue. The potential interaction of Hel308 with RecJ and MCM implicates Hel308 in DNA replication, which is consistent with the proposal that Hel308 acts in the restart of stalled replication forks.

Cdc48a/b/d proteins (HVO_2380/2700/1907)

The identification of the Cdc48 proteins is unsurprising as they are central components in many ubiquitin-mediated pathways (Baek *et al.*, 2013) and therefore, will co-purify with proteins-of-interest acting in a wide range of metabolic processes. Cdc48d was identified as a common contaminant of nickel affinity chromatography (Allers *et al.*, 2010) however, this method of protein purification was not used. Interestingly, three out of the four Cdc48 homologues in *H. volcanii* where identified via mass spectrometry, suggesting that Cdc48 is being recruited either directly or indirectly, via the remodelling of the damage DNA, by Hel308. It is possible that Cdc48 is acting to disassemble resultant replication forks.

Orc5 (HVO_1725)

Orc5 is one of several Cdc6/Orc1 homologues present in *H. volcanii.* The main chromosome of the wild isolate, DS2, has three origins of replication, *oriC1-3*, each with a distinct associated Orc protein. Orc5 is associated with the *oriC2* replication origin (Norais *et al.*, 2007; Hawkins *et al.*, 2013). The potential interaction of Orc5 and Hel308 places Hel308 at the site of DNA replication, which is consistent with the proposal that Hel308 acts in the restart of stalled replication forks.

PriS (HVO_2697)

PriS is the small catalytic subunit of the eukaryotic-type DNA primase. DNA primases synthesise RNA primers on both the leading and lagging strands in order to initiate DNA synthesis (Le Breton *et al.*, 2007). The potential interaction of PriS and Hel308 places Hel308 at the site of DNA replication, which is consistent with the proposal that Hel308 acts in the restart of stalled replication forks.

Hel308 may interact with a variety DNA repair proteins in *H. volcanii*

Homologous Recombination Proteins

RecJ3/4, RadA, RadB, and DlnA (HVO_1018/2889, HVO_0104, HVO_2383, and HVO_2382)

Bacterial RecQ interacts with RecJ to generate ssDNA tails during homologous recombination, this is known as the end resection step. (Morimatsu and Kowalczykowski, 2014). These ssDNA tails are then coated with the RecA-family recombinase nucleoprotein filament. Hel308 is a RecQ family helicase

therefore, the potential interaction with RecJ and RadA (the RecA-family recombinase homologue in *H. volcanii*) suggests that Hel308 could act analogously to bacterial RecQ.

Hel308 is proposed to be a regulator of homologous recombination in an anti-recombinase manner. The identification of RadA as an interacting partner to Hel308 supports the proposal that Hel308 is involved in homologous recombination. Protein-pull down assays of the over-expressed Hel308-D145N mutant did find RadB co-purifying with Hel308 however, this interaction was not seen with the over-expressed wild-type protein nor the natively expressed Hel308. This could be a consequence of the low level of RadB in the cell, the levels of RadB are approximately 200 times lower that the levels of RadA in *P. furiosus* (Komori *et al.*, 2000). The D145N mutation is found in the Walker B motif of Hel308 therefore, the mutant protein is ATPase-null and consequently helicase-dead. It has been shown that the D145N mutation decreases the level of recombination in *H. volcanii* (see Chapter 3: Separation of function in Hel308 for further details) suggesting that the anti-recombinase function of Hel308 is helicase-independent. On the other hand, RadB promotes homologous recombination. The presence of RadB suggests high levels of homologous recombination however, RadA was not seen in this particular assay. These data support the "road-block/speed-bump" theory whereby Hel308 first binds DNA tightly. On binding ATP, Hel308 changes conformation and loosens its grip on the DNA but remains static. Hydrolysis of the ATP causes a second conformational change and a release of energy resulting in the paddling motion and translocation along the DNA (see Chapter 3: Separation of function in Hel308 for further details). The Hel308-D145N protein is able to successfully complete DNA and ATP binding but remains static on the ssDNA, blocking access for DNA replication and repair proteins. Due to the relatively loose grip that Hel308-D145N has on the DNA, the helicase can be "knocked-off" granting access to DNA replication and repair proteins. It is possible that RadB is present in order to promote homologous recombination but Hel308-D145N bound to the ssDNA is preventing RadA nucleoprotein filament formation.

Most interestingly, the overexpression of Hel308 (WT) and Hel308-D145N yields little of the target protein but co-purifies a large quantity of a 30kDa protein identified as DlnA (HVO_2382) by mass spectrometry. DlnA is found in the highly conserved RadB neighbourhood (Figure 4.32) and belongs to the YqgF family of endonucleases. In *E. coli*, YqgF has been proposed to function as an alternative to RuvC (Aravind, Makarova and Koonin, 2000), which acts to resolve Holliday junctions (Rocha, Cornet and Michel, 2005). In *H. volcanii*, the role of DlnA is yet to be determined but this interaction with Hel308 may give further insight to how both proteins are functioning. For this reason, it would be of interest to repeat the overexpression of Hel308 in a strain deleted for DlnA. The co-purification of DlnA is not seen during the overexpression of Hel308-

F316A however, this could be explained by the difference in the strain background. The only copy of the *hel308* gene in the overexpression of Hel308 WT and Hel308-D145N is the copy on the overexpression plasmid. For the overexpression of Hel308-F316A, the chromosomal copy of *hel308* is still present. Therefore, it would be of interest to repeat this pull-down in the same genetic background as Hel308 (WT) and Hel308-D145N.

Topoisomerases

Both type I (TopA, HVO_0681) and type II (Topo VI, HVO_1570/1571 and GyrA HVO_1573) topoisomerases where identified as interacting partners of Hel308a *in vivo.* Type I topoisomerases release supercoiling and torsional tension by cleaving one strand of DNA and passing the cleaved strand around the unbroken strand (Changela, DiGate and Mondragón, 2001). By contrast, type II topoisomerases operate by cleaving both strands of DNA and passing a second segment of DNA through the break in an ATP-dependent manner, this is known as the "strand-passage mechanism" (Stuchinskaya *et al.*, 2009). Topoisomerases are able to interconvert supercoiled and relaxed DNA states by the introduction of positive and negative supercoiling. Subsequently, topoisomerases can also introduce and remove DNA catenanes and knots (Collin, Karkare and Maxwell, 2011). Topoisomerases are critical at resolving hemi-catenanes, which form as a result of the dissolution of Holliday junctions, suggesting that Hel308 may have a role in the latter stages of homologous recombination.

Mismatch repair and Nucleotide Excision repair

RecJ3/4, MutLa/b, and UvrA (HVO_1018/2889, HVO_1939/0551, and HVO_0393)

The RecJ family of proteins are typically 5'-3' exonucleases and have been implicated in several DNA repair pathways including homologous recombination, mismatch repair, and base excision repair. It is also thought that RecJ is important in the rescue of stalled replication forks (Nagata *et al.*, 2017). RecJ is responsible for producing ssDNA for mismatch repair and nucleotide excision repair (NER).

DNA Mismatch repair (MMR) corrects mismatched base pairs that arise primarily from replication errors. This is an important process for maintaining genome stability. The MMR proteins, and the overall mechanism, is highly conserved across most species of bacteria and eukaryotes however, are found in only a limited set of archaeal species. In eukaryotes, MutSα sliding clamp recognises base-base mismatches and insertions/deletions of up to two nucleotide in length. Larger nucleotide insertions/deletions are recognised by the MutSβ sliding clamp (Jiricny, 2006; Hsieh and Yamane, 2009). The

discontinuous DNA strand is nicked either upstream (5') or downstream (3') of the DNA damage by MutLα (Kadyrov *et al.*, 2006; Modrich, 2006). The damaged DNA segment is then excised by the exonuclease EXOI in concert with the single stranded DNA binding protein, RPA. DNA polymerase δ, assisted by the PCNA sliding clamp and the clamp loader RFC, fill in the gap and the resultant nick is then sealed by DNA ligase I (Genschel and Modrich, 2009). Similarly in bacteria lacking MutH, mismatched base pairs and short insertion/deletion loops are recognised by MutS and nicked by MutL. The damaged DNA segment is then excised by the exonucleases such as ExoI and RecJ, and a helicase such as UvrD, both in concert with the single stranded DNA binding protein, SSB. DNA polymerase III fills in the gap and the resultant nick is then sealed by DNA ligase. The archaeal MutL and RecJ3/4 were identified as potential interacting partners of Hel308 in this study.

Nucleotide excision repair (NER) is a highly versatile DNA damage removal pathway and is considered to be relatively 'error free'. Lesions repaired by NER tend to cause significant distortion to the DNA helix (de Laat, Jaspers and Hoeijmakers, 1999). The stages of NER are lesion recognition, 3' and 5' sequential dual incisions, unwinding, patch repair synthesis, and ligation. These general stages are conserved throughout all three domains of life. *H. volcanii* contain a mixture of bacterial UvrABC and eukaryotic XPF orthologues (White, 2003) including the NER proteins UvrA and the long version of archaeal XPF, Hef, which is found exclusively in euryarchaea and contains nuclease and helicase domains. The bacterial and eukaryotic pathways could act in tandem within the cell alternatively, the archaeal NER proteins may possess additional functions. An archaea-specific NER system has not yet been demonstrated *in vivo* or *in vitro* therefore, the archaeal NER system remains largely unknown. The NER protein UvrA was identified as a potential interacting partner of Hel308 in this co-purification assay.

These potential interactions implicate Hel308 in the nucleotide excision repair and mismatch repair pathways.

Transcription-coupled repair

RpoA1/2 and UvrA (HVO_0349/0350 and HVO_0393)

Transcription-coupled repair (TCR) is a sub-pathway of nucleotide excision repair (NER) in which RNA polymerase-arresting lesions, located in the transcribed strand of active genes, are preferentially targeted by the UvrABC complex. TCR has been observed in a variety of bacterial and eukaryotic organisms but few archaeal species. TCR was observed in *Halobacterium sp. NRC-1*, which is dependent on *uvrA* however, the same was not observed in *H. volcanii* (Stantial *et al.*, 2016). It is likely that *H. volcanii* contains a novel mechanism for TCR that is not dependent on UvrA. Hel308 acts as an anti-

recombinase therefore, it is possible that Hel308 promotes DNA repair via NER instead of homologous recombination either by acting with UvrA directly at DNA lesions or, acting during TCR at the sites of arrested RNA polymerases.

Lhr2 (HVO_1333)

Like Hel308, Lhr2 is an SF2 helicase. Phylogenetic analysis of the distribution of SF2 helicase reveals that Lhr is the most abundant helicase in archaea with Hel308/Hjm as the second most abundant member of SF2 families (Chamieh, Ibrahim and Kozah, 2015). Lhr is a 3′–5′ translocase able to unwind both RNA:DNA hybrids and DNA:DNA duplex (Fairman-Williams, Guenther and Jankowsky, 2010; Ordonez and Shuman, 2013). It is probable that Lhr2 has a similar role to Hel308 or that either protein is redundant. Further evidence to support this is that an *lhr2 hel308* double deletion mutant could not be obtained during the attempts of this study (see chapter 5: Nucleases and helicases for further details).

CRISPR-Cas

Cas7 (HVO_A0207)

Cas7 was identified as a potential interacting partner with Hel308. *H. volcanii* possesses a single type I-B CRISPR-Cas system in which Cas7 forms Cascade-like complex along with the Cas proteins 5 and 6. The Cas5/6/7 complex is important for the stability and maintenance of the CRISPR-derived RNAs (crRNAs) (Brendel *et al.*, 2014; Maier, Dyall-Smith and Marchfelder, 2015; Maier *et al.*, 2019). However, Cas7 was identified in only one band in the presence of high levels of MMC therefore, it is possible that Cas7 does not interact with Hel308 directly but rather interacts with the MMC damaged DNA.

Rad50 (HVO_0854)

Rad50 associates with Mre11 in order to target and repair DNA double-strand breaks (DSBs) via microhomology-mediated end-joining (Delmas *et al.*, 2009). MMC is an interstrand crosslinker which occasionally causes DSBs. It is unsurprising that Rad50 was identified in the presence of high levels of MMC. Interestingly, Mre11 was not observed despite forming a very stable complex with Rad50.

Hel308 activity recruits cell division proteins in *H. volcanii*?

CirA, CetZ1, and FtsZ1 (HVO_1212, HVO_2204, and HVO_0717)

Cell division proteins are crucial for successful cell division, shape, and motility. Cyanobacterial-like circadian rhythm genes act to regulate rhythmic gene expression and to control the timing of cell division. The *cirA* is one such gene in *H. volcanii*, located among a group of flagellar genes and is essential (Maniscalco *et al.*, 2014). CetZ1 and FtsZ1 are cell division GTPase proteins belonging to the FtsZ/tubulin superfamily. FtsZ1 forms a ring at the cell mid-region, indicating the future site of cell division. Inactivation of FtsZ1 renders the cell unable to divide. Conversely, inactivation of CetZ1 does not affect cell division. CetZ1 forms dynamic cytoskeletal structures in order to remodel the cell envelope from an irregular plate-shaped cell into a rod-shaped cell type that is essential for normal swimming motility (Duggin *et al.*, 2015). The identification of these proteins is more likely a consequence of increased levels of repair and replication in the presence of MMC and not a direct interaction with Hel308.

Other potential Hel308 interacting partners *in vivo*

Five other proteins were identified as potential interacting partners to Hel308 *in vivo.*

Rnj (HVO_2724)

Ribonuclease J is a bacterial RNase J-like ribonuclease. Bacterial RNase J is a member of the β-CASP family of ribonucleases involved in mRNA processing and degradation. (Phung *et al.*, 2013). Ski2 family helicases are known to unwind RNA:DNA hybrids, both Lhr2 and Hel308 have been shown to demonstrate this activity (Woodman and Bolt, 2009; Fairman-Williams, Guenther and Jankowsky, 2010; Ordonez and Shuman, 2013). It is possible that archaeal RNaseJ (aRNaseJ) is recruited to the DNA damage sites in order to degrade the resultant R-loops, any remaining bound RNA is then removed by the Ski2 family helicases.

CobN and NrdA (HVO_B0050 and HVO_2452)

Cobaltochelatase, in which CobN is one of three subunits (Cameron *et al.*, 1991; Debussche *et al.*, 1992), catalyses the insertion of Co^{2+} into sirohydrochlorin during the anaerobic biosynthesis of cobalamin (vitamin B_{12})(Yin *et al.*, 2006).

NrdA then cleaves the C-Co bond in adenosylcobalamin (Hagmann, 1997; Stubbe and Riggs-Gelasco, 1998) to produce a 5'-deoxyadenosyl radical for the formation of deoxyribonucleotides from ribonucleotides. The presence of these two proteins bridge the gap between the RNA processing enzymes and the DNA replication and repair enzymes found, by recycling the ribonucleotides released from RNA degradation and converting them to deoxyribonucleotides for new DNA synthesis and repair of legions.

Orc Proteins

Orc8 and Orc15 (HVO_C0057 and HVO_1537)

Orc8 and Orc 15 are both Cdc6/Orc1 homologues but they appear to function in completely different pathways. Orc8 is located on the mini-chromosome pHV1 and is not origin associated (Norais *et al.*, 2007) therefore, the interaction with Hel308 may indicate a repair role for Orc8. Orc8 was identified in bands c and f (Figure 4.31 B), which correspond to a protein size ~50kDa in the presence of MMC at 0.3 µg/ml and 0.5 µg/ml respectively.

Orc15, along with Orc16, has been implicated in glycerol metabolism. A high concentration of glycerol within the cell is crucial for maintaining the osmotic pressure in high salt conditions in bacteria but not archaea (Sherwood, Cano and Maupin-Furlow, 2009). However, Orc15 was only identified in band f (Figure 4.31 B) in the presence of 0.5 µg/ml MMC, which results in only 0.1% cell survival. It is possible that the presence of Orc15 is a result of high cell stress and not a true interaction with Hel308.

4.5.2 Development of Methods

Generation of pTA2118

pTA2118 is similar to pTA1369 but the *p.tna-his*$_6$ promoter tag combination has been replaced with the *his*$_7$*-strepII-strepII-* N-terminal tandem tag under the control of the *p.tnaM3* promoter, which allows for episomal over-expression of an N-terminally His$_7$-StrepII-StrepII-tagged recombinant protein of interest. The construct also features multicloning sites either side of the expression locus for the insertion of the gene of interest upstream and downstream regions that enables integration onto the chromosome and tryptophan induced expression at the gene locus (Figure 4.41).

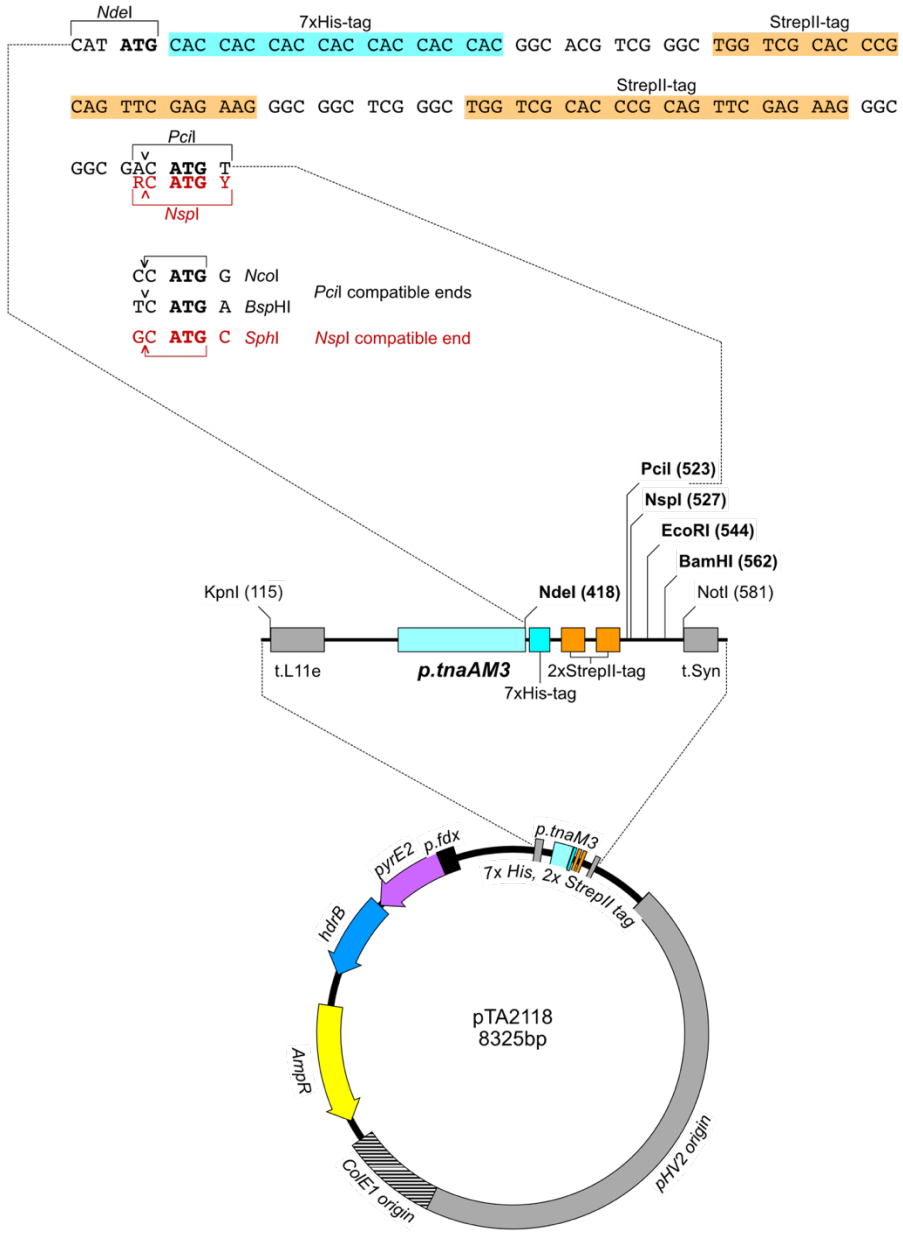


Figure 4.41: **pTA2118 for conditional overexpression of His_7-StrepII-StrepII-tagged proteins.** pTA2118 contains a *his_7-strepII-strepII-tag* downstream of the tryptophan inducible promoter *p.tnaAM3*. The genes *pyrE2* and *hdrB* allow for selection in *H. volcanii*. Restriction sites, and compatible enzymes, available for inserting genes of interest are displayed on the schematic. Abbreviations: *t.L11e* (terminator), *p.tnaAM3* (tryptophan inducible promoter with the M3 point mutation), *his_7-tag* (heptahistidine tag), *strepII*-tag (streptavidin short-affinity tag), *t.Syn* (terminator), *pHV2* origin (*H. volcanii* origin of replication), *ColE1* origin (*E. coli* origin of replication), *AmpR* (ampicillin resistance gene, *E. coli*), *hdrB* (thymidine biosynthesis, *H. volcanii*), *pyrE2* (uracil biosynthesis, *H. volcanii*), *p.fdx* (ferredoxin promoter). Y = T/C and R = A/G.

Purification of His_7-StrepII-StrepII Tagged Proteins

As anticipated, purification of *his_7-strepII-strepII* tagged *hel308* via nickel affinity chromatography only yields many histidine-rich protein contaminants in

the elutions E100-E500. Application to a StrepTactin column, following nickel affinity chromatography, removes a large number of these contaminating bands; however, recovery of Hel308 is very poor. Purification of *his$_7$-strepII-strepII* tagged *hel308* via StrepTactin affinity chromatography only has few contaminants in elutions E1-E3, the same contaminants are also present in elutions E1-E3 of the sequential purification strategy and a greater recovery of Hel308. Therefore, purification of His$_7$-StrepII-StrepII-Hel308 is most successful via StrepTactin chromatography alone. The presence of the His$_7$-tag is useful for Western blotting to verify protein purification independently of the Streptactin column.

4.6 Future Perspectives

How does Hel308 function?

This study has revealed interactions of Hel308 with proteins involved in DNA repair, replication, and RNA processing. To confirm these interactions, repetition of the protein:protein interaction assays alongside an empty vector control is required. Genuine interactions may also be determined by the reverse protein:protein interaction assays using tagged versions of the potential interaction partners. Identification of the presence of Hel308 via mass spectrometry would provide further evidence that the interaction *in vivo* is genuine.

Where possible, it would be of interest to delete the genes encoding the interaction partners of Hel308 in combination with *hel308*. Inducible complementation assays should be used where the combination of deleted genes appears to be synthetically lethal. Growth, DNA damage, and recombination phenotypes of viable strains should also be determined. These analyses would give further insight and aid understanding of the role of Hel308 in DNA repair, replication, and RNA processing.

4.7 Conclusion

Plasmids with new protein tag combinations and protein pull-down strategies have been developed to allow for the overexpression and purification of halophilic proteins in *H. volcanii.* Identification of potential Hel308 interacting proteins *in vivo* were achieved using these methods however, these data are very preliminary.

Chapter 5: Nucleases and Helicases

5.1 Background

Does Hel308 act in the Hef or Hjc pathway?

The endonucleases Hef and Hjc are thought to restart stalled replication forks in *H. volcanii* (Lestini, Duan and Allers, 2010). The Holliday junction resolvase Hjc binds specifically to Holliday junctions and symmetrically cleaves two opposing strands generating two recombinant duplexes. Hef comprises two distinct domains: an DEAH/D helicase N-terminal domain and an XPF endonuclease family C-terminal domain. Hef acts on nicked, flapped or forked DNA and is able to convert a Holliday junction into a forked structure by introduction of an incision near the branch point (Komori *et al.*, 2002; Horiuchi *et al.*, 2004). Single deletions of *hef* or *hjc* in *H. volcanii* lead to moderate to no sensitivity to DNA crosslinking agents, similarly when deleted in combination with *hel308*. However, the *hjc hef* double deletion is synthetically lethal (Lestini, Duan and Allers, 2010). Deletion of *hef* in combination with *radA* is highly deleterious however, the *hjc* and *radA* combined deletion has the same phenotype as the *radA* single deletion. This suggests that Hjc acts exclusively in homologous recombination whereas Hef is able to act in a pathway that bypasses homologous recombination (Figure 5.1). PINA, PilT N-terminal (PIN)-domain-containing ATPase, is an ATPase and DNA helicase that is highly conserved throughout Archaea (Zhai *et al.*, 2017, 2018). PINA physically and functionally interacts with Hjc to promote Holliday junction migration. In *Sulfolobus islandicus* and *Sulfolobus tokodaii*, PINA has also been shown to have a direct physical interaction with the Hel308a homologue, Hjm (Li *et al.*, 2008; Zhai *et al.*, 2018). *S. tokodaii* Hjm is able to unwind DNA in both 3’ to 5’ and 5’ to 3’ directions, unlike *H. volcanii* Hel308a which has a 3’ to 5’ polarity. The unwinding activity of *S. tokodaii* Hjm is inhibited by Hjc *in vitro* (Li *et al.*, 2008). Furthermore, *in vitro* biochemical analysis of the *S. islandicus* homologues indicates that Hjc, Hjm and PINA coordinate Holliday junction migration and cleavage in a concerted manner. These data suggest that the Hel308/Hjm family helicases, in association with Hjc endonucleases and PINA, are involved in the processing of stalled replication forks. To determine whether Hel308 acts in the Hef or Hjc pathway in *H. volcanii*, genetic interactions between *hjc, hef, pinA* and *hel308* where analysed.

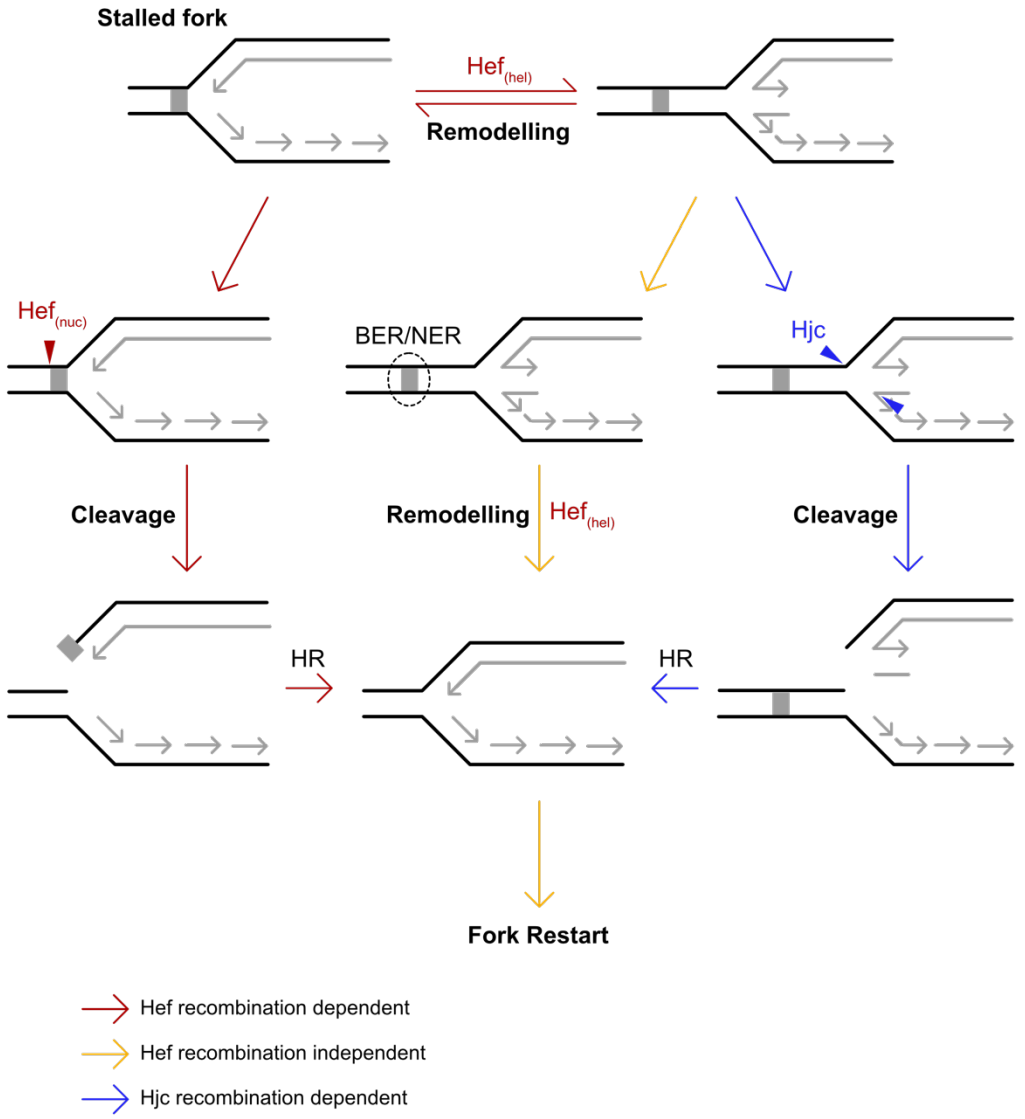


Figure 5.1: **Pathways of replication fork restart by Hef and Hjc.** Hef is proposed to work in both a recombination dependent pathway, involving the cleavage of the replication fork (red arrows), and a recombination independent pathway, involving fork remodelling (yellow arrows). $Hef_{(hel)}$ indicates Hef helicase activity and $Hef_{(nuc)}$ indicates Hef nuclease acivity. Hjc is proposed to act only in a recombination dependent pathway involving the cleavage of the replication fork (blue arrows). Figure adapted from Lestini et al 2010.

CRISPR-Cas

The presence of a clustered regularly interspaced short palindromic repeats (CRISPR)-Cas adaptive immune system is preserved in about 85% of archaea. *H. volcanii* possesses a single type I-B CRISPR-Cas system however, there are several copies of the CRISPR repeats located on the main chromosome and pHV4. CRISPR-Cas immunity proceeds in three stages adaptation, expression, and interference. An invading genetic element releases DNA into the cell which

is recognised as foreign and degraded. A degraded fragment of the nucleic acid (known as a protospacer) flanked by a protospacer adjacent motif (PAM), is selected by the Cas1-Cas2 complex and Cas4, and then integrated as new spacer into the CRISPR locus. This process is known as the adaptation stage. During the expression stage, the CRISPR locus is transcribed into a long precursor known as the pre-crRNA. The endonuclease Cas6 cleaves the pre-crRNA within the repeats generating a pool of CRISPR-derived RNAs (crRNAs) each carrying an individual spacer. The mature crRNAs are incorporated into a multi-protein complex called Cascade. Cascade comprises Cas5, Cas6, and Cas7, and is important for the stability and maintenance of the crRNAs (Brendel *et al.*, 2014; Maier, Dyall-Smith and Marchfelder, 2015). During CRISPR interference, Cascade detects the PAM sequence and the crRNA then investigates the protospacer. If the cell has previously adapted the protospacer, the crRNA spacer region base pairs with the complementary sequence of the invading DNA. This interaction allows the binding of DNA by the Cascade complex and recruits the Cas3 nuclease for degradation of the invading DNA (Maier *et al.*, 2019) (see Chapter 1: Introduction, Figure 1.22).

A CRISPR interference (CRISPRi) system has been developed for *H. volcanii* utilising these mechanisms. Pre-requisites for CRISPRi include the deletion of genes encoding the nucleases Cas3 and Cas6. By introducing crRNAs with spacers that are complementary to genomic DNA, Cascade is guided to these regions and binds the DNA. Since the *cas3* gene is deleted, the DNA cannot be cleaved and remains intact. Therefore, Cascade acts as a "road-block" by blocking access to the DNA. The *cas6* deletion prevents the maturation of endogenous crRNAs and allows all Cascade complexes in the cell to be loaded with the synthetic crRNA (Maier *et al.*, 2019).

Strains deleted for *cas3* have demonstrated an increase in recombination (Personal correspondence, Uri Gophna) therefore, it is of interest to analyse the recombination levels in strains deleted for *cas3* in combination with the hyper-recombinant *hel308* mutants and Δ*hel308*.

5.2 Aims and Objectives

Genetic analysis of *hel308*, in combination with other helicases and nucleases such as *pinA, hef, hjc,* and *lhr* will give insights into the role that Hel308 plays in these DNA repair pathways. Additionally, genetic analysis of *hel308* in combination with c*as3,* should identify the role, if any, that Hel308 plays in the *H. volcanii* CRISPR-Cas adaptive immune system.
The objectives of this chapter are to:

- Generate strains deleted for *hel308* in combination with genes encoding helicase and nuclease proteins.
- Analyse the generation time of strains deleted for helicases and nucleases.
- Analyse the phenotypes of *hel308* deletion strains after treatment with DNA damaging agents such as UV, which causes double strand breaks, and MMC, an interstrand crosslinking agent.
- Generate strains deleted for *cas3* in combination with Δ*hel308, hel308-D145N, hel308-F316A,* and *hel308-R743A.*
- Analyse the recombination frequency of strains deleted for *cas3* in combination with Δ*hel308, hel308-D145N, hel308-F316A,* and *hel308-R743A,* including the crossover: non-crossover product ratio.

5.3 Materials

5.3.1 Strains

Table 5.1 ***Haloferax volcanii*** **strains used in this chapter**

Parent: H26 (*ΔpyrE2*)			
Strain	**Parent**	**Genotype**	**Notes**
H178	H158	*Δhjc*	Constructed by GN
H358	H338	Δ*hef*	Constructed by ZD
H1391	H1379	Δ*hel308*	Constructed by TA

Parent: H53 (*ΔpyrE2, ΔtrpA*)			
Strain	**Parent**	**Genotype**	**Notes**
H1467	H1427	Δ*hjc,* Δ*hel308::trpA+*	Constructed by TA
H1468	H1427	Δ*hef,* Δ*hel308::trpA+*	Constructed by TA
H2085	H2071	*Δhel308*	Constructed by DW
H4249	H4248	Δ*pinA::trpA+*	
H4363	H4362	Δ*pinA*	
H4441	H4440	Δ*pinA,* Δ*hef::trpA+*	
H4443	H4442	Δ*pinA,* Δ*hel308::trpA+*	

Parent: H164 (*ΔpyrE2, bgaHa-Bb, leuB-Ag1, ΔtrpA)*			
Strain	**Parent**	**Genotype**	**Notes**
H2117	H2098	Δ*hel308::trpA+*	Constructed by RG-M
H2397	H2356	*hel308-F316A*	Constructed by RG-M
H2398	H2357	*hel308-R743A*	Constructed by RG-M
H2400	H2213	*hel308-D145N*	Constructed by RG-M
H4386	H4380	Δ*cas3::trpA+*	
H4388	H4382	*hel308-F316A,* Δ*cas3::trpA+*	
H4389	H4383	*hel308-R743A,* Δ*cas3::trpA+*	
H4445	H4444	Δ*hel308::trpA+,* Δ*cas3*	
H4447	H4446	*hel308-D145N,* Δ*cas3::trpA+*	

GN = Greg (Hien-Ping) Ngo, ZD= Zhenhong Duan, DW = David Wallbank, RG-M = Rebecca Gamble-Milner

5.3.2 Plasmids

Table 5.2 **Plasmids used in this chapter.**

Name	Use	Notes
pBluescriptII SK+	Standard *E. coli* vector with functional blue/white screening capability	From Stratagene
pTA131	For making deletions in a *ΔpyrE2* background	Constructed by TA (Allers *et al.* 2004)
pTA163	Integrative vector containing *leuB-Ag1* and *pyrE2*. For use in recombination assays.	Constructed by GN
pTA298	*trpA* genomic clone with flanking *Bam*HI sites	Constructed by TA
pTA370	*hef* deletion construct	Constructed by ZD
pTA1062	Δ*hjc::trpA+*	Constructed by RL
pTA1064	Δ*hef::trpA+*	Constructed by RL
pTA1273	*hel308::trpA+* deletion construct	Constructed by TA
pTA1316	*hel308* genomic clone	Constructed by TA
pTA1335	*hel308-D145N* gene replacement construct	Constructed by TA
pTA1591	*hjc* overexpression construct	Constructed by DW
pTA1592	*hef* overexpression construct	Constructed by DW
pTA1642	*hel308-F316A* gene replacement construct	Constructed by RG-M
pTA1643	*hel308-R743A* gene replacement construct	Constructed by RG-M
pTA1921	*pinA* genomic clone	
pTA1923	*pinA* deletion construct	
pTA1970	*pinA::trpA+* deletion construct	
p131-Cas3up/do	Δ*cas3*	Anita Marchfelder
pTA131-iCas3KO-trpA	Δ*cas3::trpA+*	Anita Marchfelder
pTA927-cas3FLAGN	*FLAG-cas3* expression construct used to generate radiolabelled probe	Anita Marchfelder

TA = Thorsten Allers, GN = Greg (Hien-Ping) Ngo, ZD = Zhenhong Duan, RL = Roxane Lestini, DW = David Wallbank, RG-M = Rebecca Gamble-Milner.

pTA370 (*hef* deletion construct)

The *hef* gene deletion construct, pTA370, was constructed prior to this study by Zhenhong Duan in 2004 (Figure 5.3).

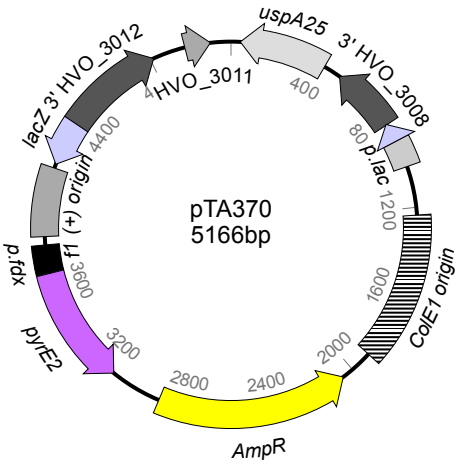


Figure 5.3: **pTA370.** The *hef* gene deletion construct. Constructed by Zhenhong Duan in 2004.

pTA1062 (*hjc::trpA*+ deletion construct)

The *hjc::trpA*+ gene deletion construct, pTA1062, was constructed prior to this study by Roxanne Lestini in 2008 (Figure 5.4).

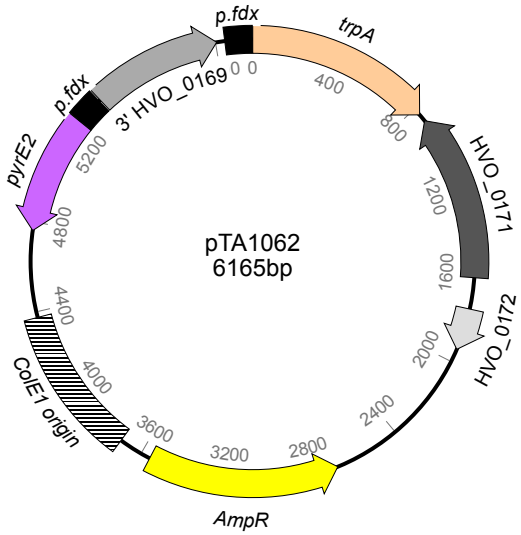


Figure 5.4: **pTA1062.** The *hjc::trpA*+ gene deletion construct. Constructed by Roxanne Lestini in 2008.

pTA1064 (*hef::trpA*+ deletion construct)

The *hef::trpA*+ gene deletion construct, pTA1064, was constructed prior to this study by Roxanne Lestini in 2008 (Figure 5.5).

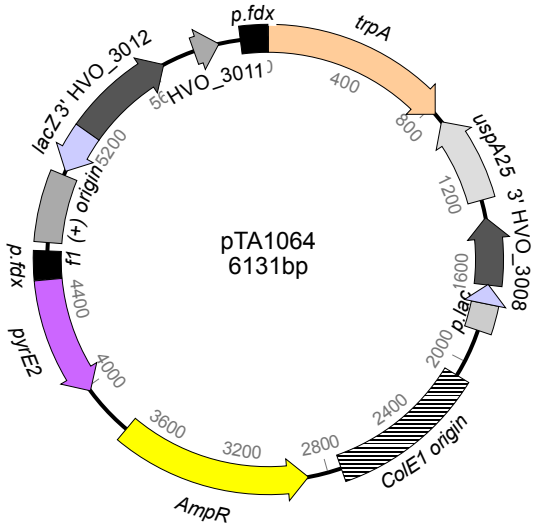


Figure 5.5: **pTA1064.** The *hef::trpA*+ gene deletion construct. Constructed by Roxanne Lestini in 2008.

pTA1591 (*hjc* overexpression construct)

The *hjc* overexpression construct, pTA1591, was constructed prior to this study by David Wallbank in 2015 (Figure 5.6).

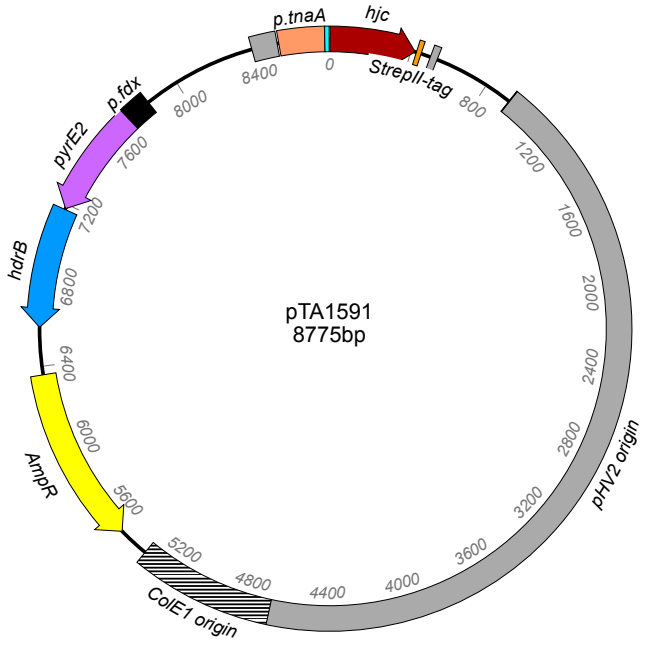


Figure 5.6: **pTA1591.** The *hjc* overexpression construct. Constructed by David Wallbank in 2015.

pTA1592 (*hef* overexpression construct)

The *hef* overexpression construct, pTA1592, was constructed prior to this study by David Wallbank in 2015 (Figure 5.7).

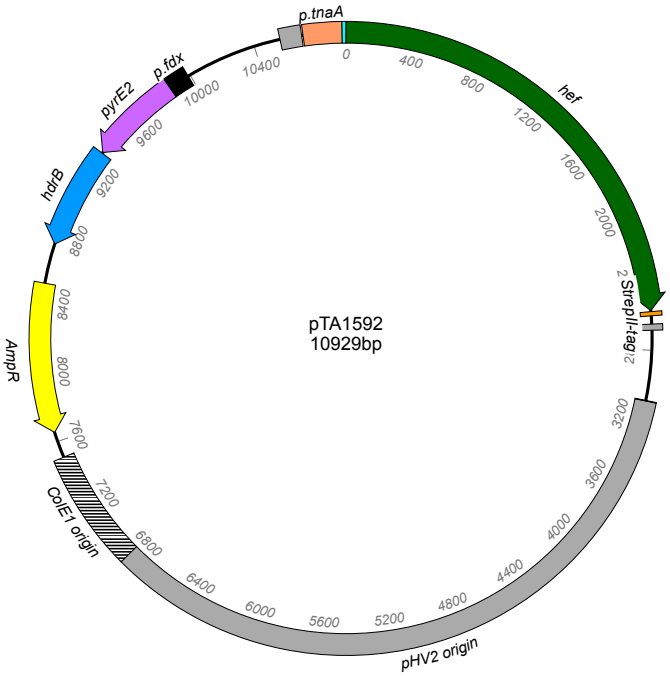


Figure 5.7: **pTA1592.** The *hef* overexpression construct. Constructed by David Wallbank in 2015.

p131-Cas3up/do (*cas3* deletion construct)

The *cas3* deletion construct, p131-Cas3up/do, was provided by Anita Marchfelder (Figure 5.8).

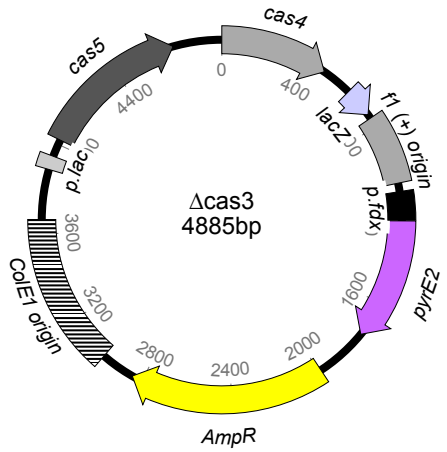


Figure 5.8: **p131-Cas3up/do.** The *cas3* deletion construct**.** Provided by Anita Marchfelder.

pTA131-iCas3KO-trpA (*cas3::trpA+* deletion construct)

The *cas3::trpA+* deletion construct, pTA131-iCas3KO-trpA, was provided by Anita Marchfelder (Figure 5.9).

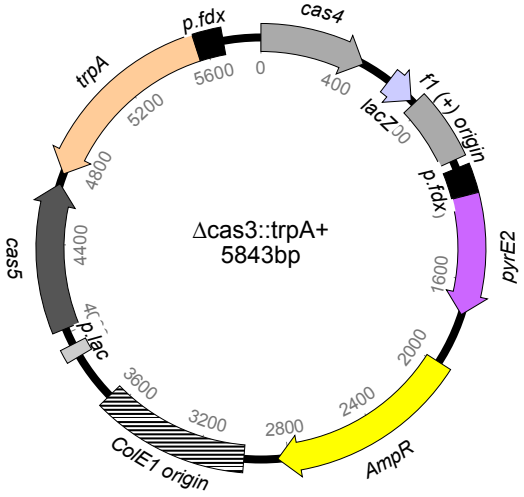


Figure 5.9: **pTA131-iCas3KO-trpA.** The *cas3::trpA+* deletion construct. Provided by Anita Marchfelder.

pTA927-cas3FLAGN (*FLAG-cas3* expression construct)

The *FLAG-cas3* expression construct, pTA927-cas3FLAGN, used to generate the radiolabelled *cas3+* probe was provided by Anita Marchfelder (Figure 5.10).

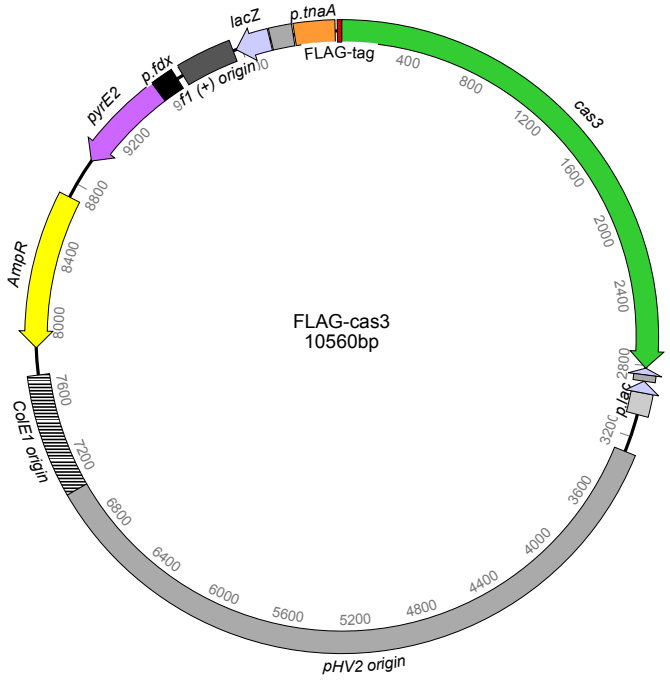


Figure 5.10: **pTA927-cas3FLAGN.** The *FLAG-cas3* expression construct. Provided by Anita Marchfelder.

5.3.3 Oligonucleotides

Table 5.3 **Oligonucleotides used in this chapter.**

Name	Sequence (5'-3')	Description	Used to construct
pBSF2	TTAAGTTGGGTAACGCCAGGG	PCR primer that binds to the pBluescript backbone. Sequencing primer	pTA1923, pTA1951, pTA1991, K53A plasmids
pBSR3	ACCCCAGGCTTTACACTTTATGC	PCR primer that binds to the pBluescript backbone. Sequencing primer	pTA1923, pTA1951, pTA1991
PinA_F	CGAGTAGTTCGTCGAGCCCTCGG	Amplification of 5861 bp region of chromosome containing *pinA*	pTA1921
PinA_R	CGTCGAGGAAGTAGAAGTCGCGGTCG	Amplification of 5861 bp region of chromosome containing *pinA*	pTA1921
delpinABglF	CGCAACaGaTCTCTTACGACACCGACTGC	Forward inverse PCR primer to delete *pinA*	pTA1923
delpinANdeR2	CCGGCACTATCTTCATatGCGTGAGTTTGGGCGGG	Reverse inverse PCR primer to delete *pinA*	pTA1923
BglLinkNdeR	GATCTGGGAATTCCGACGCGTCGCA	Annealed to NdeLinkBglF to form NdeLinkBgl linker oligo with *Nde*I and *Bgl*II ends	pTA1927, pTA1951, pTA1970
NdeLinkBglF	TATGCGACGCGTCGGAATTCCCA	Annealed to BglLinkNdeR to form NdeLinkBgl linker oligo with *Nde*I and *Bgl*II ends	pTA1927, pTA1951, pTA1970

5.4 Results

5.4.1 Generation of *hel308, hef, hjc*, and *pinA* deleted stains

Prior to this study, single and double deletions of the genes *hel308, hef,* and *hjc*, were all generated, with the exception of the *hjc hef* double deletion that is synthetically lethal (Lestini, Duan and Allers, 2010).

5.4.1.1 Plasmid construction

pTA1921 (*pinA* genomic clone)

Genomic DNA extracted from the wildtype strain H26 was used as a template to PCR amplify the *pinA* gene and flanking regions using the primers PinA_F and PinA_R. The product was digested with *Bst*BI and *Apa*I and ligated into the *Cla*I and *Apa*I sites in pTA131 (once passaged though *dam- E. coli,* as *Cla*I site is blocked by overlapping *dam* methylation) to create the *pinA* genomic clone, pTA1921. The construct was confirmed via DNA sequencing (Figure 5.11).

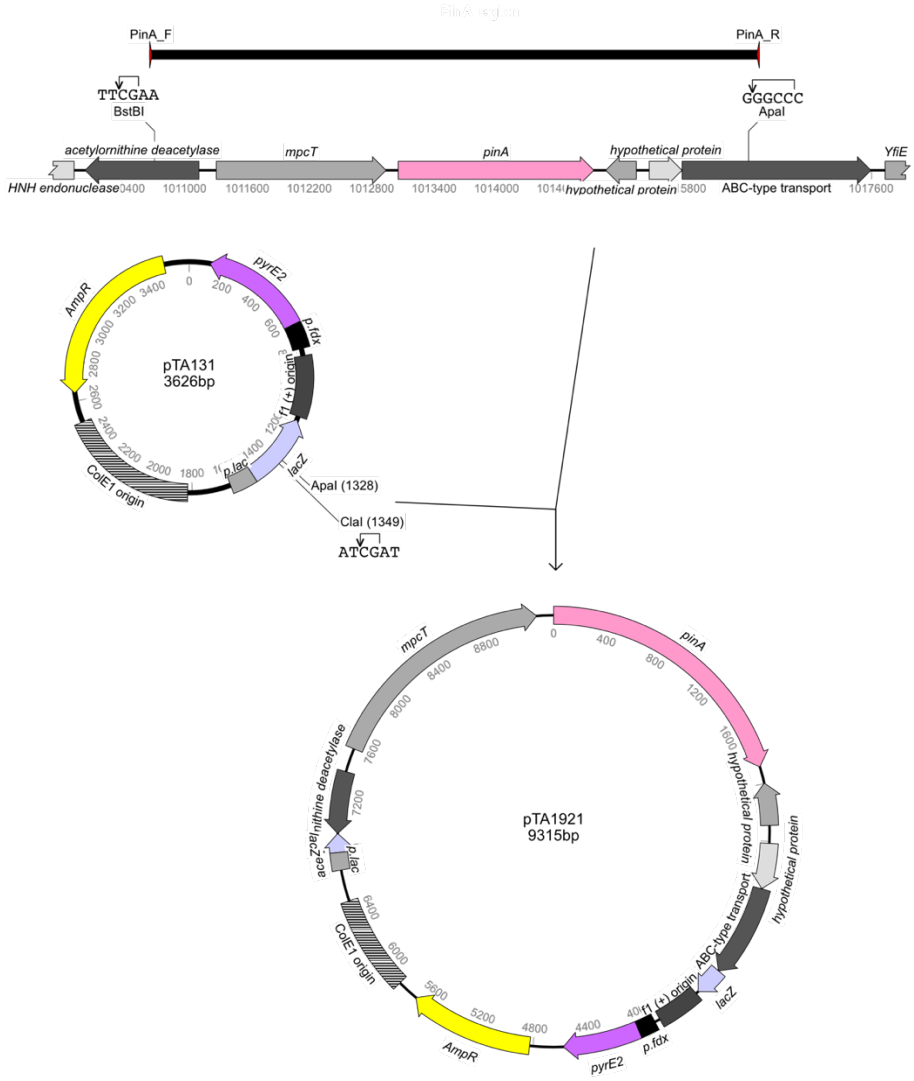


Figure 5.11 **Generation of *pinA* genomic clone, pTA1921.** Genomic DNA extracted from H26 was used as a template to PCR amplify PinA_F to PinA_R The product was digested with *Bst*BI and *Apa*I and ligated into the *Cla*I and *Apa*I sites of pTA131 to create the *pinA* genomic clone, pTA1921.

pTA1970 (*pinA::trpA*+ deletion construct)

Attempts to inverse PCR the *pinA* genomic clone, pTA1921, were unsuccessful therefore the upstream and downstream regions were amplified separately, joined together by means of the BglLinkNde linker, and ligated into pTA131. pTA1921 was used as a template to PCR amplify the upstream region of *pinA* using the primers pBSR3 and delpinANdeR2, generating a 2574 bp product that was then digested with *Nde*I and *Kpn*I to enable the ligation with the BglLinkNde linker. pTA1921 was again used as a template to PCR amplify the downstream region of *pinA* using the primers pBSF2 and delpinABglF, generating a 1536 bp product that was then digested with *Bgl*II and *Xba*I to enable the ligation with the BglLinkNde linker. Once the upstream and downstream regions were ligated together, the fragment was ligated into the *Kpn*I and *Xba*I sites of pTA131 to generate the *pinA* deletion construct, pTA1923. pTA298 was digested with *Bam*HI to generate a 965 bp fragment containing the *trpA* CDS and the *ferredoxin* promoter, which was ligated into the *Bgl*II site of pTA1923. Restriction digest was used to screen for orientation of the *trpA* marker, the same orientation as *pinA* was preferable, and was further confirmed via DNA sequencing to generate the *pinA::trpA*+ deletion construct, pTA1970 (Figure 5.12).

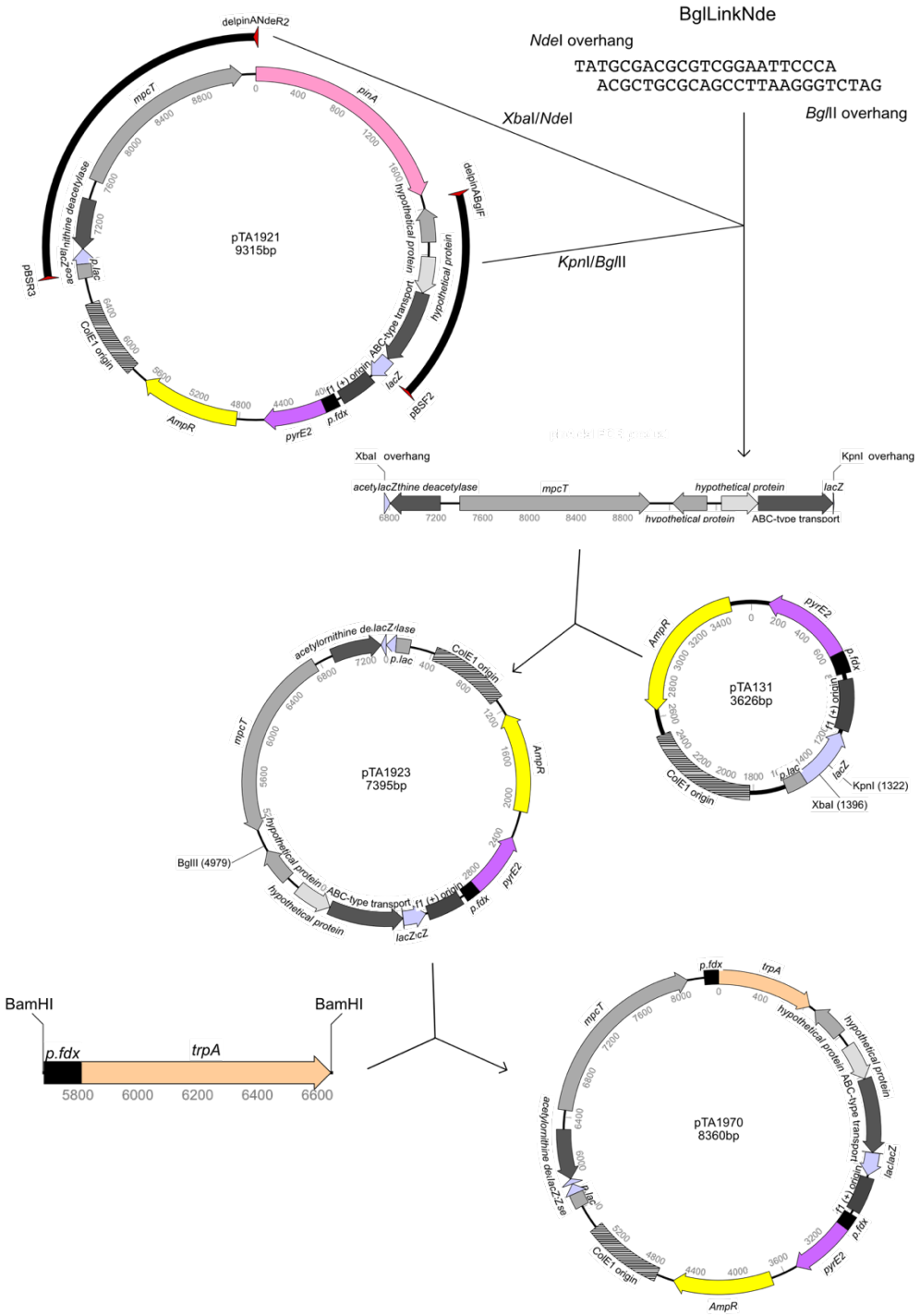


Figure 5.12 **Generation of *pinA::trpA*+ deletion construct, pTA1970.** pTA1921 was used as a template to PCR amplify pBSR3 to delpinANdeR2, generating an 2574 bp product that was then digested with *Nde*I and *Kpn*I. pTA1921 was used as a template to PCR amplify pBSF2 to delpinABglF, generating a 1536 bp product that was then digested with *Bgl*II and *Xba*I. The *pinA* upstream and downstream regions were ligated together. The USlinkDS fragment was ligated into the *Kpn*I and *Xba*I sites of pTA131 to generate the *pinA* deletion construct, pTA1923. The *trpA* marker digested from pTA298 using *Bam*HI was ligated into the *Bgl*II site of pTA1923 to generate the *pinA::trpA*+ deletion construct, pTA1970.

5.4.1.2 Generation of *pinA* deletion strains

Δ*pinA::trpA+*

The *pinA::trpA+* deletion construct, pTA1970, was introduced into H53 (*ΔpyrE2, ΔtrpA*) to generate the pop-in strain H4248. Strains were grown over three consecutive overnights in YPC to allow the pop-out to occur. Strains were plated on 5-FOA in order to select for the pop-out of pTA1970. Colonies were patched and screened via colony hybridisation using a *pinA+* radiolabelled probe to check for *pinA* deletions. H4249 (*ΔpyrE2, ΔtrpA,* Δ*pinA::trpA+*) was verified via Southern blot (Figure 5.13).

Δ*pinA*

The *trpA+* marker was then removed from H4249 by transformation with the unmarked deletion construct, pTA1923, generating the pop-in strain H4362. Strains were grown over three consecutive overnights in YPC to allow the pop-out to occur. Strains were plated on 5-FOA +Trp in order to select for the pop-out of pTA1923 and the loss of the *trpA+* marker. Colonies were patched and screened via *trp-* phenotype (data not shown) to generate the strain, H4363 (*ΔpyrE2, ΔtrpA,* Δ*pinA*).

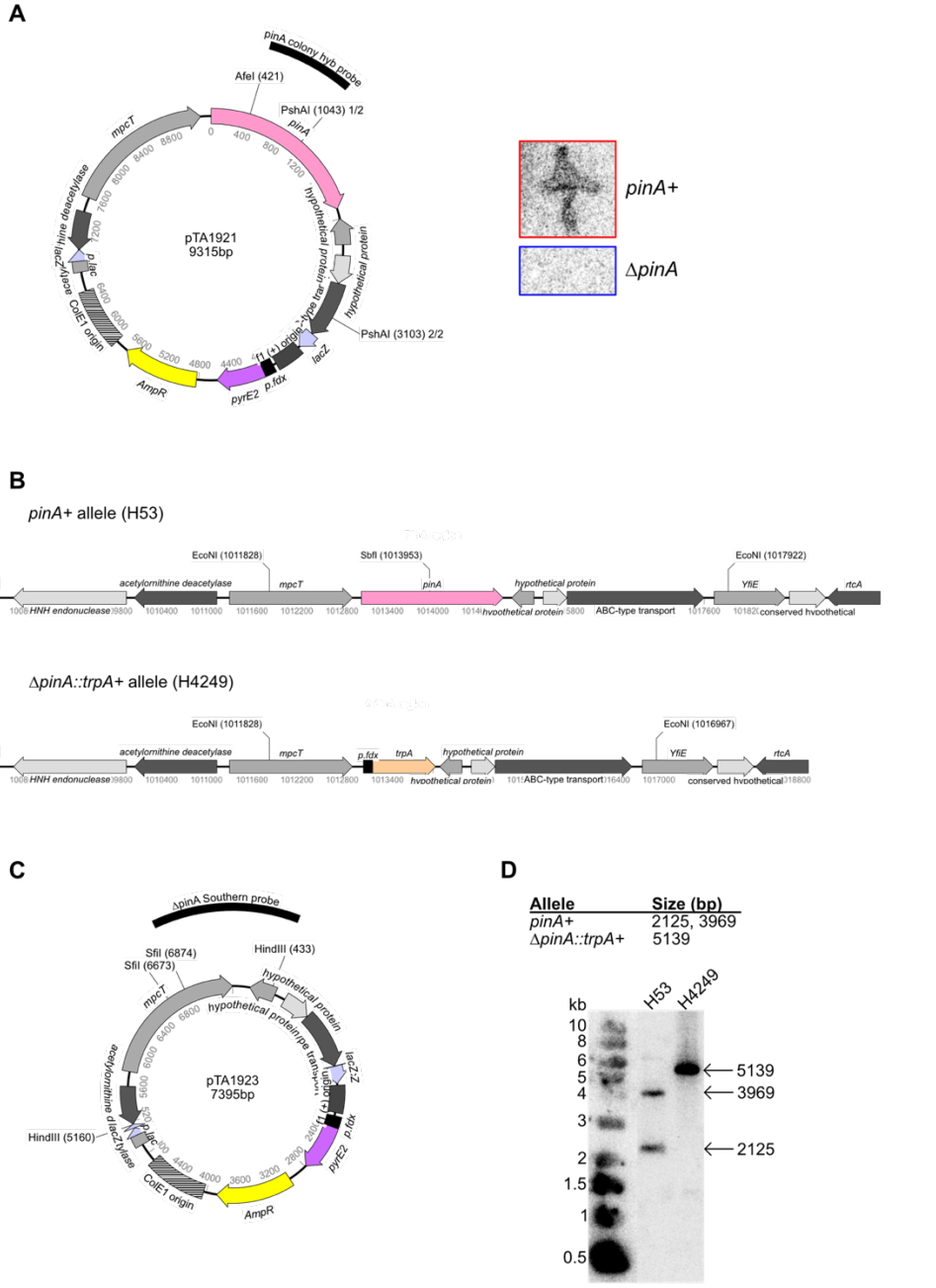


Figure 5.13 **Deletion of *pinA* from H53 (Δ*pyrE2,* Δ*trpA*).** A) The 622bp *pinA*+ probe for colony hybridisation was made by digesting pTA1921 with *Afe*I and *Psh*AI, and the corresponding hyb. B) Allele for *pinA*+ and Δ*pinA::trpA*+. The *Sbf*I and *Eco*NI sites used to digest genomic DNA are shown. C) A 954bp Δ*pinA* probe for southern hybridisation was made by digesting pTA1923 with *Sfi*I and *Hind*III. D) Predicted fragment size and Southern blot showing H4249 is Δ*pinA::trpA*+.

5.4.1.3 Generation of Δ*pinA* in combination with *hel308, hef,* and *hjc* deletions

The order in which genes are deleted in *H. volcanii* often affects the ease of the deletion process and the viability of the strain. Therefore, both orders of gene deletion were attempted to obtain the double deletion strains. However, the deletion of the *pinA* gene was only successful when attempted first.

Δ*pinA,* Δ*hef::trpA+*

The *hef::trpA+* deletion construct, pTA1064, was introduced into H4363 (*ΔpyrE2, ΔtrpA,* Δ*pinA*) to generate the pop-in strain H4440. Strains were grown over three consecutive overnights in YPC to allow the pop-out to occur. Strains were plated on 5-FOA in order to select for the pop-out of pTA1064. Colonies were patched and screened via colony hybridisation using a *hef+* radiolabelled probe to check for *hef* deletions. H4441 (*ΔpyrE2, ΔtrpA,* Δ*pinA,* Δ*hef::trpA+*) was verified via Southern blot (Figure 5.14).

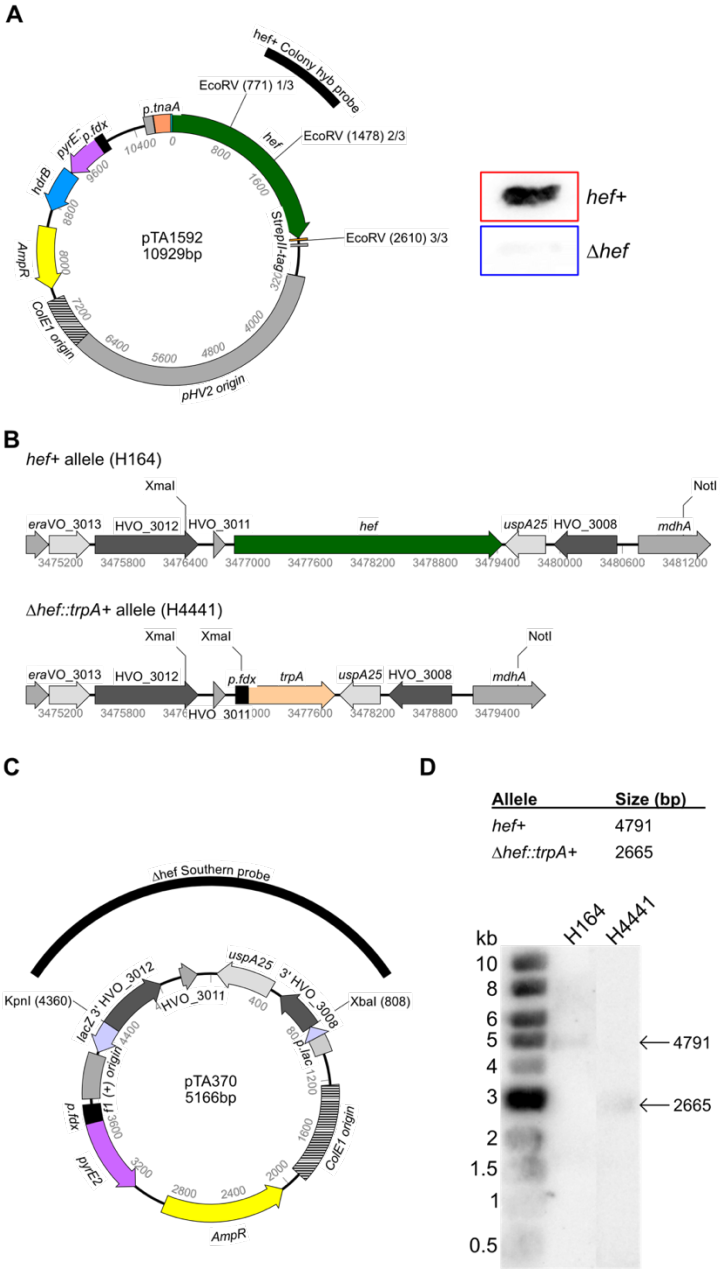


Figure 5.14 **Deletion of *hef* from H4363 (Δ*pyrE2,* Δ*trpA,* Δ*pinA*).** A) The 707bp *hef*+ probe for colony hybridisation was made by digesting pTA1592 with *Eco*RV, and the corresponding hyb. B) Allele for *hef*+ and Δ*hef::trpA*+. The *Xma*I and *Not*I sites used to digest genomic DNA are shown. C) A 1614bp Δ*pinA* probe for southern hybridisation was made by digesting pTA370 with *Kpn*I and *Xba*I. D) Predicted fragment size and Southern blot showing H4441 is Δ*hef::trpA*+.

Δ*pinA,* Δ*hel308::trpA*+

The *hel308::trpA*+ deletion construct, pTA1273, was introduced into H4363 (*ΔpyrE2, ΔtrpA,* Δ*pinA*) to generate the pop-in strain H4442. Strains were grown over three consecutive overnights in YPC to allow the pop-out to occur. Strains

were plated on 5-FOA in order to select for the pop-out of pTA1273. Colonies were patched and screened via colony hybridisation using a *hel308+* radiolabelled probe to check for *hel308* deletions. H4443 (*ΔpyrE2, ΔtrpA, ΔpinA, Δhel308::trpA+*) was verified via Southern blot (Figure 5.15).

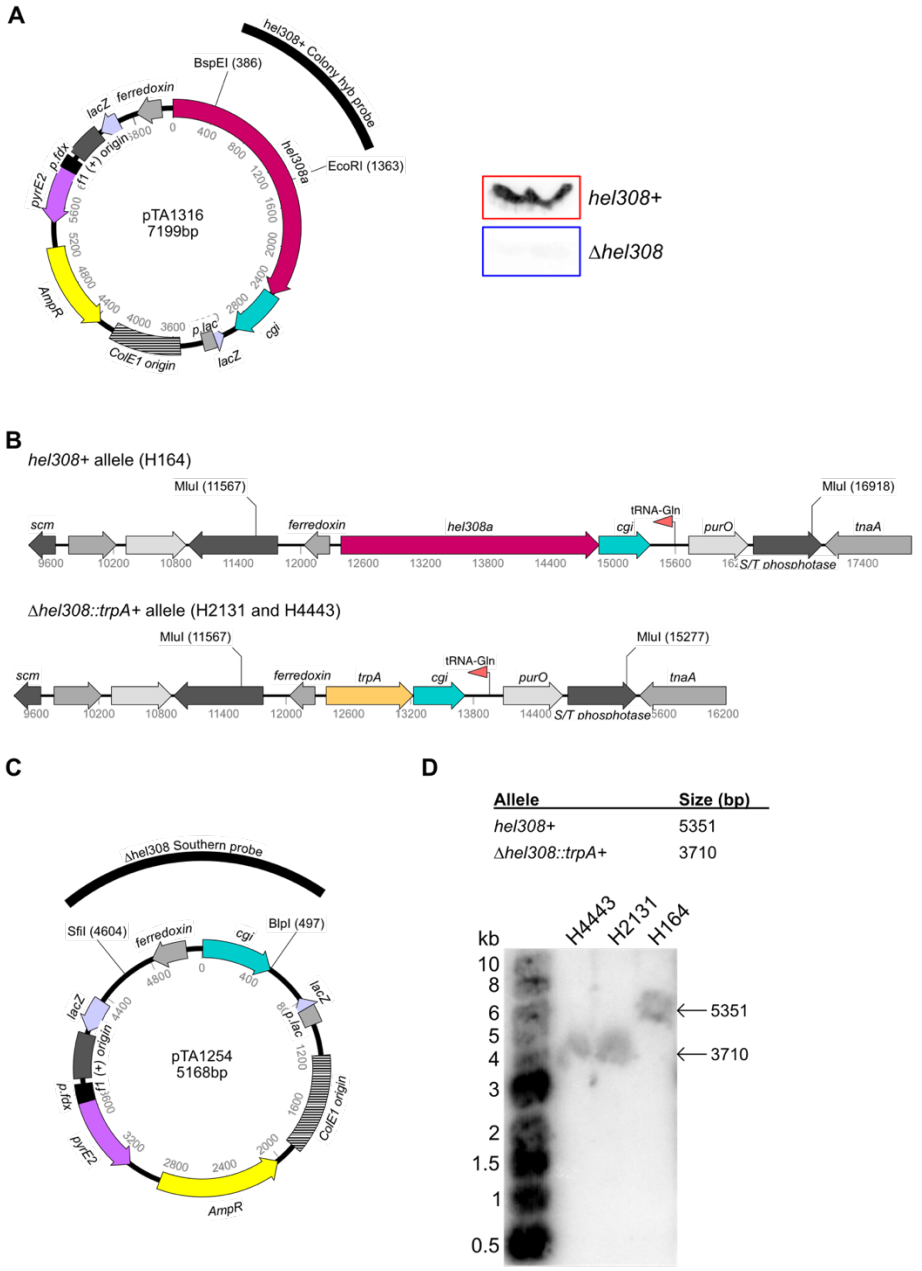


Figure 5.15 **Deletion of *hel308* from H4363 (*ΔpyrE2, ΔtrpA, ΔpinA*).** A) The 977bp *hel308+* probe for colony hybridisation was made by digesting pTA1316 with *Bsp*EI and *Eco*RI, and the corresponding hyb. B) Allele for *hel308+* and *Δhel308::trpA+*. The *Mlu*I sites used to digest genomic DNA are shown. C) A 1061bp *Δhel308* probe for southern hybridisation was made by digesting pTA1254 with *Sfi*I and *Blp*I. D) Predicted fragment size and Southern blot showing H4443 is *Δhel308::trpA+*.

Δ*pinA, Δhjc::trpA+*

The *hjc::trpA+* deletion construct, pTA1062, was introduced into H4363 (*ΔpyrE2, ΔtrpA,* Δ*pinA*) to generate the pop-in strain H4438. Strains were grown over three consecutive overnights in YPC to allow the pop-out to occur. Strains were plated on 5-FOA in order to select for the pop-out of pTA1062. Colonies were patched and screened via colony hybridisation using a *hjc+* radiolabelled probe to check for *hjc* deletions. Out of 40 colonies screened, no *hjc* deletions were found in a H4363 (Δ*pyrE2,* Δ*trpA,* Δ*pinA*) background. Due to time constraints, a Δ*pinA,* Δ*hjc::trpA+* was not generated in this study. (Figure 5.16).

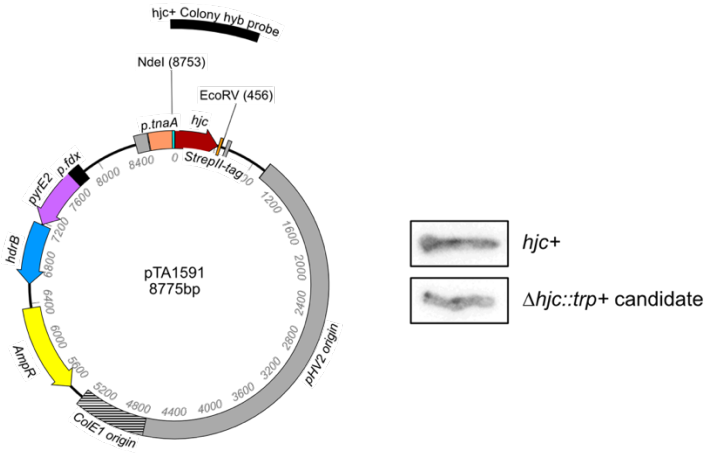


Figure 5.16 **Attempted Deletion of *hjc* from H4363 (Δ*pyrE2,* Δ*trpA,* Δ*pinA*).** The 478bp *hjc+* probe for colony hybridisation was made by digesting pTA1591 with *Eco*RV and *Nde*I, and the corresponding hyb showing that *hjc* is not deleted.

5.4.2 Characterisation of *hel308, hef, hjc*, and *pinA* deleted stains

Growth Rate

Standard growth rates in liquid media were quantified for strains deleted for *hel308, hef, hjc* and *pinA.* Strains were grown over two consecutive overnights to form a vigorously growing culture. The exponential phase of cell growth was calculated using an Epoch2 Microplate Spectrophotometer (BioTek) by measuring the A_{600} of the culture at 15 minutes intervals over a 96 hour time course. A growth curve was plotted and the generation time during exponential phase for each strain was calculated (Figure 5.17). Single and double deletions, where viable, were directly compared within the same experiment.

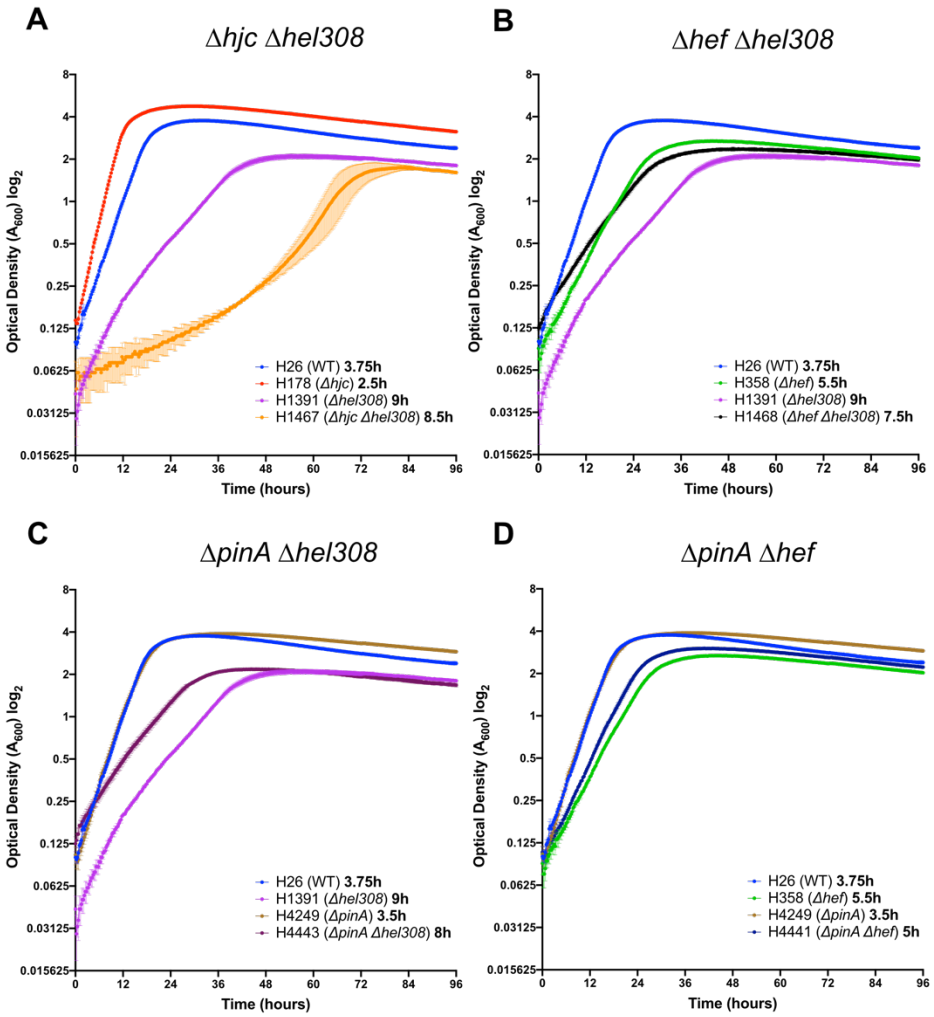


Figure 5.17 **Exponential growth rate of *hjc, hef, pinA* and *hel308* delete strains.** A) H26 (WT), H178 (Δ*hjc*), H1391 (Δ*hel308*), and H1467 (Δ*hjc,* Δ*hel308::trpA*+). B) H26 (WT), H358 (Δ*hef*), H1391 (Δ*hel308*), and H1468 (Δ*hef,* Δ*hel308::trpA*+). C) H26 (WT), H1391 (Δ*hel308*), H4249 (Δ*pinA::trpA*+), and H4443 (Δ*pinA,* Δ*hel308::trpA*+). D) H26 (WT), H358 (Δ*hef*), H4249 (Δ*pinA::trpA*+), and H4441 (Δ*pinA,* Δ*hef::trpA*+). Strains were grown in liquid culture for two consecutive overnights. The strains were then diluted and grown for 96 hours. Growth was measured by A_{600}. Generation time is indicated next to each strain. All strains were incubated on the same 96-well plate and measured simultaneously using an Epoch2 Microplate Spectrophotometer (BioTek). Data points are a mean of two technical repeats within two biological repeats. Standard deviation is shown.

In this experiment the wildtype strain H26 had a slower than expected generation time of 3.75h, demonstrating the importance of making comparisons between sets of strains within the same experiment only. The single gene-deleted strains H178 (Δ*hjc*) and H4249 (Δ*pinA::trpA*+) had slightly decreased generation times compared to H26 (WT) with generation times of 2.5h and 3.5h respectively. The single gene-deleted strain H358 (Δ*hef*) had a slightly increased generation time compared with the wildtype strain, H26, with a generation time of 5.5h. As

expected, the *hel308*-deleted strain, H1391, had a greatly increased generation time compared with the wildtype strain, of 9h. The double gene-deleted strains H1467 (Δ*hjc,* Δ*hel308::trpA*+), H1468 (Δ*hef,* Δ*hel308::trpA*+), and H4443 (Δ*pinA,* Δ*hel308::trpA*+) all had marginally faster generation times compared to the *hel308*-deleted strain, H1391, with generation times of 8.5h, 7.5, and 8h respectively. The double gene-deleted strain H4441 (Δ*pinA,* Δ*hef::trpA*+) had a comparable generation time with the *hef*-single delete strain, H358, with a generation time of 5h.

Survival Following Treatment with DNA-damaging Agents

Strains deleted for *hel308* are sensitive to MMC in an H26 (Δ*pyrE2*) and H164 (*ΔpyrE2, bgaHa-Bb, leuB-Ag1, ΔtrpA*) genetic backgrounds. *hel308*-delete strains in an H164 genetic background are also sensitive to UV irradiation. Therefore it was of interest to determine the phenotype following UV irradiation or treatment with MMC for the combined *hjc, hef, pinA* and *hel308* deleted strains. (Figure 5.18 and 5.19).

There was no observed difference in survival fraction for any strain compared to wildtype H26 following UV irradiation with the exception of H178 (Δ*hjc*), which shows an increase in survival fraction compared to wildtype (Figure 5.18, A). There was no observed difference in survival fraction for H178 (Δ*hjc*), H358 (Δ*hef*), H1468 (Δ*hef,* Δ*hel308::trpA*+), or H4441 (Δ*pinA,* Δ*hef::trpA*+) compared to wildtype H26 following treatment with MMC. As expected, most strains deleted for *hel308,* with the exception of the aforementioned Δ*hef*Δ*hel308* strain, demonstrated a significant decrease in survival fraction following MMC treatment compared to wildtype. In particular the Δ*hjc* Δ*hel308* strain, H1467, showed a decrease in survival compared to the single Δ*hel308* strain. Conversely, H4249 (Δ*pinA::trpA*+) shows an increase in survival fraction following MMC treatment compared to wildtype.

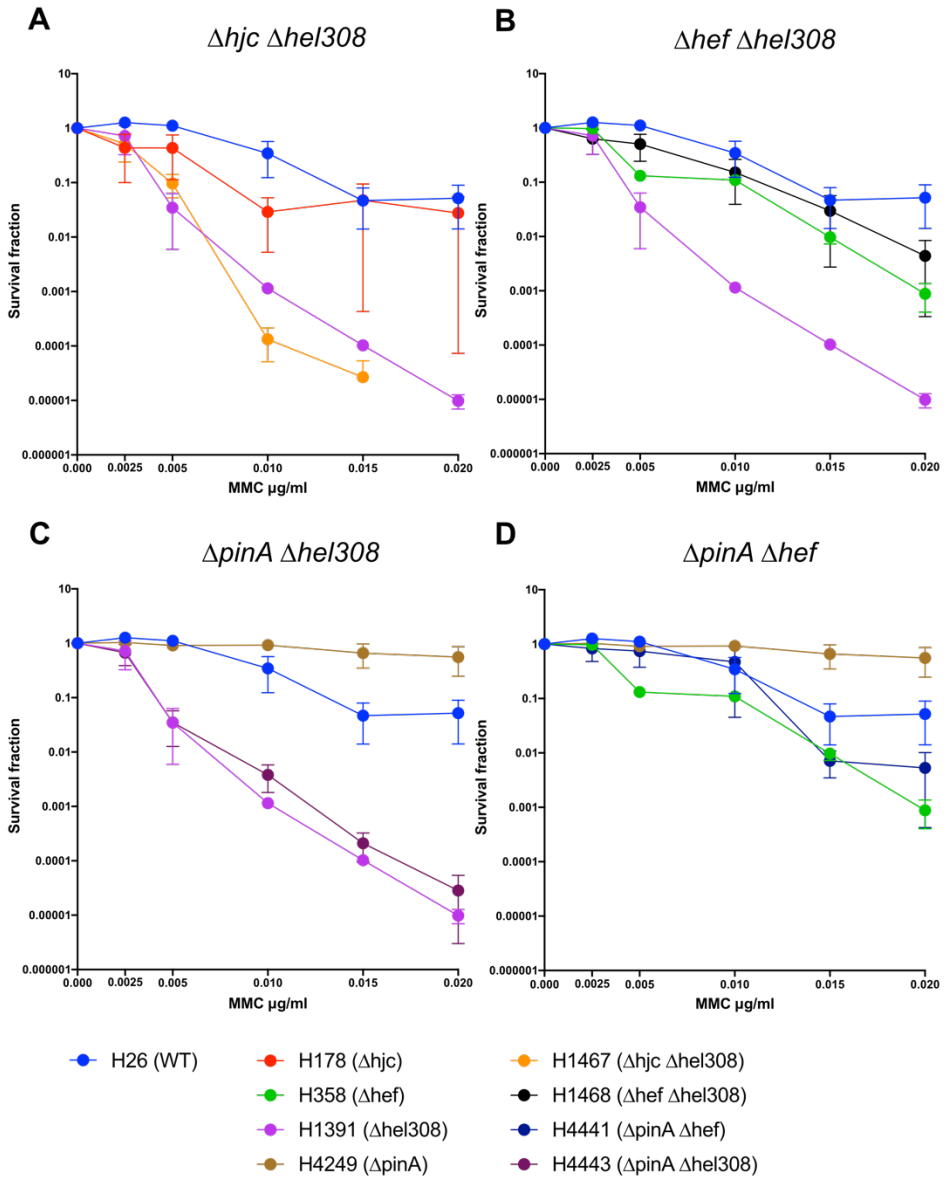


Figure 5.18: **Survival frequency of *hjc, hef, pinA* and *hel308* delete strains following UV irradiation.** A) H26 (WT), H178 (Δ*hjc*), H1391 (Δ*hel308*), and H1467 (Δ*hjc,* Δ*hel308::trpA*+). B) H26 (WT), H358 (Δ*hef*), H1391 (Δ*hel308*), and H1468 (Δ*hef,* Δ*hel308::trpA*+). C) H26 (WT), H1391 (Δ*hel308*), H4249 (Δ*pinA::trpA*+), and H4443 (Δ*pinA,* Δ*hel308::trpA*+). D) H26 (WT), H358 (Δ*hef*), H4249 (Δ*pinA::trpA*+), and H4441 (Δ*pinA,* Δ*hef::trpA*+). No observed difference in survival fraction for any strain compared to wildtype H26 following UV irradiation with the exception of H178 (Δ*hjc*), which shows an increase in survival fraction compared to wildtype. Survival fraction is calculated relative to un-treated control. Each data point is generated as the mean of at least 3 independent trials. Standard deviation is shown.

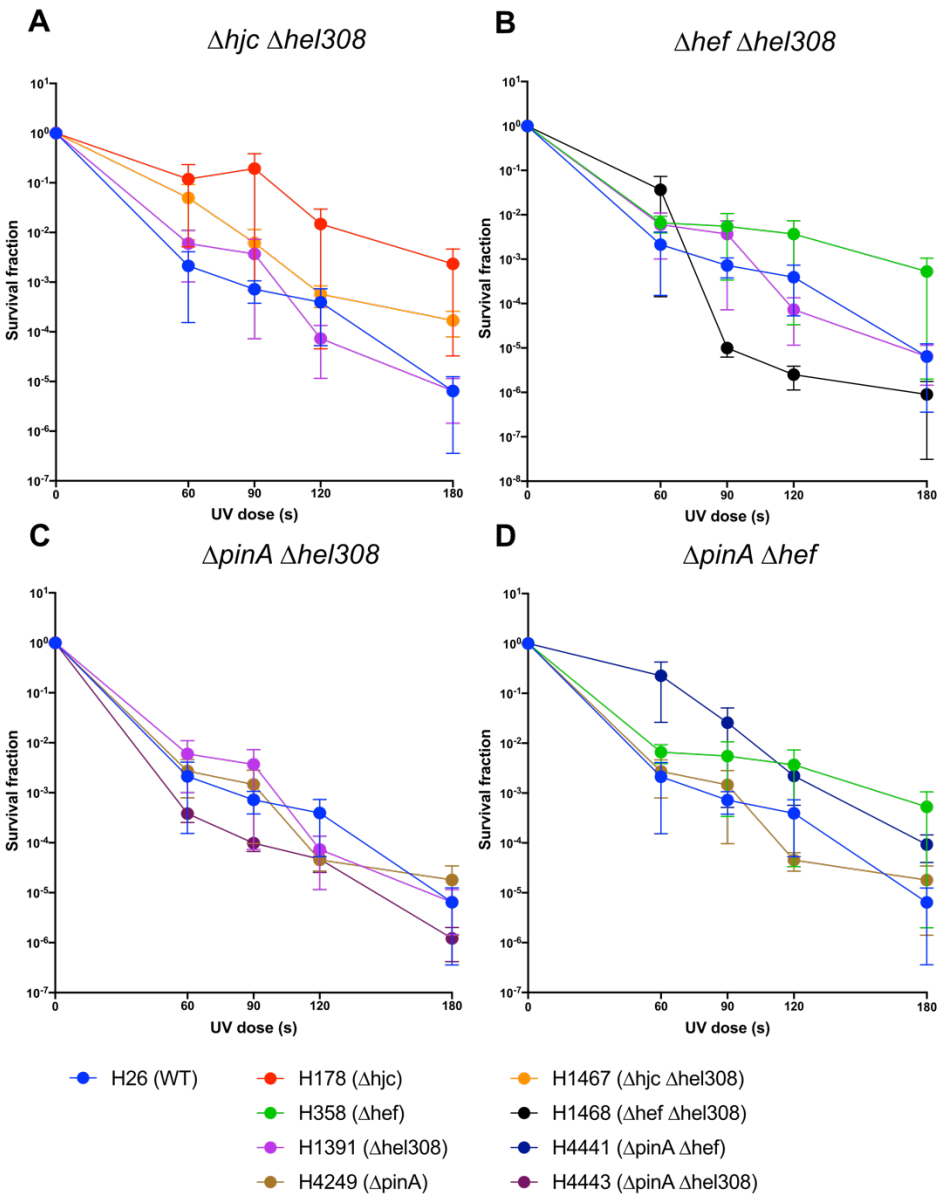


Figure 5.19: **Survival frequency of *hjc, hef, pinA* and *hel308* delete strains following treatment with MMC.** A) H26 (WT), H178 (Δ*hjc*), H1391 (Δ*hel308*), and H1467 (Δ*hjc, Δhel308::trpA+*). B) H26 (WT), H358 (Δ*hef*), H1391 (Δ*hel308*), and H1468 (Δ*hef, Δhel308::trpA+*). C) H26 (WT), H1391 (Δ*hel308*), H4249 (Δ*pinA::trpA+*), and H4443 (Δ*pinA, Δhel308::trpA+*). D) H26 (WT), H358 (Δ*hef*), H4249 (Δ*pinA::trpA+*), and H4441 (Δ*pinA, Δhef::trpA+*). No observed difference in survival fraction for H178 (Δ*hjc*), H358 (Δ*hef*), H1468 (Δ*hef, Δhel308::trpA+*), or H4441 (Δ*pinA, Δhef::trpA+*) compared to wildtype H26 following treatment with MMC. H1391 (Δ*hel308*), H1467 (Δ*hjc, Δhel308::trpA+*), and H4443 (Δ*pinA, Δhel308::trpA+*) demonstrate a significant decrease in survival fraction following MMC treatment compared to wildtype. H4249 (Δ*pinA::trpA+*) shows an increase in survival fraction compared to wildtype. Survival fraction is calculated relative to un-treated control. Each data point is generated as the mean of at least 3 independent trials. Standard deviation is shown.

5.4.3 Generation of *hel308* mutant and *cas3* deleted stains

The *FLAG-cas3* expression construct (used to generate *cas3*+ radiolabelled probe), and Δ*cas3* and Δ*cas3::trpA*+ deletion constructs were kindly provided by Anita Marchfelder. In order to analyse the effect of the *cas3* deletion on recombination levels, all strains were generated in the 'recombination assay' genetic background, H164 (*ΔpyrE2, bgaHa-Bb, leuB-Ag1, ΔtrpA*).

Δ*cas3::trpA+*

The *cas3::trpA*+ deletion construct, pTA131-iCas3KO-trpA, was introduced into H164 (*ΔpyrE2, bgaHa-Bb, leuB-Ag1, ΔtrpA*) to generate the pop-in strain H4380. Strains were grown over three consecutive overnights in YPC to allow the pop-out to occur. Strains were plated on 5-FOA in order to select for the pop-out of pTA131-iCas3KO-trpA. Colonies were patched and screened via colony hybridisation using a *cas3*+ radiolabelled probe to check for *cas3* deletions. H4386 was verified via Southern blot (Figure 5.20).

hel308-F316A, Δ*cas3::trpA+*

The *cas3::trpA*+ deletion construct, pTA131-iCas3KO-trpA, was introduced into H2397 (*ΔpyrE2, bgaHa-Bb, leuB-Ag1, ΔtrpA, hel308-F316A*) to generate the pop-in strain H4382. Strains were grown over three consecutive overnights in YPC to allow the pop-out to occur. Strains were plated on 5-FOA in order to select for the pop-out of pTA131-iCas3KO-trpA. Colonies were patched and screened via colony hybridisation using a *cas3*+ radiolabelled probe to check for *cas3* deletions. H4388 was verified via Southern blot (Figure 5.21).

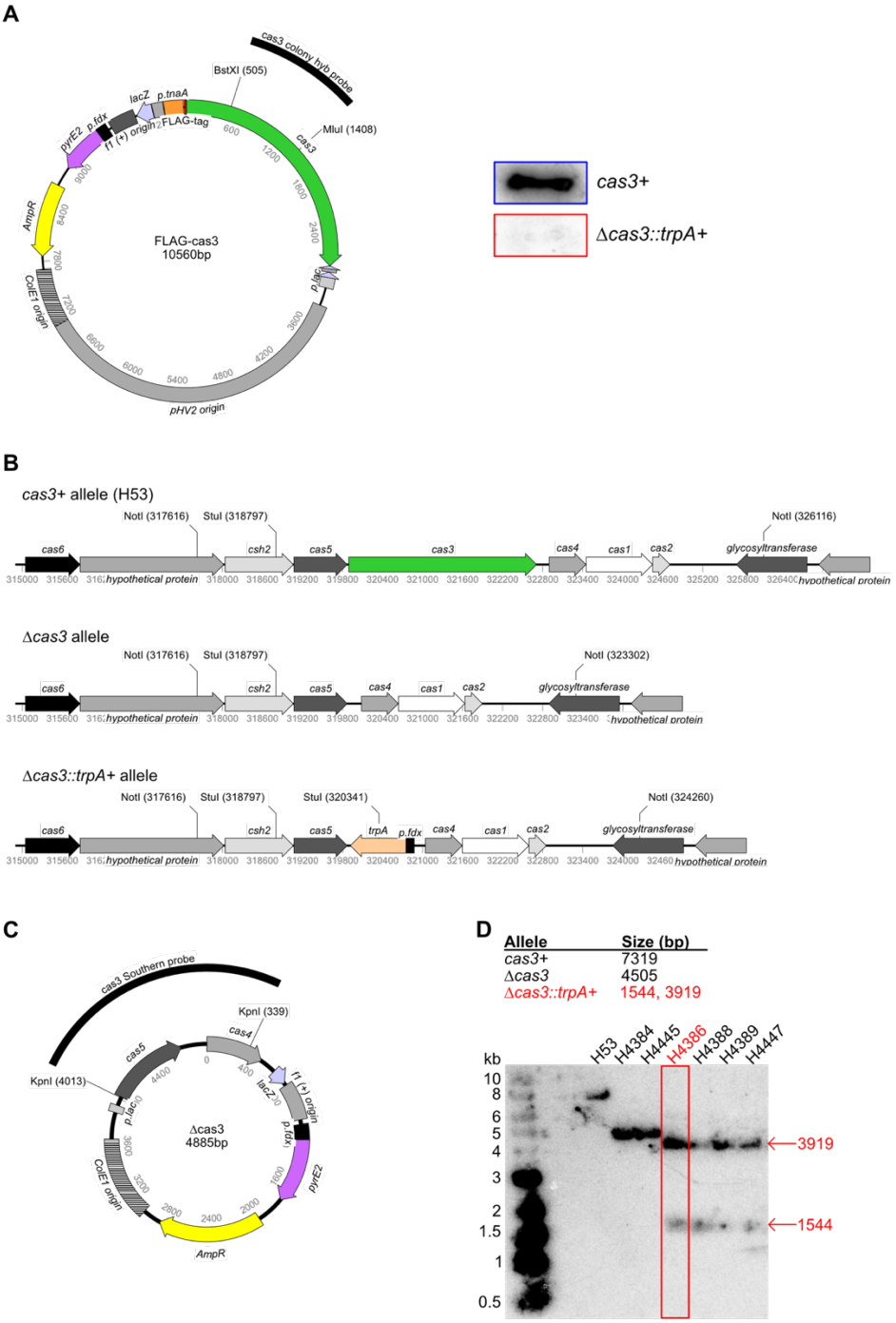


Figure 5.20 **Deletion of *cas3* from H164 (Δ*pyrE2, bgaHa-Bb, leuB-Ag1,* Δ*trpA*).** A) The 903 bp *cas3*+ probe for colony hybridisation was made by digesting pTA927-cas3FLAGN with *Bst*XI and *Mlu*I, and the corresponding hyb. B) Allele for *cas3*+, Δ*cas3* and Δ*cas3::trpA*+. The *Not*I and *Stu*I sites used to digest genomic DNA are shown. C) A 1211 bp Δ*cas3* probe for southern hybridisation was made by digesting p131-Cas3up/do with *Kpn*I. D) Predicted fragment size and Southern blot, highlighted and boxed in red, showing H4386 is Δ*cas3::trpA*+.

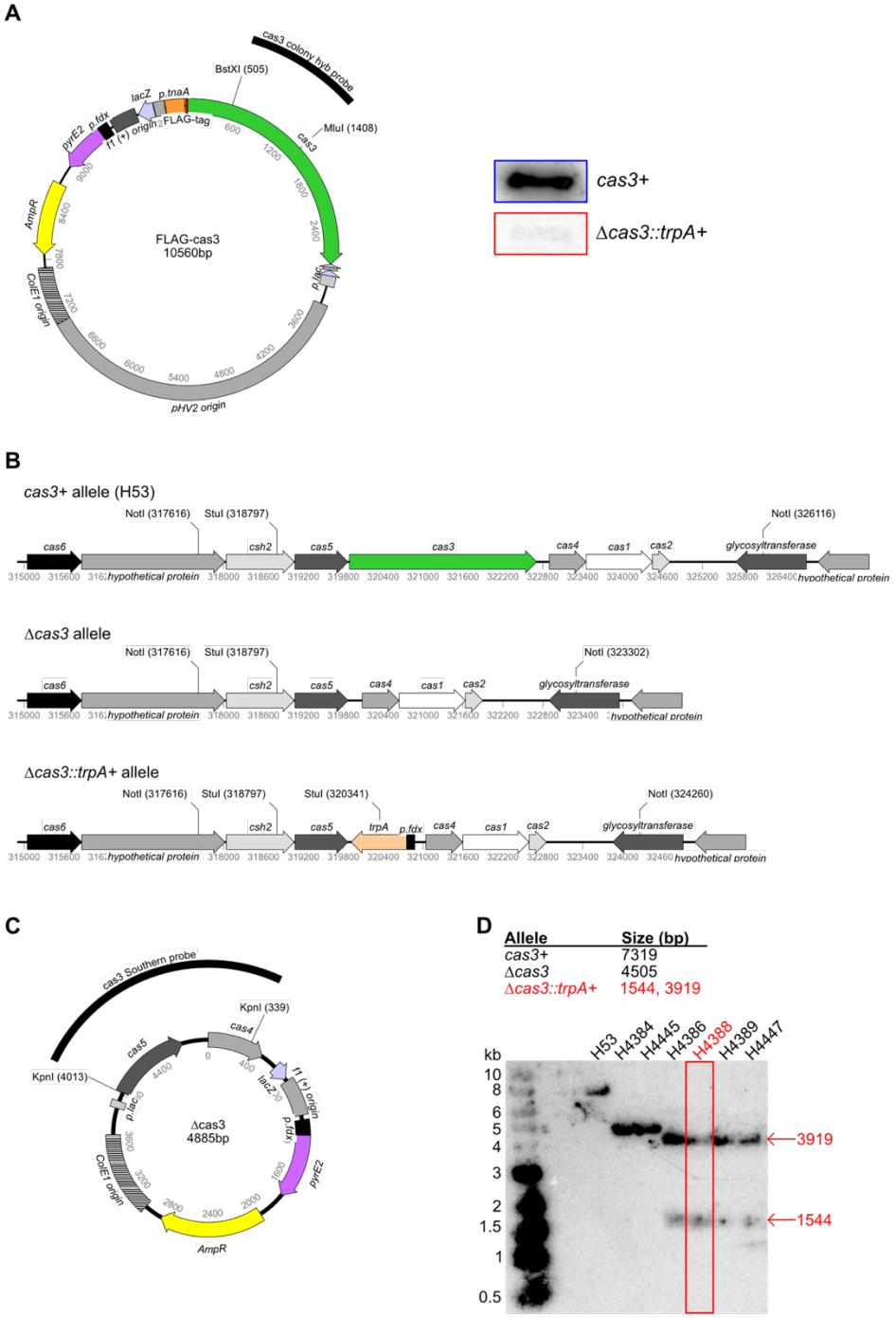


Figure 5.21 **Deletion of *cas3* from H2397 (Δ*pyrE2, bgaHa-Bb, leuB-Ag1*, Δ*trpA, hel308-F316A*).** A) The 903 bp *cas3*+ probe for colony hybridisation was made by digesting pTA927-cas3FLAGN with *Bst*XI and *Mlu*I, and the corresponding hyb. B) Allele for *cas3*+, Δ*cas3* and Δ*cas3::trpA*+. The *Not*I and *Stu*I sites used to digest genomic DNA are shown. C) A 1211 bp Δ*cas3* probe for southern hybridisation was made by digesting p131-Cas3up/do with *Kpn*I. D) Predicted fragment size and Southern blot, highlighted and boxed in red, showing H4388 is Δ*cas3::trpA*+.

hel308-R743A, Δcas3::trpA+

The *cas3::trpA+* deletion construct, pTA131-iCas3KO-trpA, was introduced into H2398 (*ΔpyrE2, bgaHa-Bb, leuB-Ag1, ΔtrpA, hel308-R743A*) to generate the pop-in strain H4383. Strains were grown over three consecutive overnights in YPC to allow the pop-out to occur. Strains were plated on 5-FOA in order to select for the pop-out of pTA131-iCas3KO-trpA. Colonies were patched and screened via colony hybridisation using a *cas3+* radiolabelled probe to check for *cas3* deletions. H4389 was verified via Southern blot (Figure 5.22).

Δhel308::trpA+, Δcas3

The *cas3* deletion construct, p131-Cas3up/do, was introduced into H2117 (*ΔpyrE2, bgaHa-Bb, leuB-Ag1, ΔtrpA, Δhel308*) to generate the pop-in strain H4444. Strains were grown over three consecutive overnights in YPC to allow the pop-out to occur. Strains were plated on 5-FOA in order to select for the pop-out of p131-Cas3up/do. Colonies were patched and screened via colony hybridisation using a *cas3+* radiolabelled probe to check for *cas3* deletions. H4445 was verified via Southern blot (Figure 5.23).

hel308-D145N, Δcas3::trpA+

The *cas3::trpA+* deletion construct, pTA131-iCas3KO-trpA, was introduced into H2400 (*ΔpyrE2, bgaHa-Bb, leuB-Ag1, ΔtrpA, hel308-D145N*) to generate the pop-in strain H4446. Strains were grown over three consecutive overnights in YPC to allow the pop-out to occur. Strains were plated on 5-FOA in order to select for the pop-out of pTA131-iCas3KO-trpA. Colonies were patched and screened via colony hybridisation using a *cas3+* radiolabelled probe to check for *cas3* deletions. H4447 was verified via Southern blot (Figure 5.24).

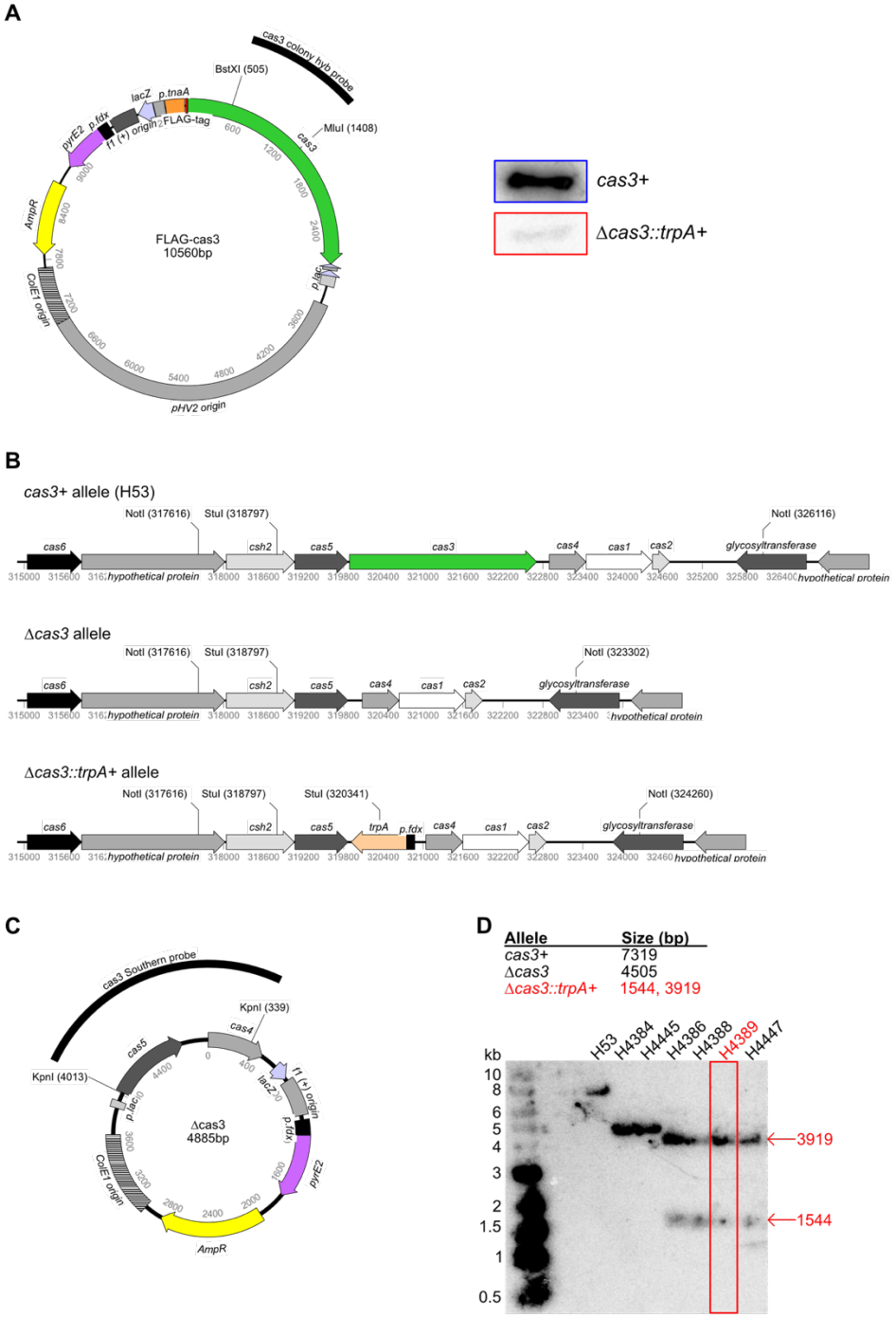


Figure 5.22 **Deletion of *cas3* from H2398 (Δ*pyrE2, bgaHa-Bb, leuB-Ag1*, Δ*trpA, hel308-R743A*).** A) The 903 bp *cas3*+ probe for colony hybridisation was made by digesting pTA927-cas3FLAGN with *Bst*XI and *Mlu*I, and the corresponding hyb. B) Allele for *cas3*+, Δ*cas3* and Δ*cas3::trpA*+. The *Not*I and *Stu*I sites used to digest genomic DNA are shown. C) A 1211 bp Δ*cas3* probe for southern hybridisation was made by digesting p131-Cas3up/do with *Kpn*I. D) Predicted fragment size and Southern blot, highlighted and boxed in red, showing H4389 is Δ*cas3::trpA*+.

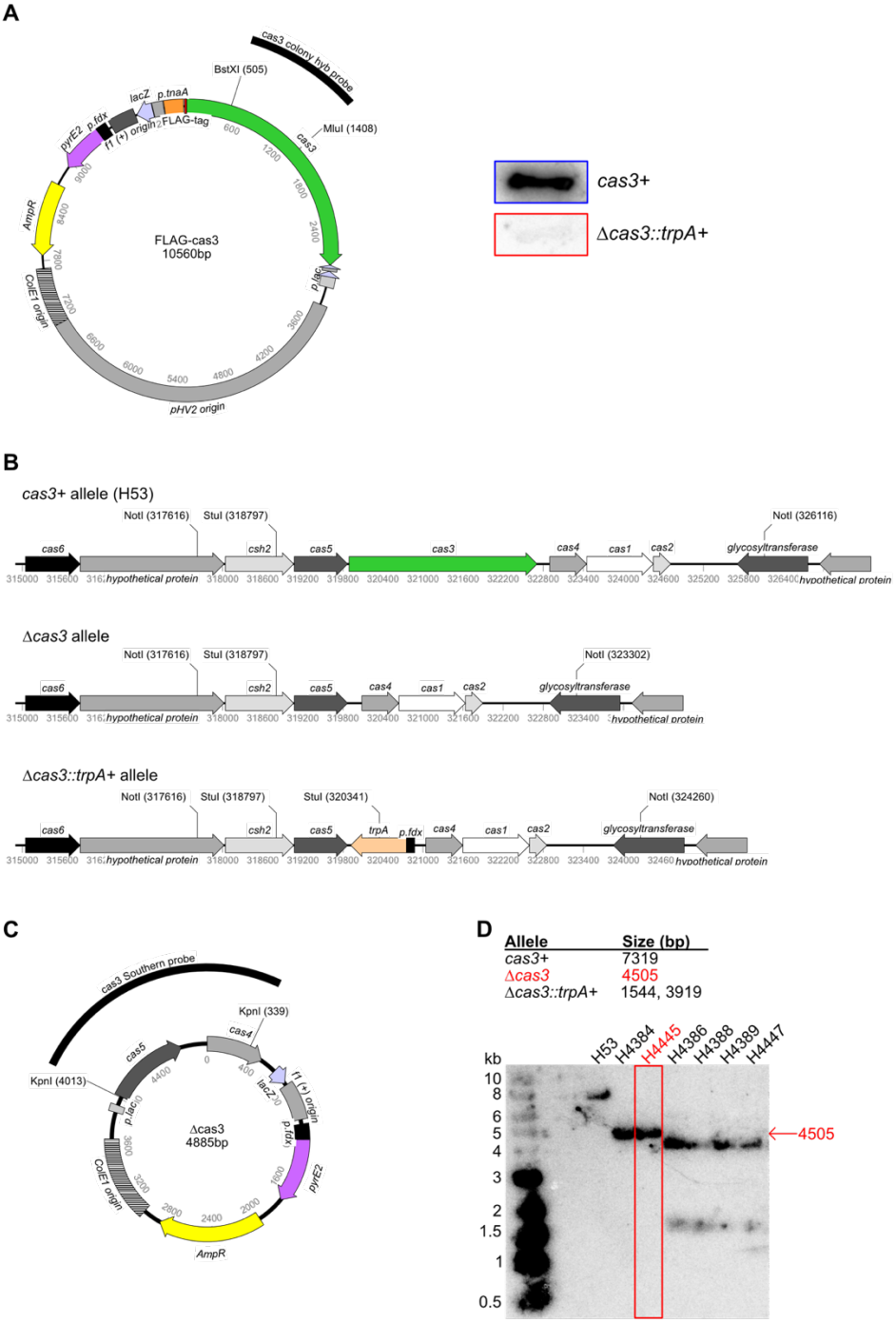


Figure 5.23 **Deletion of *cas3* from H2117 (Δ*pyrE2, bgaHa-Bb, leuB-Ag1*, Δ*trpA*, Δ*hel308*).** A) The 903 bp *cas3*+ probe for colony hybridisation was made by digesting pTA927-cas3FLAGN with *Bst*XI and *Mlu*I, and the corresponding hyb. B) Allele for *cas3*+, Δ*cas3* and Δ*cas3::trpA*+. The *Not*I and *Stu*I sites used to digest genomic DNA are shown. C) A 1211 bp Δ*cas3* probe for southern hybridisation was made by digesting p131-Cas3up/do with *Kpn*I. D) Predicted fragment size and Southern blot, highlighted and boxed in red, showing H4445 is Δ*cas3*.

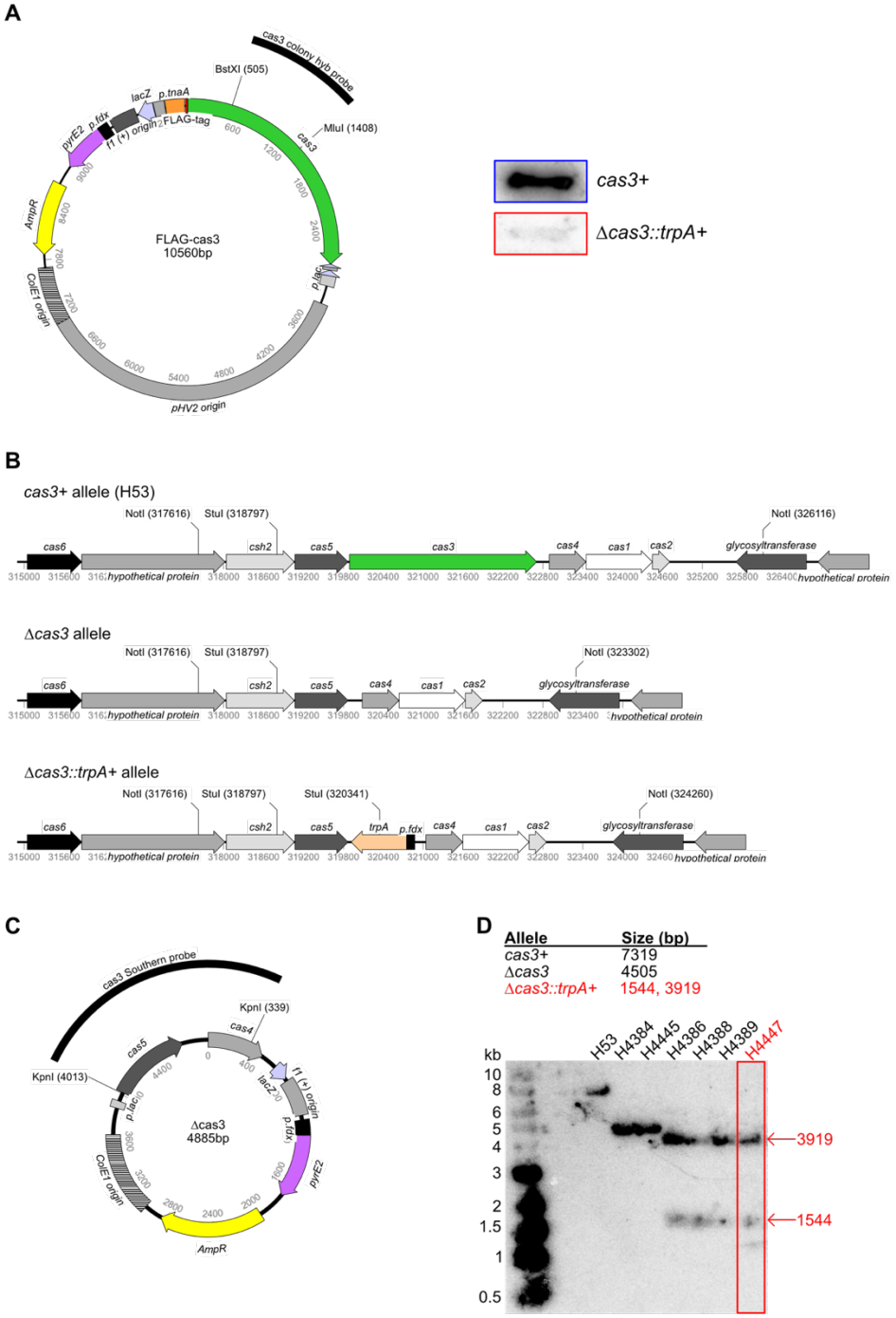


Figure 5.24 **Deletion of *cas3* from H2400 (Δ*pyrE2, bgaHa-Bb, leuB-Ag1*, Δ*trpA, hel308-D145N*).** A) The 903 bp *cas3*+ probe for colony hybridisation was made by digesting pTA927-cas3FLAGN with *Bst*XI and *Mlu*I, and the corresponding hyb. B) Allele for *cas3*+, Δ*cas3* and Δ*cas3::trpA*+. The *Not*I and *Stu*I sites used to digest genomic DNA are shown. C) A 1211 bp Δ*cas3* probe for southern hybridisation was made by digesting p131-Cas3up/do with *Kpn*I. D) Predicted fragment size and Southern blot, highlighted and boxed in red, showing H4447 is Δ*cas3::trpA*+.

5.4.4 Characterisation of *hel308* mutant and *cas3* deleted stains

Growth Rate

The standard growth rates, in liquid media, were quantified for strains deleted for *cas3* in combination with Δ*hel308, hel308-D145N, hel308-F316A,* and *hel308-R743A.* Strains were grown over two consecutive overnights to form a vigorously growing culture. The exponential phase of cell growth was calculated using an Epoch2 Microplate Spectrophotometer (BioTek) by measuring the A_{600} of the culture at 15 minutes intervals over a 72 hour time course. A growth curve was plotted and the generation time during exponential phase for each strain was calculated (Figure 5.25). Single and double mutants were directly compared within the same experiment.

In this experiment, the wildtype strain H164 had a generation time of 2.75h. The generation times can be separated into two groups, wildtype-like and Δ*hel308*-like. The Hel308-F316A, Hel308-R743A, and Δ*cas3* strains group with the wildtype strain with comparable generation times of 2.75h, 3h, and 2.75h respectively. Similarly, the combined mutant strains H4388 (Δ*cas3, hel308-F316A*) and H4389 (Δ*cas3, hel308-F316A*) group with the wildtype strain with comparable generation times of 3h and 2.75h respectively. However, all strains containing the D145N point mutation group with the Δ*hel308* strain. H2117 (Δ*hel308*) has a generation time of 5.75h. H2400 (*hel308-D145N*), H4445 (Δ*cas3,* Δ*hel308*), and H4447 (Δ*cas3, hel308-D145N*) have the same or marginally faster generation times of 5.75h, 4.75h, and 3.75h respectively.

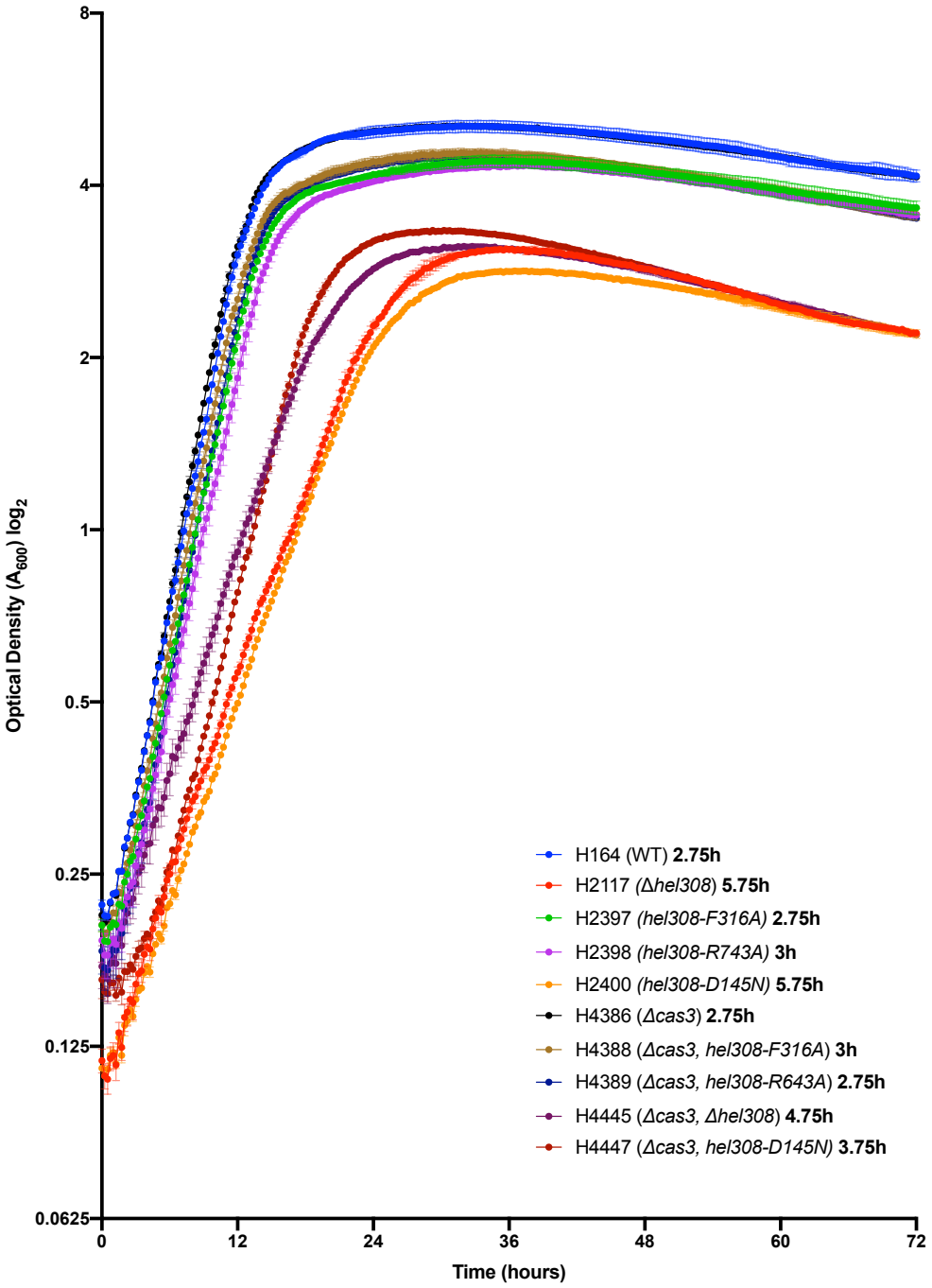


Figure 5.25 **Exponential growth rate of ATPase-null and hyper-recombinant mutant *hel308* strains, and Δ*cas3* strains.** H164 (*hel308*+), H2117 (Δ*hel308*), H2397 (*hel308-F316A*), H2398 (*hel308-R743A*), H2400 (*hel308-D145N*), H4386 (Δ*cas3*), H4388 (Δ*cas3, hel308-F316A*), H4389 (Δ*cas3, hel308-R743A*), H4445 (Δ*cas3,* Δ*hel308*), and H4447 (Δ*cas3, hel308-D145N*), were grown in liquid culture for two consecutive overnights. The strains were then diluted and grown for 48 hours. Growth was measured by A_{600}. Generation time is indicated next to each strain. Boxes indicate strain groups, wildtype-like (blue) and Δ*hel308*-like (red). All strains were incubated on the same 96-well plate and measured simultaneously using an Epoch2 Microplate Spectrophotometer (BioTek). Data points are a mean of two technical repeats within two biological repeats. Standard deviation is shown.

Recombination frequency

A hyper-recombinant phenotype is an increase in the levels of recombination compared to wild-type. Strains deleted for *cas3* have demonstrated an increase in recombination (personal correspondence, Uri Gophna) therefore, it is of interest to analyse the recombination levels in strains deleted for *cas3* in combination with the hyper-recombinant *hel308* mutants, ATPase-null *hel308* mutant, and Δ*hel308* (See chapter 3: Separation of function of Hel308, for more detail) (Table 5.4).

Strain	**H164**	**H2117**	**H2397**	**H2398**	**H2400**
	hel308+	*Δhel308*	*hel308-F316A*	*hel308-R743A*	*hel308-D145N*
Relative RF normalised by TE	$1.43x10^{-2}$	$7.85x10^{-2}$	$2.37x10^{3}$	$5.07x10^{2}$	$7.26x10^{-3}$
	1	**5.5x**	**166000x**	**35500x**	**0.51x**
CO fraction	16.5% (200)	21.9% (64)	<0.5% (200)	<0.5% (200)	35.7% (14)
NCO fraction	83.5% (200)	78.1% (64)	>99.5% (200)	>99.5% (200)	64.3% (14)

Strain	**H4386**	**H4445**	**H4388**	**H4389**	**H4447**
	Δcas3	*Δcas3, Δhel308*	*Δcas3, hel308-F316A*	*Δcas3, hel308-R743A*	*Δcas3, hel308-D145N*
Recombination Frequency (RF)	$7.25x10^{-4}$ ($+/-3.54x10^{-4}$)	$5.94x10^{-4}$ ($+/-2.72x10^{-4}$)	$9.29x10^{0}$ ($+/-6.18x10^{0}$)	$8.08x10^{0}$ ($+/-3.25x10^{0}$)	$7.94x10^{-5}$ ($+/-2.74x10^{-5}$)
Transformation Efficiency (TE)	$4.49x10^{-3}$	$1.48x10^{-2}$	$4.91x10^{-3}$	$1.99x10^{-2}$	$2.66x10^{-1}$
Relative RF normalised by TE	$1.61x10^{-1}$	$4.00x10^{-2}$	$1.89x10^{3}$	$4.07x10^{2}$	$2.98x10^{-4}$
	1.1x	**2.8x**	**132000x**	**28500x**	**0.02x**
CO fraction	21% (200)	<6% (17)	<0.5% (200)	<0.5% (200)	2.6% (38)
NCO fraction	79% (200)	>94% (17)	>99.5% (200)	>99.5% (200)	97.4% (38)

Table 5.4: **Recombination frequencies of Δ*cas3* and mutant *hel308* strains.** Top) Summary of recombination frequencies of WT, Δ*hel308, hel308-D145N, hel308-F316A* and *hel308-R743A* strains. Bottom) Recombination frequencies of Δ*cas3* strains. Values in bold indicate the amount of recombination relative to wildtype H164 (*hel308+*). Values are generated as an average of 5 independent trials, +/- standard error is shown in brackets. Blue shaded cells indicate a recombination defect, whilst red indicates hyper-recombination. Proportion of CO and NCO events are represented as a percentage, with the number of assayed colonies indicated in brackets underneath. Significant differences from the wildtype (P=0.05) are shaded in blue for a decrease and red for an increase.

H4386 (Δ*cas3*) has a very small increase in the recombination levels at 1.1x that of the wildtype H164. Crossover recombination events were marginally higher than wild type. H4445 (Δ*cas3, Δhel308*) has a hyper-recombinant phenotype with recombination levels 2.8x that of wild type H164 (*cas3+, hel308+*). This data is comparable with the *hel308*-delete strain, H2117, with a recombination

level of 5.5x that of wildtype. Unlike the H2117 strain, no crossover events were observed for H4445 and non-crossover events were significantly higher than wild type. Both H4388 (Δ*cas3, hel308-F316A*) and H4389 (Δ*cas3, hel308-R743A*) have an extreme hyper-recombinant phenotype with recombination levels of 132000x and 28500x respectively. No crossover events were observed for these strains and non-crossover events were significantly higher than wild type. These data are comparable to the H2397 (*hel308-F316A*) and H2398 (*hel308-R743A*) strains. A decrease in the levels of recombination is observed for H4447 (Δ*cas3, hel308-D145N*), at 0.02x that of wild type. Crossover recombination events were decreased compared with wild type. These data suggest that *cas3* does not play a role in the regulation of homologous recombination.

5.5 Discussion

Hel308 may act in the Hjc/PINA pathway.

Deletion of Hel308 in combination with Hef and Hjc, and the helicase PINA, has no significant effect on generation time compared to the single *hel308* strain. The lack of a synthetic defect suggests that Hel308 does not function in the later stages of homologous recombination alongside Hef and Hjc/PINA.

Following treatment with MMC, the Δ*pinA*Δ*hel308* strain shows survival similar to the single Δ*hel308* strain whereas the Δ*hel308*Δ*hjc* strain shows a decrease in survival. Conversely, the Δ*hel308* Δ*hef* strain shows survival similar to the single Δ*hef* strain. These data suggest that Hel308 does not act alongside Hef during the repair of interstrand cross links in *H. volcanii;* however, Hel308 may act in the Hjc/PINA pathway. This proposal is supported by yeast two-hybrid analyses in *Sulfolobus tokodaii* revealing that the Hel308 homologue Hjm physically interacts with Hjc (Li *et al.*, 2008).

Interestingly, an Δ*hjc* Δ*pinA* double deleted strain could not be generated during this study. PINA physically and functionally interacts with Hjc to promote Holliday junction migration. Furthermore, *in vitro* biochemical analysis of the *S. islandicus* homologues indicated that Hjc, Hjm and PINA coordinate Holliday junction migration and cleavage in a concerted manner. These data suggest that the Hel308/Hjm family helicases, in association with Hjc endonucleases and PINA, are involved in the processing of stalled replication forks (Li *et al.*, 2008; Zhai *et al.*, 2018). Further investigation into this relationship in *H. volcanii* is required.

Cas3 does not regulate homologous recombination

The *cas3*-deleted strain has a very slight increase in the recombination levels at 1.1x that of the wildtype H164, but this is not significant. Crossover recombination events in this strain were marginally higher than wild type, but again not significant. Similarly, deletion of *cas3* in combination with Δ*hel308* or *hel308*-point mutants does not increase the levels of recombination; if anything, a decrease in recombination levels is observed. Interestingly, no or very few crossover events were observed in these strains and non-crossover events were significantly higher than wild type. These data contradict the proposal that the absence of *cas3* significantly increases the overall level of recombination in the cell. Deletion of *cas3* in combination with Δ*hel308* or hyper-recombinant *hel308*-mutants has no effect on generation time. Therefore, *cas3* does not appear to play a role in the regulation of homologous recombination in *H. volcanii*.

5.6 Future Perspectives

Hjc, PINA, and Hef

To further investigate the relationship between the *hel308, hjc,* and *pinA* genes and the possible role that Hel308 plays in the Hjc/PINA pathway, it would be of interest to reattempt the double deletion of *hjc* and *pinA*, and to attempt the triple deletion of *hel308, hjc,* and *pinA.* The Δ*hjc*Δ*hef* is known to be synthetically lethal therefore, it would also be of interest to attempt the triple deletions of *hjc, hef,* and *pinA;* and *hel308, hef,* and *pinA.* Furthermore, synthetic lethal combinations should be analysed via complementation assays.

Lhr and AshA

Hel308a is an SF2 helicase. The SF2 is one of the largest superfamily of helicases, members of which participate in almost all steps of DNA and RNA metabolism. Phylogenetic analysis of the distribution of SF2 helicase reveals that Lhr is the most abundant helicase in archaea with Hel308/Hjm as the second most abundant member of the SF2 family (Chamieh, Ibrahim and Kozah, 2015). Lhr is a 3′–5′ translocase able to unwind both RNA:DNA hybrids and DNA:DNA duplex (Fairman-Williams, Guenther and Jankowsky, 2010; Ordonez and Shuman, 2013). There are two Lhr proteins encoded in *H. volcanii*, Lhr2, which is counterintuitively found in the Lhr-1 clade, and Lhr1, which is similarly found in the Lhr-2 clade (Figure 5.26, adapted from Chamieh, Ibrahim and Kozah, 2015).

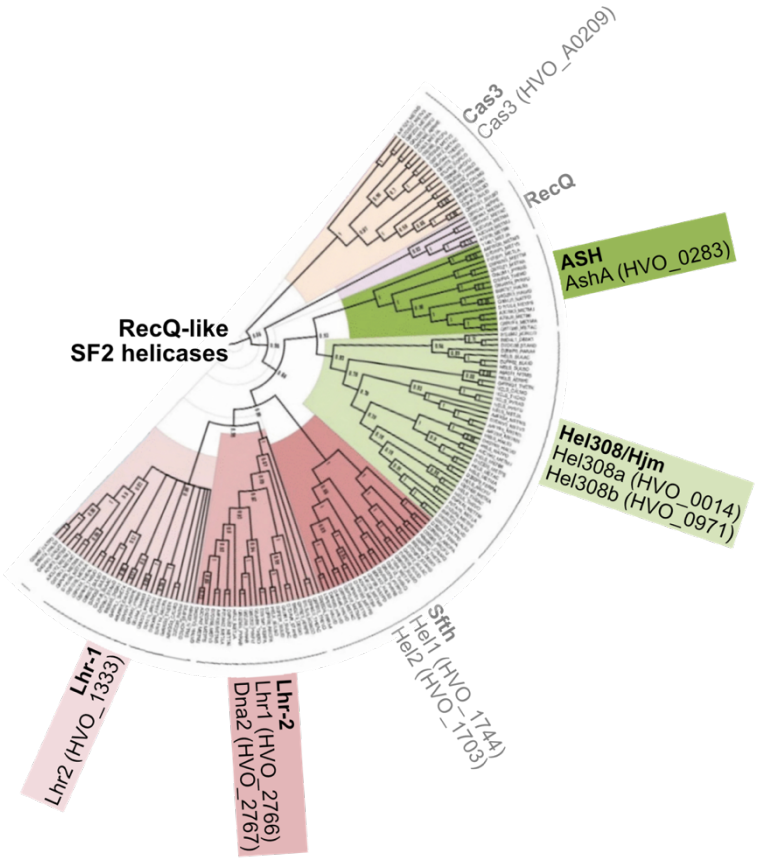


Figure 5.26: **Distribution of Archaeal the SF2 helicases Lhr and AshA in relation to Hel308.** Phylogenetic distribution of the archaeal RecQ-like helicases. *Haloferax volcanii* genes are annotated next to the clade in which they are found. Lhr, ASH and Hel308/Hjm helicases are highlighted. Figure adapted from Chamieh, Ibrahim and Kozah, 2015.

A second helicase, Dna2, is found in the Lhr-2 clade alongside Lhr1. Archaea also contain a large group of DEAD/DEAH helicases known as the ASH helicase (Archaeal Specific Helicase, AshA in *H. volcanii*). Similarly, ASH is an SF2 helicase. ASH evolved in Euryarchaeota and is evolutionary related to Hel308/Hjm from the Ski2-like family. It is probable that Hel308a has a similar role to Lhr or AshA, or that either protein is redundant. Redundancy between two or more proteins functions is not uncommon in *H. volcanii*, and identifying such redundancies gives insights into the role of the proteins. The single deletion strains *hel308*, *lhr1, lhr2* and *dna2,* and the double deletion strains *lhr1* and *dna2,* and *lhr1* and *lhr2* were generated prior to the commencement of this study. The *hel308::trpA+* deletion construct was introduced into all of these genetic backgrounds by Hannah Smallwood (MSci student) under my supervision, generating the strains H3690 (Δ*pyrE2,* Δ*trpA,* Δ*dna2,* Δ*hel308::trpA+*), H3691 (Δ*pyrE2,* Δ*trpA,* Δ*lhr1,* Δ*lhr2,* Δ*hel308::trpA+*), and H3692 (Δ*pyrE2,* Δ*trpA,* Δ*dna2 lhr1,* Δ*hel308::trpA+*). However, these strains are yet to be characterised. Hannah also generated the pop-in strains H3251 (Δ*pyrE2,* Δ*trpA,* Δ*lhr1,*

hel308+::[Δhel308::trpA+, pyrE2+]) and H3253 (Δ*pyrE2,* Δ*trpA,* Δ*lhr2, hel308+::[Δhel308::trpA+, pyrE2+]*), but pop-out strains are yet to be obtained. It would be of interest to introduce the Δ*ashA* into these genetic backgrounds and to characterise all viable strains in order to determine any redundancy present.

Hel308b

In addition to the canonical Hel308, a second *hel308* gene known as *hel308b* (HVO_0971) is present in *Haloferax volcanii*. Like Lhr and AshA, Hel308b is an SF2 helicase (Figure 5.26). Hel308b lacks domain 5 found in Hel308a, which is key in the regulation of ATP hydrolysis. Mutants deleted for *hel308b* show no sensitivity to DNA-crosslinking agents unlike the *Δhel308* mutant (Gamble-Milner, 2016). However, similar to the *Δhel308* mutant, the *Δhel308b* mutant does exhibit a recombination defect compared to the wild type (Gamble-Milner, 2016). This suggests that Hel308b may have a regulatory role in homologous recombination. The lack of domain 5 implies that the exact role of Hel308b is independent of helicase function, which supports the separation of function of Hel308a between helicase-independent recombination and helicase-dependent DNA repair (See Chapter 3: Separation of function in Hel308 for further details). Mutants deleted for both *hel308a* and *hel308b* demonstrate a sensitivity to DNA-crosslinking agents and a recombination defect compared to the wild type similar to that of the *hel308a* single delete mutants (Gamble-Milner, 2016). Due to the conserved nature of Hel308b, it would be of interest to introduce the Δ*ashA,* Δ*lhr1,* Δ*lhr2,* and Δ*dna2* into the *hel308b*-single and the *hel308a, hel308b*-double delete genetic backgrounds, and to characterise all viable strains in order to determine any redundancy present.

5.7 Conclusion

Hel308 might play a role in the Hjc/PINA fork restart or Holliday junction migration pathway however, the mechanism of this is not known. The double deletion of Hjc and PINA may be synthetically lethal, suggesting that Hjc and PINA have overlapping but distinct functions. Unlike Hel308, Cas3 is not a regulator of recombination.

Chapter 6: Genetic analysis of the *recJ* genes

6.1 Background

6.1.1 The CMG complex and DNA replication

Archaeal Replication

Archaeal replication resembles its eukaryotic counterpart, for which the CMG complex is essential. The CMG complex is highly conserved throughout eukaryotes and archaea, comprising Cdc45, the replicative helicase MCM, and GINS (Makarova, Koonin and Kelman, 2012). Archaeal RecJ, known as GINS-associated nuclease (GAN) in several archaeal species, is proposed to be the Cdc45 homologue. However, few archaeal RecJ-MCM-GINS complexes have been isolated biochemically. Whilst *H. volcanii* MCM and GINS homologues are known, a Cdc45 homologue has yet to be identified.

H. volcanii Replication

DNA replication is typically origin-dependent. Recombination-dependent replication is an alternative pathway possible in *Haloferax volcanii* due in part to the organism's polyploidy status. The formation and remodelling of a single D-loop generates a replication fork. D-loops form as a response to double strand breaks and similar DNA damage. DNA damage can of course arise anywhere on the genome, thus recombination-dependent replication eliminates the requirement of an origin. Recombination-dependent replication allows *H. volcanii* to grow faster when all of the main chromosomal origins are deleted. The machinery driving replication may differ between origin-dependent and recombination-dependent replication, however MCM has been identified as essential for both pathways (Marriott, 2017). GINS is believed to be essential for DNA replication pathways, but this is yet to be tested in an originless strain.

6.1.2 *H. volcanii* RecJ

Archaeal RecJ/Gan

Bioinformatic and structural studies have revealed that human Cdc45 and archaeal RecJ are evolutionarily related to bacterial RecJ (Makarova, Koonin and Kelman, 2012; Pellegrini, 2017). The RecJ family of proteins are typically 5'-3' exonucleases and have been implicated in several DNA repair pathways including homologous recombination, mismatch repair, and base excision repair.

It is also thought that RecJ is important in the rescue of stalled replication forks (Nagata, Simons, *et al.*, 2017). However, some RecJ proteins are exceptions to the rule. For example, *Thermoplasma acidophilum* has 2 RecJ-like proteins, TaRecJ1 and TaRecJ2. Unlike the majority of the RecJ family, TaRecJ2 has 3'-5' exonuclease activity which preferentially cuts RNA over DNA. In *T. acidophilum*, it was shown that this unusual TaRecJ2 formed a stable complex with TaGINS, but the typical TaRecJ1 did not. The TaRecJ2-TaMCM-TaGINS complex was observed *in vivo* however, TaRecJ2 does not directly interact with TaMCM and was therefore not required for the helicase activation (Ogino *et al.*, 2017). Similarly, *Methanocaldococcus jannaschii* has two RecJ-like proteins that exhibit 3'-5' exonuclease activity and therefore preferentially cuts ssRNA (Yi *et al.*, 2017).

H. volcanii RecJ

Haloferax volcanii has four RecJ homologues, designated RecJ1 (HVO_0073), RecJ2 (HVO_1147), RecJ3 (HVO_1018), and RecJ4 (HVO_ 2889). RecJ1 and RecJ3 have maintained potential 5'-3' exonuclease activity and are therefore typical RecJ-like proteins. In contrast, RecJ2 and RecJ4 have lost the residues essential for exonuclease activity, which is similar to eukaryotic Cdc45 (personal correspondence, Julie Maupin, and Figure 6.1).

Prior to the commencement of this study, a strain where the chromosomal copy of *hel308* is replaced with an N-terminally *his$_6$-tagged* and C-terminally *strepII-tagged* version of *hel308* was used to explore protein:protein interaction partners of Hel308 *in vivo*. However, it was determined that the presence of the StrepII-tag at the C-terminus of the Hel308 has a detrimental effect on the growth of the strain and therefore, the *in vivo* protein:protein interaction assay was repeated with an improved tagged Hel308 strain (See Chapter 4: Biochemical Analysis of Hel308 for further details). Both assays identified RecJ4 as a potential protein interacting partner of Hel308. Furthermore, RecJ4 was identified as a protein interacting partner of Orc1 in a similar *in vivo* protein:protein interaction assay (Darya Ausiannikava, unpublished data). These data suggest that RecJ4 plays a role in DNA repair and replication pathways and give further insight into the role of Hel308.

6.2 Aims and Objectives

It is of interest to further explore the roles of the four *H. volcanii* RecJ proteins and their relationship with Hel308a, and to identify the *H. volcanii* Cdc45 homologue. The objectives of this chapter were to:

- Generate strains deleted for all four *H. volcanii recJ* genes, individually and in combination.
- Analyse the growth and determine generation time of strains deleted for *recJ* genes.
- Generate chromosomal tagged *recJ* constructs in order to screen for *in vivo* protein:protein interactions of the four RecJ proteins expressed natively in *H. volcanii*. This may give further insight into the cellular role of the four RecJ proteins.

6.3 Materials

6.3.1 Strains

Table 6.1 ***Haloferax volcanii* strains used in this chapter.**

Parent: H164 (*ΔpyrE2, bgaHa-Bb, leuB-Ag1, ΔtrpA*)		
Strain	**Parent**	**Genotype**
H3929	H3901	Δ*recJ1::trpA+*
H3931	H3905	Δ*recJ3::trpA+*
H3932	H3907	Δ*recJ4::trpA+*
H4099	H4092	Δ*recJ1*
H4103	H4095	Δ*recJ3*
H4105	H4097	Δ*recJ4*
H4170	H4164	Δ*recJ1,* Δ*recJ3::trpA+*
H4173	H4167	Δ*recJ3,* Δ*recJ4::trpA+*
H4174	H4168	Δ*recJ4,* Δ*recJ1::trpA+*
H4190	H4176	Δ*recJ1,* Δ*recJ3*
H4192	H4179	Δ*recJ3,* Δ*recJ4*
H4193	H4180	Δ*recJ4,* Δ*recJ1*
H4273	H4268	Δ*recJ1,* Δ*recJ3,* Δ*recJ4::trpA+*
H4275	H4270	Δ*recJ3,* Δ*recJ4,*Δ*recJ1::trpA+*
H4376	H4367	Δ*recJ1,* Δ*recJ3,* Δ*recJ4*

6.3.2 Plasmids

Table 6.2 **Plasmids used in this chapter.**

Name	Use	Notes
pBluescriptII SK+	Standard *E. coli* vector with functional blue/white screening capability	From Stratagene
pTA1716	*ginS* genomic clone	Constructed by HM
pTA1882	*recJ4* genomic clone	
pTA1887	*recJ1* genomic clone	
pTA1888	*recJ2* genomic clone	
pTA1889	*recJ3* genomic clone	
pTA1905	*recJ2* genomic clone with *pyrE2* marker	
pTA1912	*recJ1* genomic clone with *pyrE2* marker	
pTA1913	*recJ3* genomic clone with *pyrE2* marker	
pTA1924	*recJ1* deletion construct	
pTA1927	*recJ3* deletion construct	
pTA1951	*recJ2* deletion construct	
pTA1958	*recJ1::trpA+* deletion construct	
pTA1960	*recJ3::trpA+* deletion construct	
pTA1991	*recJ4* deletion construct	
pTA1996	*recJ2::trpA+* deletion construct	
pTA1997	*recJ4::trpA+* deletion construct	
pTA2082	Intermediate construct for creating pTA2090	
pTA2083	Intermediate construct for creating pTA2091	
pTA2084	Intermediate construct for creating pTA2130	
pTA2085	Intermediate construct for creating pTA2095	
pTA2090	N-terminally his6-tagged and a C-terminally StrepII-tagged recJ1 gene replacement construct	
pTA2091	N-terminally his6-tagged and a C-terminally StrepII-tagged recJ2 gene replacement construct	
pTA2095	N-terminally his6-tagged and a C-terminally StrepII-tagged recJ4 gene replacement construct	
pTA2130	Intermediate construct for creating N-terminally his6-tagged and a C-	

	terminally StrepII-tagged recJ3 gene replacement construct
pTA2184	Intermediate construct for creating pTA2315
pTA2315	*ginS* partial deletion construct maintaining ~100bp C-terminus (suspected promoter for *bcp1*)
pTA2335	Δ*ginS::trpA*+ partial deletion construct maintaining ~100bp C-terminus (suspected promoter for *bcp1*)

HM = Hannah Marriott

6.3.3 Oligonucleotides

Table 6.3 **Oligonucleotides used in this chapter.**

Name	Sequence (5'-3')	Description	Used to construct
pBSF2	TTAAGTTGGGTAACGCCAGGG	Amplification of the US region to multiple genes: *recJ2, recJ4, hel308a*. Sequencing primer	pTA1923, pTA1951, pTA1991, K53A plasmids
pBSR3	ACCCCAGGCTTTACACTTTATGC	Amplification of the DS region to multiple genes: *recJ2, recJ4*. Sequencing primer	pTA1923, pTA1951, pTA1991
RecJ1NcoI_F	CGTGAGTACCACACCcATGGACGGACCCGTCCCC	Incorporation of an *Nco*I site at the *recJ1* start codon	pTA2082
RecJ1NheI_R	CGGTCGACGAgctagCCGCGTTTTCAGCCAACTGC	Incorporation of an *Nhe*I at the *recJ1* stop codon	pTA2082
RecJ2PciI_F	CCTCGAAGAATAACTacATGTCCGTGAGCCCCGC	Incorporation of a *Pci*I site at the *recJ2* start codon	pTA2083
RecJ2NheI_R	GGGTCGCACGCCGGCgCtagcGCGCACCTCCC	Incorporation of an *Nhe*I at the *recJ2* stop codon	pTA2083
RecJ3BspHI_F	CCTAGCGGGAtcATGAGCGACGAGCACGCCGGGG	Incorporation of a *Bsp*HI site at the *recJ3* start codon	pTA2084
RecJ3NheI_R	CGCCGGCgctaGCCGTCGTCGACAGCTCTTCG	Incorporation of an *Nhe*I at the *recJ3* stop codon	pTA2084
RecJ4NcoI_F	GCTCAACGCcATGGATTGGATTACGCACGAGGAGG	Incorporation of an *Nco*I site at the *recJ4* start codon	pTA2085

RecJ4NheI_R	GGATTCGctagcGAACTGCTCGGCGGCGGCGTCG	Incorporation of an *Nhe*I at the *recJ4* stop codon	pTA2085
RecJ4probeF	CGAGTACCACTACTTCACCCGTCGCCCGC	Amplification of internal region of *recJ4*	RecJ4 colony hyb probe
RecJ4probeR1	CCTGATACGACTGGTAGTAGGCTTCC	Amplification of internal region of *recJ4*	RecJ4 colony hyb probe
RecJ2probeF	GCACGACACCGAAGGAAGACACCGACC	Amplification of internal region of *recJ2*	RecJ2 colony hyb probe
RecJ2probeR	GGCGAGGACAGGTCACAGGAGAGCG	Amplification of internal region of *recJ2*	RecJ2 colony hyb probe
RecJ1_F	CGAAGCGGGAGAGGAGGCGAGAGTC	Amplification of 4232 bp region of chromosome containing *recJ1*	pTA1887
RecJ1_R	CCGTCTCTGTTTGTACCTCCG	Amplification of 4232 bp region of chromosome containing *recJ1*	pTA1887
RecJ2_F	GGTAGAACGTGAGGTAGTTGTCGGC	Amplification of 4009 bp region of chromosome containing *recJ2*	pTA1888
RecJ2_R	CGTACCTGCCCTTTCCCTCTTACG	Amplification of 4009 bp region of chromosome containing *recJ2*	pTA1888
RecJ3_F	GTCCTGCGTGAAGTCGAAGAACG	Amplification of 5198 bp region of chromosome containing *recJ3*	pTA1889
RecJ3_R	TTCGTTCTCTTTCTCTCGCAGTCGCTCGC	Amplification of 5198 bp region of chromosome containing *recJ3*	pTA1889
delrecJ4NdeR	CCAATCCATatgGTTGAGCAAACAACCCCTAACTAAAACGATTGTCG	Amplification of the upstream region to *recJ4*	pTA1991
delrecJ4BamF	AGCCGgATCCAGACTCCTCCTCGCGCGG	Amplification of the downstream region to *recJ4*	pTA1991
delrecJ1BamF	GCAGTTGGCTGgAtcCGCGGACTAATCGTCGACCG	Forward inverse PCR primer to delete *recJ1*	pTA1924
delrecJ1NdeR	GGGTCCGTCCATatGTGTGGTACTCACGCCC	Reverse inverse PCR primer to delete *recJ1*	pTA1924
delrecJ3BglF	CGTTCagaTCTCAGACTGCGACCGCTCATTTTTCG	Forward inverse PCR primer to delete *recJ3*	pTA1927

delrecJ 3NdeR	CGTCGCTCATAtg CCCGCTAGGTCA GG	Reverse inverse PCR primer to delete *recJ3*	pTA1927
delrecJ 2BglF	GCCGTCAGaGAtc TGCGCCGATGA GCCG	Amplification of the downstream region to *recJ2*	pTA1951
delrecJ 2NdeR	GCTCACGGACA TatgGTTATTCTT CGAGG	Amplification of the upstream region to *recJ2*	pTA1951
gintagD SXbaR	ACGCtCTaGACG AAATCGGCGCG AGCGTCACCG	Amplification of the downstream region to ginS and introduction of an XbaI site	pTA2315
ginCter mBamF	GGTGCGgATCcC CGCCGACGTGG GGACCATCCTCG	Amplification of C-terminal region of *ginS* (~100 bp) and introduction of a *Bam*HI site	pTA2315
TrpABg lF	GCTCGGTACCCG GaGATCtCGTGG ATAAAACCCC	Amplification of *trpA* incorporating a *Bgl*II site	pTA2335
TrpABg lR	CCAAGCTTGCAT GCCaGATCtGTTA TGTGCGTTCC	Amplification of *trpA* incorporating a *Bgl*II site	pTA2335
NdeLin kBamF	TATGCGACGCGT CGGAATTCCCG	NdeLinkBam linker oligo with *Nde*I and *Bam*HI ends	pTA1924, pTA1991
BamLin kNdeR	GATCCGGGAAT TCCGACGCGTCG CA	NdeLinkBam linker oligo with *Nde*I and *Bam*HI ends	pTA1924, pTA1991
BglLink NdeR	GATCTGGGAATT CCGACGCGTCG CA	NdeLinkBgl linker oligo with *Nde*I and *Bgl*II ends	pTA1927, pTA1951
NdeLin kBglF	TATGCGACGCGT CGGAATTCCCA	NdeLinkBgl linker oligo with *Nde*I and *Bgl*II ends	pTA1927, pTA1951
TRPAR	CGTTGCGACGC GCCCGCTACC	PCR screen for upstream pop-in of pTA1996 at RecJ2 locus	

6.4 Results

6.4.1 Phylogenetic analysis of Archaeal *recJ*

Domain analysis of *H. volcanii* RecJ Proteins

The domain structure of the four *H. volcanii* RecJ proteins were analysed by protein sequence alignment with RecJ from *Thermus thermophilus*, GAN from *Thermococcus kodakaraensis*, and Cdc45 from *Homo sapiens.* Alignment of the proteins was carried out in MacVector using T-Coffee (Myers Miller; penalty for open gap = -50; extend gap = -50) (Figure 6.1).

Predicted domains based on the *Thermococcus kodakarensis* (TKO) GAN, *Homo sapiens* Cdc45, and *Thermus thermophilus* (TTH) RecJ crystal structures (Pellegrini, 2017) are indicated as follows: DHH in blue, DHHA1 in red, and the central helix domain connecting the DHH and DHHA1 domains is in turquoise. The highly conserved DHH motif within the DHH domain is also highlighted, note that *H. volcanii* (HVO) RecJ2 and RecJ4 have lost the conserved histidine residues and therefore almost certainly lost exonuclease activity.

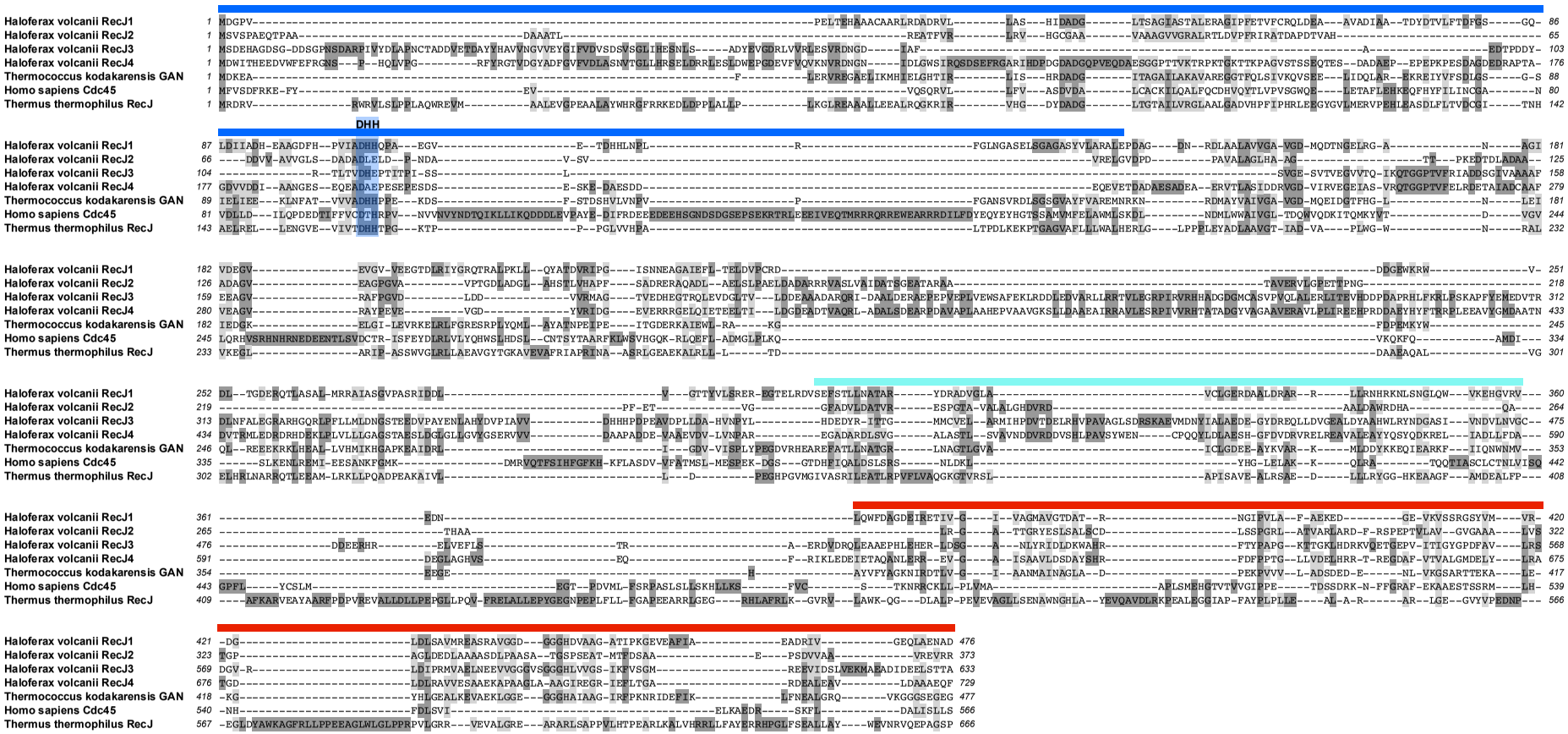


Figure 6.1: **RecJ protein sequence alignment.** Multiple sequence alignment of RecJ1, RecJ2, RecJ3, and RecJ4 from Haloferax volcanii, GAN from Thermococcus kodakarensis, Cdc45 from Homo sapiens, and RecJ from Thermus thermophilus. Predicted domains based on the GAN crystal structure are indicated as follows: DHH in blue (see conserved DHH motif), central helix in green, and DHHA1 in red. Alignment carried out in MacVector using T-Coffee (Myers Miller; penalty for open gap = -50; extend gap = -50)

To further analyse the *H. volcanii* RecJ proteins domain structure, the protein sequences were analysed via Pfam domain search (Figure 6.2).

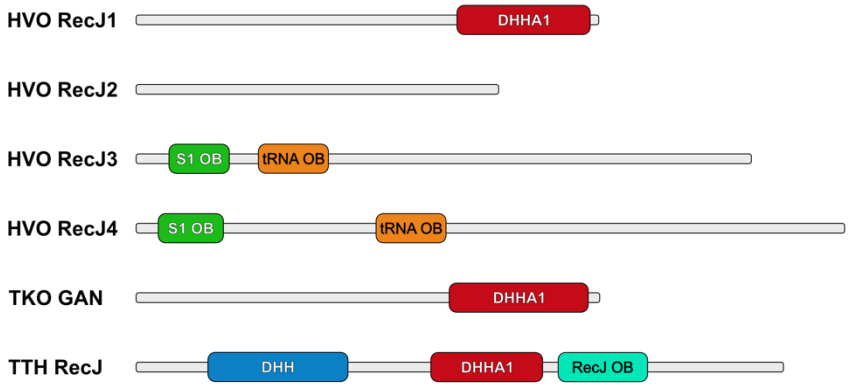


Figure 6.2: **Pfam Domain analysis of RecJ proteins.** Pfam domain analysis of RecJ1, RecJ2, RecJ3, and RecJ4 from *Haloferax volcanii* (HVO), GAN from *Thermococcus kodakarensis* (TKO), and RecJ from *Thermus thermophilus* (TTH). Domain search carried out on Pfam (Pfam-A E-value cut-off = 1.0)

Pfam domain analysis reveals that RecJ1 is most similar to TKO GAN, with both containing a DHHA1 domain (PF02272). The DHHA1 domain spans approximately 60 residues, which contain a conserved GG motif. The DHHA1 domain is named for the DHH associated domain thus, is often found adjacent to a DHH domain (PF01368) as seen in TTH RecJ. The DHHA1 domain is also found in alanyl tRNA synthetase, which suggests an RNA binding function. It is predicted that the entire DHH family of proteins perform a phosphoesterase role. Whilst Pfam analysis reveals a lack of a DHH domain in TKO GAN, *in vitro* experiments show that TKO GAN has DNA nuclease activity (Oyama *et al.*, 2016) therefore, it is possible the archaeal equivalent for the DHH domain is yet to be identified. The final annotated domain for TTH RecJ is the RecJ OB domain (PF17768). The OB-fold (oligonucleotide/oligosaccharide-binding fold) architecture comprises of a five/six-stranded closed beta-barrel formed by 70-80 residues. The loops connecting the strands vary in length and form the functional binding face of the protein. Different OB-fold proteins use this binding face to bind a variety of substrates such as oligosaccharides, oligonucleotides, proteins, metal ions and catalytic substrates (Mitchell *et al.*, 2019). The RecJ OB-fold is found in RecJ proteins and is responsible for binding ssDNA. However, Pfam does not recognise the presence of this domain in TKO GAN or any of the four HVO RecJ proteins. As stated previously, TKO GAN is known to bind DNA *in vitro* and therefore, it is likely that the DHHA1 domain is able to bind to ssDNA as well as RNA, or that the domain responsible is yet to be identified.

Though HVO RecJ3 and RecJ4 are lacking a RecJ OB-fold, two different OB folds have been found in both, these are the S1 RNA binding domain (PF00575) and the tRNA OB-fold nucleic acid binding domain (PF01336). The S1 domain has an OB-fold structure and is found in a wide range of RNA associated proteins. The OB-fold nucleic acid binding domain family is composed of a variety of OB-fold domains, which includes the anti-codon binding domain of aminoacyl-tRNA synthetases that catalyse the addition of an amino acid to the appropriate tRNA molecule, part of the RecG helicase involved in DNA repair, a subunit of replication factor A, and bacterial DNA polymerase III alpha chain. It is probable that HVO RecJ3 and RecJ4 are able to bind single stranded DNA and RNA via these domains. Pfam was unable to identify any known domains in HVO RecJ2.

Phyre2 was used to predict 3D structure (Kelley *et al.*, 2015). Phyre2 utilises PSI-BLAST to analyse and align a protein sequence to a database of known 3D structures. This technique is able to predict 3D structure based on remote homology (Figure 6.3). Phyre2 predicted 3D structures for HVO RecJ1, RecJ3, and RecJ4 that closely resembled the known crystal structures of TKO GAN, TTH RecJ, and human Cdc45. However, the 3D structure of RecJ2 remained ambiguous.

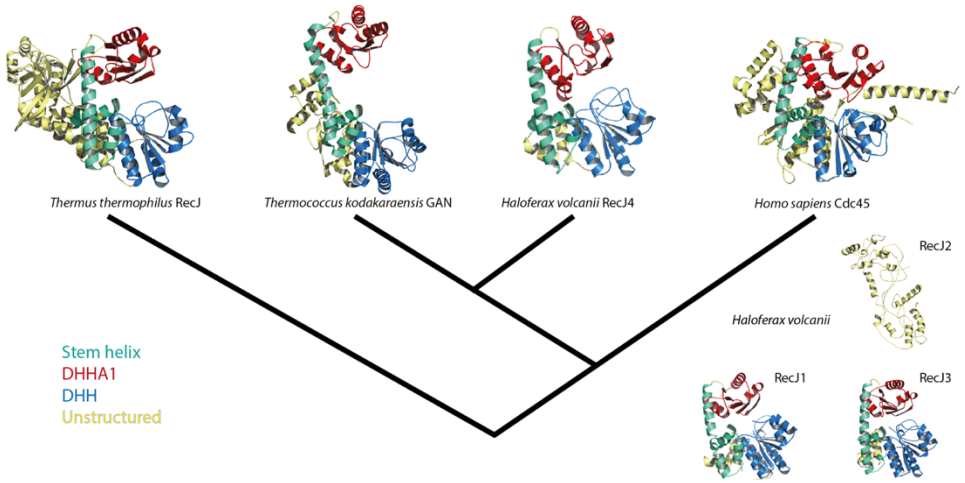


Figure 6.3: **Predicted structure of the four *H. volcanii* RecJ proteins.** Structures predicted using Phyre2 compared with the known crystal structures of *Thermus thermophilus* RecJ (PDB ID 2ZXO), *Thermococcus kodakarensis* GAN (PDB ID 5GHT), and *Homo sapiens* Cdc45 (PDB ID 5DGO). The structures are drawn as ribbon diagrams and coloured to highlight the RecJ/Cdc45 fold. The N-terminal DHH domain is in blue, the C-terminal DHHA1 domain is in red and the stem domain connecting the two DHH domains is in green. The 3D predicted structures of *H. volcanii* RecJ1, RecJ3, and RecJ4 reveal a close resemblance to the known crystal structures of *T. thermophilus* RecJ, *T. kodakarensis* GAN, and human Cdc45. Little information is known about the 3D structure of RecJ2. Figure adapted from Pellegrini 2016.

Distribution of RecJ/GAN

In order to determine the distribution of RecJ/GAN within the archaea, a BLAST search was initially carried out using the RecJ4 protein sequence from *H. volcanii* (RecJ4 UniProt accession number: D4GXM0). RecJ4 was chosen as the protein was identified as an interacting partner of Hel308 (See Chapter 6: Biochemical analysis of Hel308 for further details). Subsequent BLAST searches using RecJ1, RecJ2, and RecJ3 sequences from *H. volcanii* (RecJ1, 2, and 3 UniProt accession numbers: D4GYR9, D4GWB0, and D4GVJ5 respectively) were carried out and potential homologues were cross-compared (Figure 6.4).

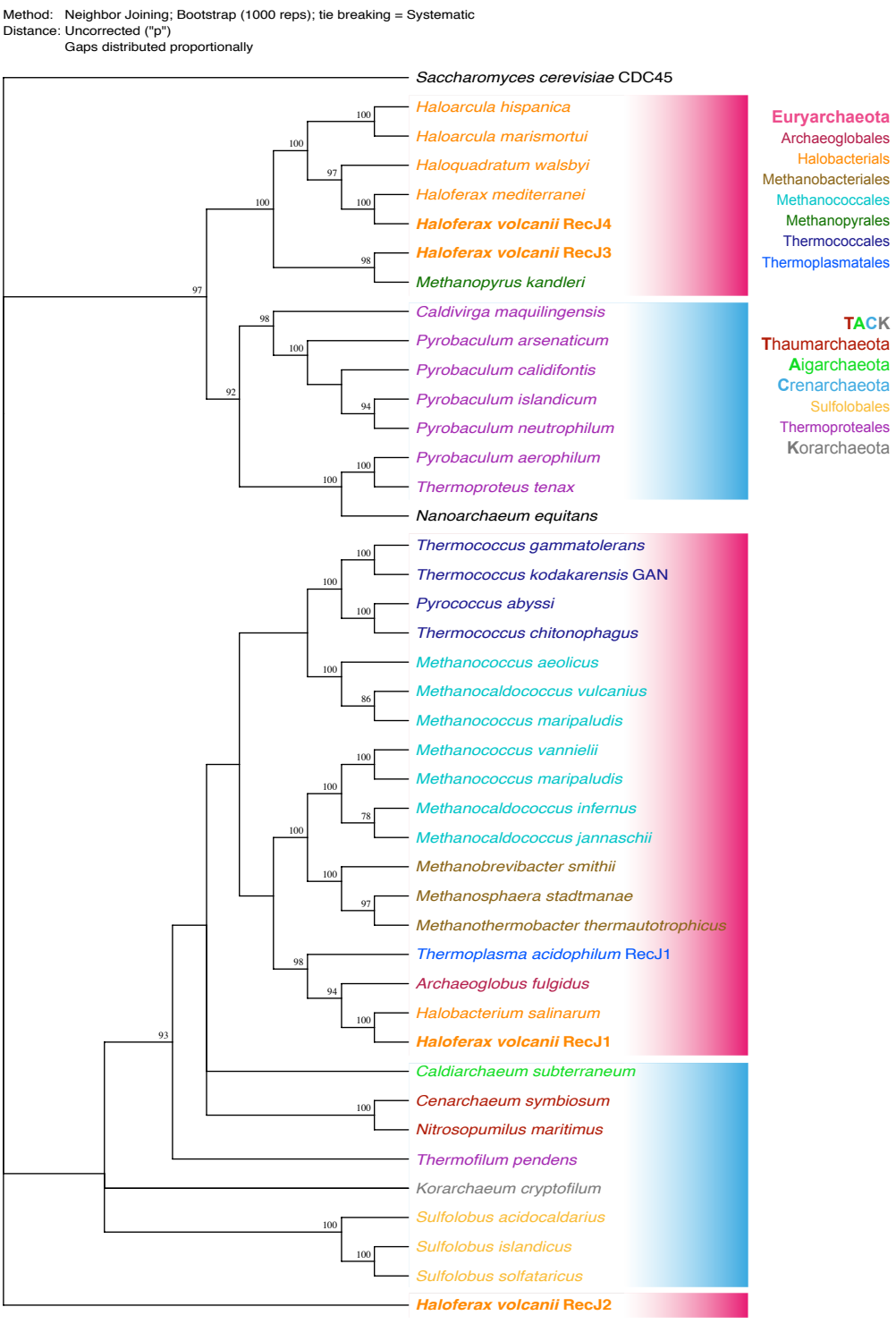


Figure 6.4 **Phylogenetic tree of RecJ**. Rooted phylogenetic tree of RecJ from Archaea and *Saccharomyces cerevisiae*. Tree rooted to *Saccharomyces cerevisiae*. Euryarchaeota are highlighted in pink and the TACK archaea are highlighted in blue. The four *H. volcanii* RecJ proteins are in **bold**. Calculated using Neighbour Joining method (Bootstrap 1000 reps). Numbers above branches indicate the percentage likelihood supporting the nodes. Generated using MacVector 15.5.4

The distribution of the four HVO RecJ proteins indicate that RecJ3 and RecJ4 are closely related to one another and group with the majority of other halophilic archaeal RecJ proteins. RecJ1 is less similar to the halophilic RecJ proteins however, RecJ1 still groups with other Euryarchaeota RecJ proteins including *Halobacterium salinarum*. HVO RecJ2 is different from the other three HVO RecJ proteins and does not fall into a logical group.

6.4.2 Generation of *recJ* and *ginS* Deleted Strains

Gene deletion constructs for all four *H. volcanii recJ* genes were developed by first generating genomic clones of each of the *recJ* genes and the gene flanking regions. The genomic clone for *ginS* was made prior to this study.

pTA1882 (*recJ4* genomic clone)

A PCR product was difficult to obtain due to significant secondary DNA structure within the *recJ4* gene and the flanking regions, predicted using Mfold software (Zuker, 2003). Therefore, the more traditional technique of generating a genomic clone via restriction digest was employed. Genomic DNA extracted from the wildtype strain H26 (see section 2.2.3 Genomic DNA Extraction from *Haloferax volcanii* for further details) was digested with the restriction enzymes *Bgl*II and *Not*I and was ligated into the *Bam*HI and *Not*I sites in pTA131 to create a genomic library. The ligation mixture was then digested with *Xba*I, in order to remove any self-ligated pTA131 vector, before transformation of the genomic library into *E. coli.* Transformants were plated on LB solid media supplemented with ampicillin (Amp), 5-Bromo-4-chloro-3-indolyl β-D-galactopyranoside (X-Gal), and Isopropyl β-D-1-thiogalactopyranoside (IPTG) to allow for blue/white screening. The insertion of a genomic region at the *Bam*HI and *Not*I sites in pTA131 should disrupt the *lacZ* gene producing white colonies, whereas blue colonies are indicative of a pTA131 vector self-ligation. White colonies were patched and screened for *recJ4* via colony lift and hybridisation. An internal gene probe for *recJ4* was generated via PCR amplification of the genomic DNA extracted from H26 using the primers RecJ4probeF and RecJ4probeR1. Plasmid DNA was extracted from *recJ4* positive *E. coli* colonies, confirmed via DNA sequencing to generate the *recJ4* genomic clone, pTA1882 (Figure 6.5).

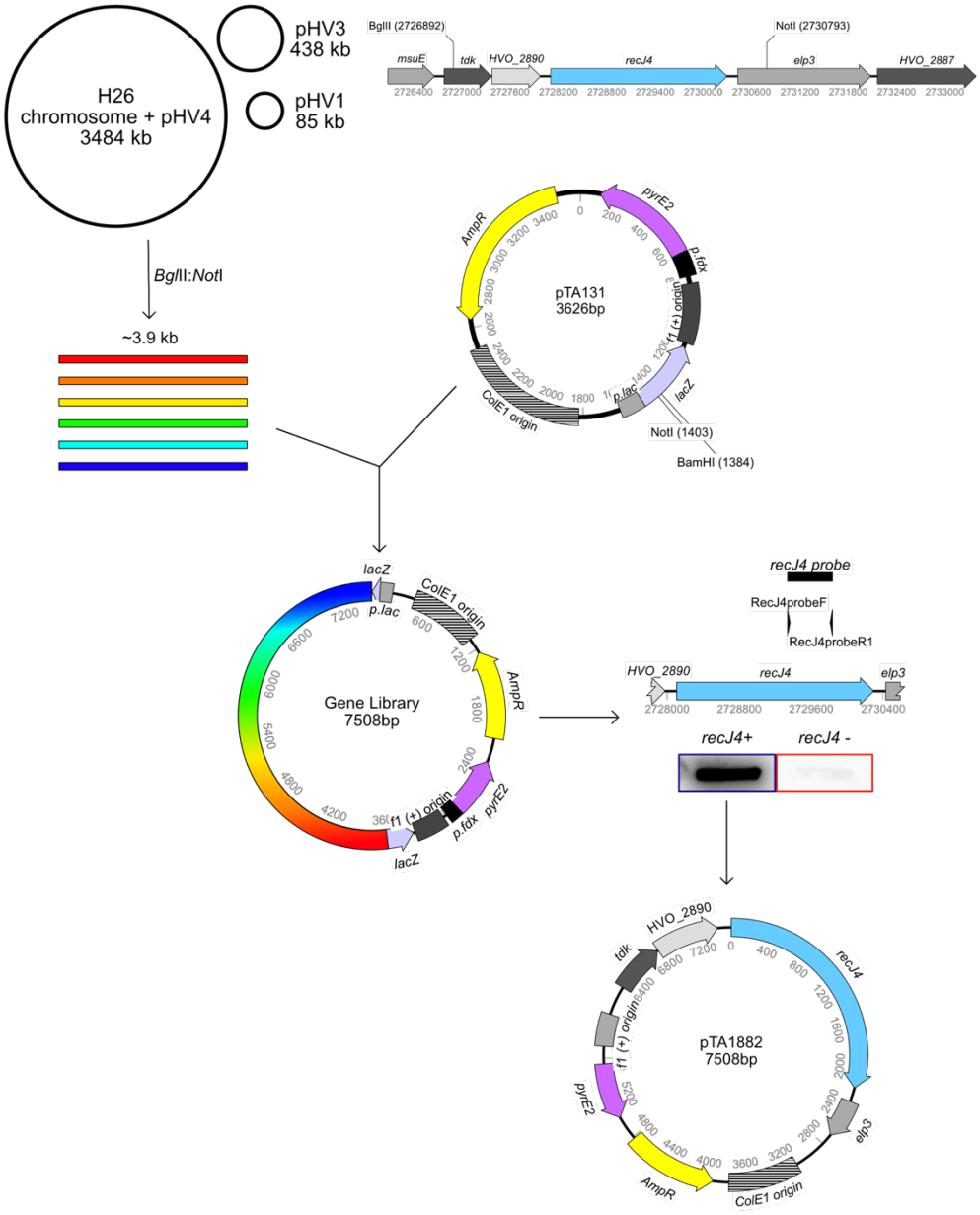


Figure 6.5 **Generation of *recJ4* genomic clone, pTA1882.** Genomic DNA extracted from H26 was digested with *Bgl*II and *Not*I and ligated into the *Bam*HI and *Not*I sites in pTA131 to create a gene library. The library was transformed into *E. coli* and candidate colonies were screened with a radiolabelled RecJ4probeF to RecJ4probeR1. PCR product of *recJ4* from H26 genomic DNA. Colonies that did hybridise were *recJ4*+. DNA sequencing confirmed the *recJ4* genomic clone, pTA1882.

pTA1887 (*recJ1* genomic clone)

Genomic DNA extracted from the wildtype strain H26 was used as a template to PCR amplify the *recJ1* gene and flanking regions using the primers RecJ1_F and RecJ1_R. The product was digested with *Mlu*I and *Sph*I and ligated into the *Bss*HII and *Nsp*I sites in pBluescriptII SK+ to create the *recJ1* genomic clone, pTA1887. The construct was confirmed via DNA sequencing (Figure 6.6).

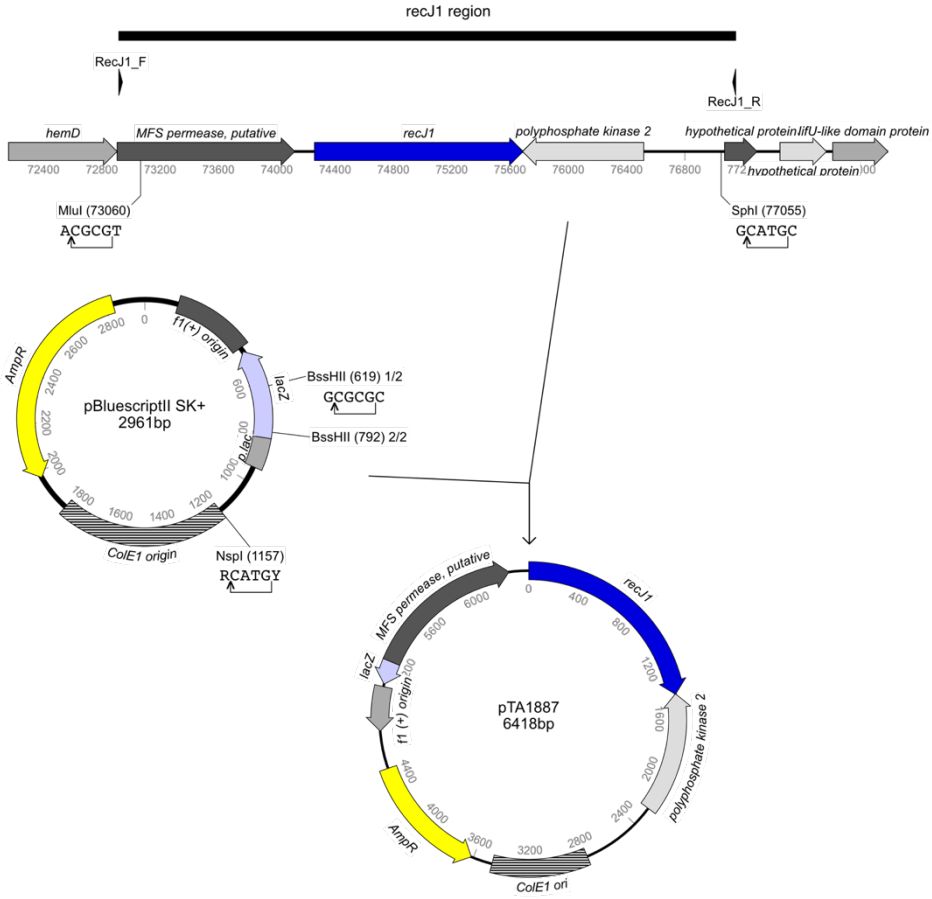


Figure 6.6 **Generation of *recJ1* genomic clone, pTA1887.** Genomic DNA extracted from H26 was used as a template to PCR amplify RecJ1_F to RecJ1_R. The product was digested with *Mlu*I and *Sph*I and ligated into the *Bss*HII (1/2) and *Nsp*I sites of pBluescriptII SK+ to create the *recJ1* genomic clone, pTA1887.

pTA1888 (*recJ2* genomic clone)

Genomic DNA extracted from the wildtype strain H26 was used as a template to PCR amplify the *recJ2* gene and flanking regions using the primers RecJ2_F and RecJ2_R. The product was digested with *Afe*I and *Age*I and ligated into the *Eco*RI and *Xma*I sites in pBluescriptII SK+ to create the *recJ2* genomic clone, pTA1888. The construct was confirmed via DNA sequencing (Figure 6.7).

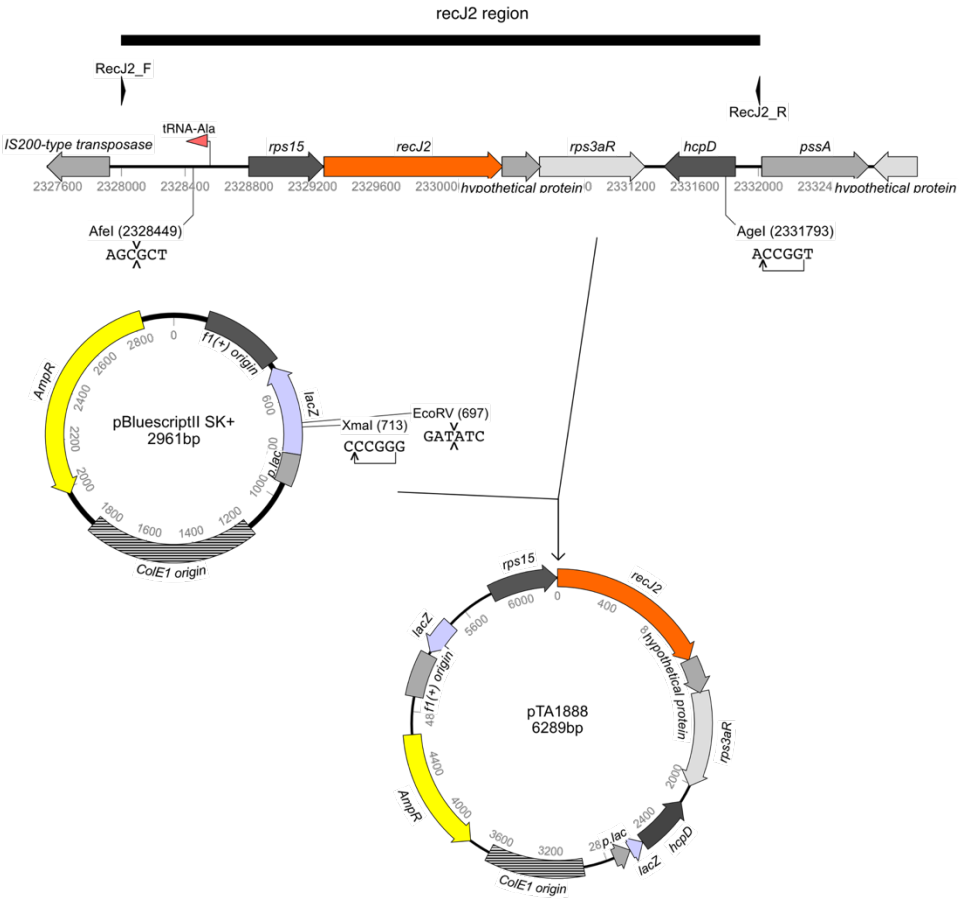


Figure 6.7 **Generation of *recJ2* genomic clone, pTA1888.** Genomic DNA extracted from H26 was used as a template to PCR amplify RecJ2_F to RecJ2_R. The product was digested with *Afe*I and *Age*I and ligated into the *Eco*RV and *Xma*I sites of pBluescriptII SK+ to create the *recJ2* genomic clone, pTA1888.

pTA1889 (*recJ3* genomic clone)

Genomic DNA extracted from the wildtype strain H26 was used as a template to PCR amplify the *recJ3* gene and flanking regions using the primers RecJ3_F and RecJ3_R. The product was digested with *Not*I and *Xma*I and ligated into the *Not*I and *Xma*I sites in pBluescriptII SK+ to create the *recJ3* genomic clone, pTA1889. The construct was confirmed via DNA sequencing (Figure 6.8).

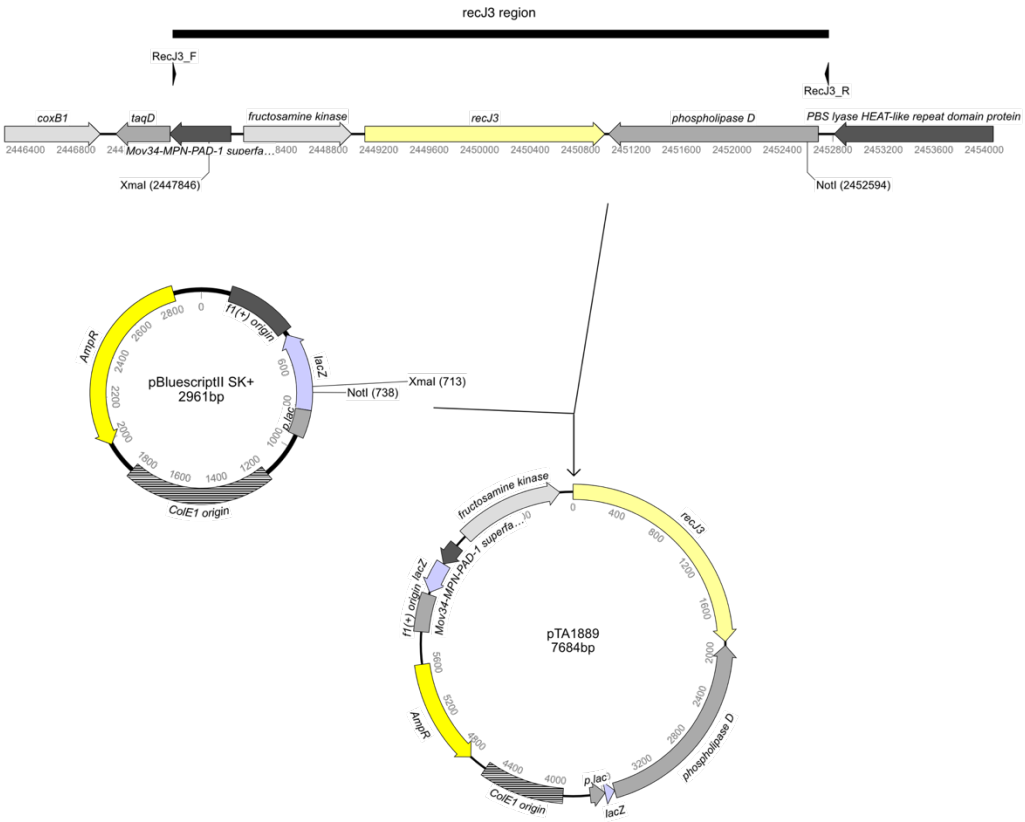


Figure 6.8 **Generation of *recJ3* genomic clone, pTA1889.** Genomic DNA extracted from H26 was used as a template to PCR amplify RecJ3_F to RecJ3_R. The product was digested with *Not*I and *Xma*I and ligated into the *Not*I and *Xma*I sites of pBluescriptII SK+ to create the *recJ3* genomic clone, pTA1889.

pTA1958 (*recJ1::trpA*+ deletion construct)

To allow for selection in *H. volcanii*, the *pyrE2* marker was first introduced to the *recJ1* genomic clone. pGB70 was digested using the restriction enzymes *Bam*HI and *Xba*I to generate a 658 bp fragment containing the *pyrE2* CDS and the *ferredoxin* promoter (Bitan-banin *et al.*, 2003). The fragment was then blunt-end filled using Klenow and ligated into the *Psi*I site of pTA1887 creating the *recJ1, pyrE2*+ genomic clone, pTA1912. pTA1912 was used as a template for inverse PCR amplification generating a 5682 bp PCR product containing the gene flanking regions, *ColE1* origin, *AmpR* and *pyrE2* markers. The primers used for the inverse PCR were delrecJ1BamF and delrecJ1NdeR introducing *Bam*HI and *Nde*I sites respectively. The BamLinkNde annealed oligo linker, made by annealing oligonucleotides NdeLinkBamF and BamLinkNdeR, was ligated into the *Bam*HI and *Nde*I digested PCR product creating the *recJ1* deletion construct, pTA1924. pTA298 was digested with *Bam*HI to generate a 965 bp fragment containing the *trpA* CDS and the *ferredoxin* promoter, which was ligated into the *Bam*HI site of pTA1924. Restriction digest was used to screen for orientation of the *trpA* marker, the same orientation as *recJ1* was preferable, and was further confirmed via DNA sequencing to generate the *recJ1::trpA*+ deletion construct, pTA1958 (Figure 6.9).

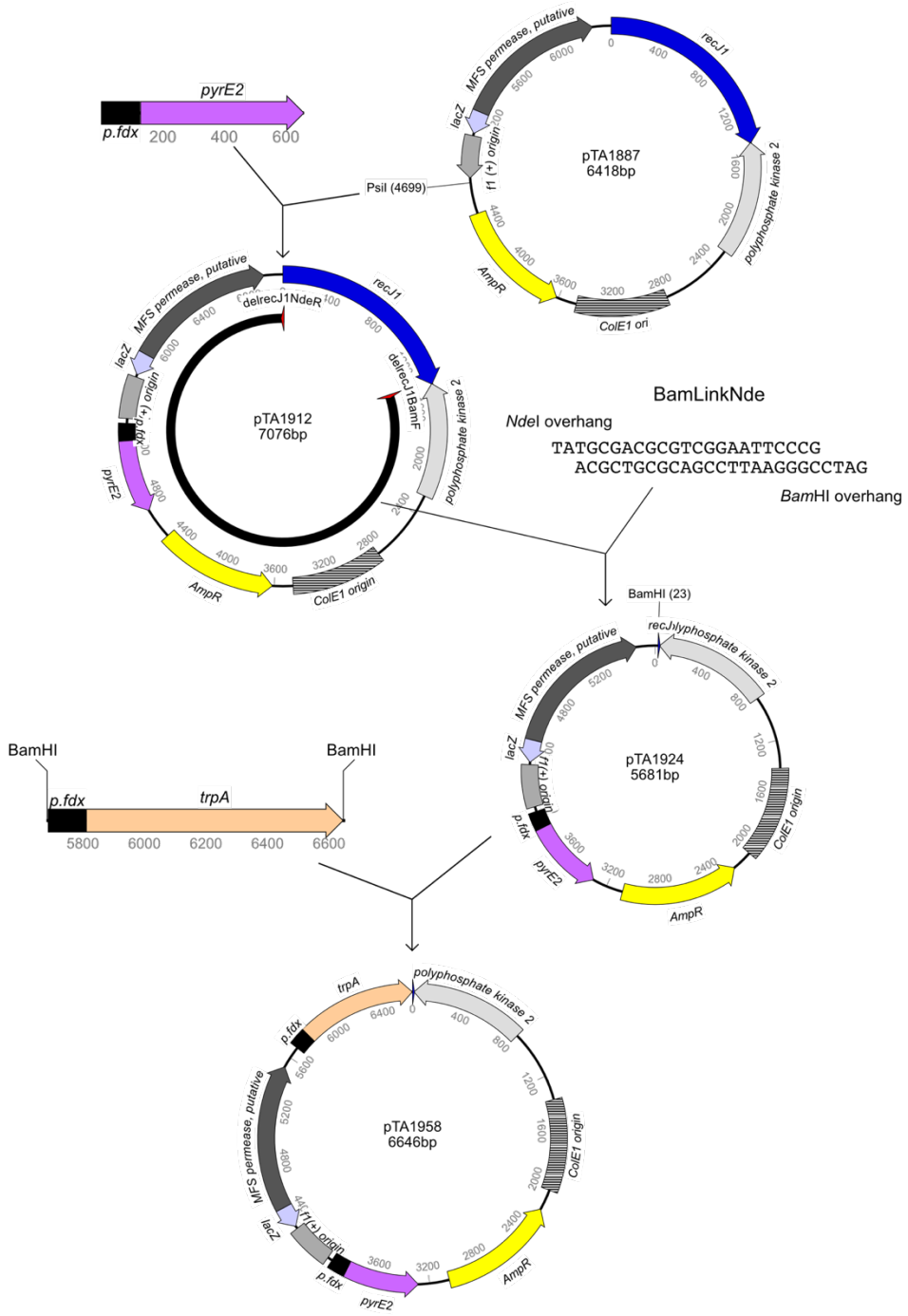


Figure 6.9 **Generation of *recJ1::trpA+* deletion construct, pTA1958.** The *pyrE2* marker digested from pTA52 using *Bam*HI and *Xba*I was blunt-end filled using Klenow and ligated into the *Psi*I site of pTA1887 creating the *recJ1, pyrE2+* genomic clone, pTA1912. pTA1912 was used as a template to PCR amplify delrecJ1BamF to delrecJ1NdeR introducing *Bam*HI and *Nde*I sites respectively. The BamLinkNde annealed oligo linker was ligated into the *Bam*HI and *Nde*I digested PCR product creating the *recJ1* deletion construct, pTA1924. The *trpA* marker digested from pTA298 using *Bam*HI was ligated into the *Bam*HI site of pTA1924 to generate the *recJ1::trpA+* deletion construct, pTA1958.

pTA1960 (*recJ3::trpA*+ deletion construct)

Similar to the *recJ1* deletion cloning strategy, the *pyrE2* marker was first introduced into the *recJ3* genomic clone. pGB70 was digested using the restriction enzymes *Bam*HI and *Xba*I to generate a 658 bp fragment containing the *pyrE2* CDS and the *ferredoxin* promoter (Bitan-banin *et al.*, 2003). The fragment was then blunt-end filled using Klenow and ligated into the *Psi*I site of pTA1889 creating the *recJ3, pyrE2*+ genomic clone, pTA1913. pTA1913 was used as a template for inverse PCR amplification generating a 6445 bp PCR product containing the gene flanking regions, *ColE1* origin, *AmpR* and *pyrE2* markers. The primers used for the inverse PCR were delrecJ3BglF and delrecJ3NdeR introducing *Bgl*II and *Nde*I sites respectively. The BglLinkNde annealed oligo linker, made by annealing oligonucleotides NdeLinkBglF and BglLinkNdeR, was ligated into the *Bgl*II and *Nde*I digested PCR product creating the *recJ3* deletion construct, pTA1927. pTA298 was digested with *Bam*HI to generate a 965 bp fragment containing the *trpA* CDS and the *ferredoxin* promoter, which was ligated into the *Bgl*II site of pTA1927. Restriction digest was used to screen for orientation of the *trpA* marker. Though the same orientation as *recJ3* was preferable, this was less crucial due to the presence of genes encoded in both directions forming the *recJ3* flanking regions. The construct was further confirmed via DNA sequencing to generate the *recJ3::trpA*+ deletion construct, pTA1960 (Figure 6.10).

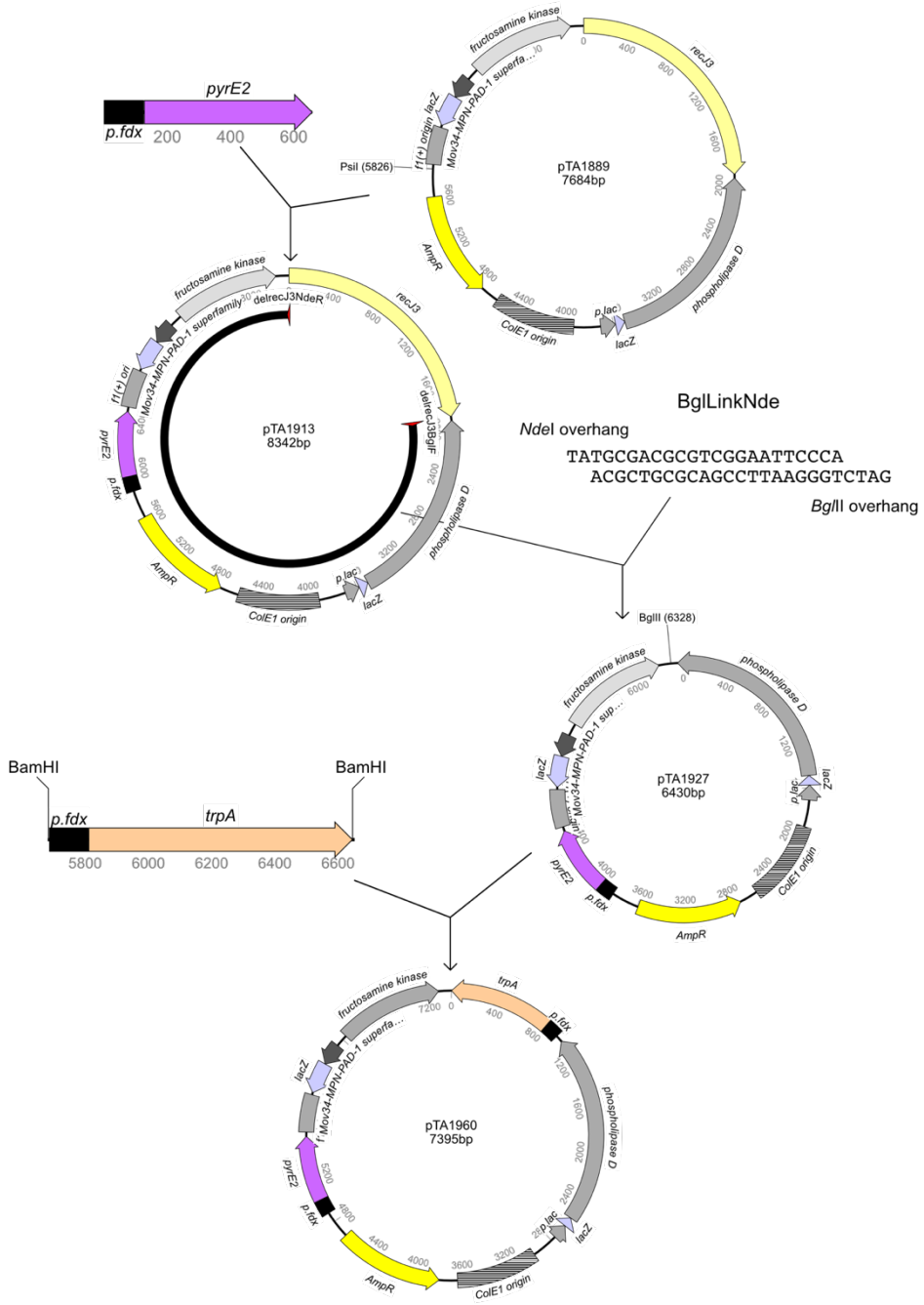


Figure 6.10 **Generation of *recJ3::trpA*+ deletion construct, pTA1960.** The *pyrE2* marker digested from pTA52 using *Bam*HI and *Xba*I was blunt-end filled using Klenow and ligated into the *Psi*I site of pTA1889 creating the *recJ3, pyrE2*+ genomic clone, pTA1913. pTA1913 was used as a template to PCR amplify delrecJ3BglF to delrecJ3NdeR introducing *Bgl*II and *Nde*I sites respectively. The BglLinkNde annealed oligo linker was ligated into the *Bgl*II and *Nde*I digested PCR product creating the *recJ3* deletion construct, pTA1927. The *trpA* marker digested from pTA298 using *Bam*HI was ligated into the *Bgl*II site of pTA1927 to generate the *recJ3::trpA*+ deletion construct, pTA1960.

pTA1996 (*recJ2::trpA*+ deletion construct)

The *recJ2* genomic clone, pTA1888, was digested with *Bam*HI and *Xmn*I to generate a 3210 bp fragment containing *recJ2* and flanking regions. The fragment was ligated into the *Bam*HI and *Eco*RV sites in pTA131 to generate the *recJ2, pyrE2*+ genomic clone, pTA1905. Not only did this method introduce the *pyrE2* marker required for *H. volcanii* selection, it also removed a region of genomic rearrangement discovered when pTA1888 was DNA sequenced. Consequently, the length of the upstream and downstream regions of *recJ2* are asymmetric in pTA1905 at 679 bp and 1403 bp respectively. Attempts to inverse PCR pTA1905 were unsuccessful therefore the upstream and downstream regions were amplified separately, joined together by means of the BglLinkNde linker, and ligated into pTA131. pTA1905 was used as a template to PCR amplify the upstream region of *recJ2* using the primers pBSF2 and delrecJ2NdeR, generating an 833 bp product that was then digested with *Nde*I to enable the ligation with the BglLinkNde linker. pTA1905 was again used as a template to PCR amplify the downstream region of *recJ2* using the primers pBSR3 and delrecJ2BglF, generating a 1609 bp product that was then digested with *Bgl*II to enable the ligation with the BglLinkNde linker. Once the upstream and downstream regions were ligated to the BglLinkNde linker, the fragment was digested with *Kpn*I and *Xba*I and ligated into the *Kpn*I and *Xba*I sites of pTA131 to generate the *recJ2* deletion construct, pTA1951. pTA298 was digested with *Bam*HI to generate a 965 bp fragment containing the *trpA* CDS and the *ferredoxin* promoter, which was ligated into the *Bgl*II site of pTA1951. Restriction digest was used to screen for orientation of the *trpA* marker, the same orientation as *recJ2* was preferable, and was further confirmed via DNA sequencing to generate the *recJ2::trpA*+ deletion construct, pTA1996 (Figure 6.11).

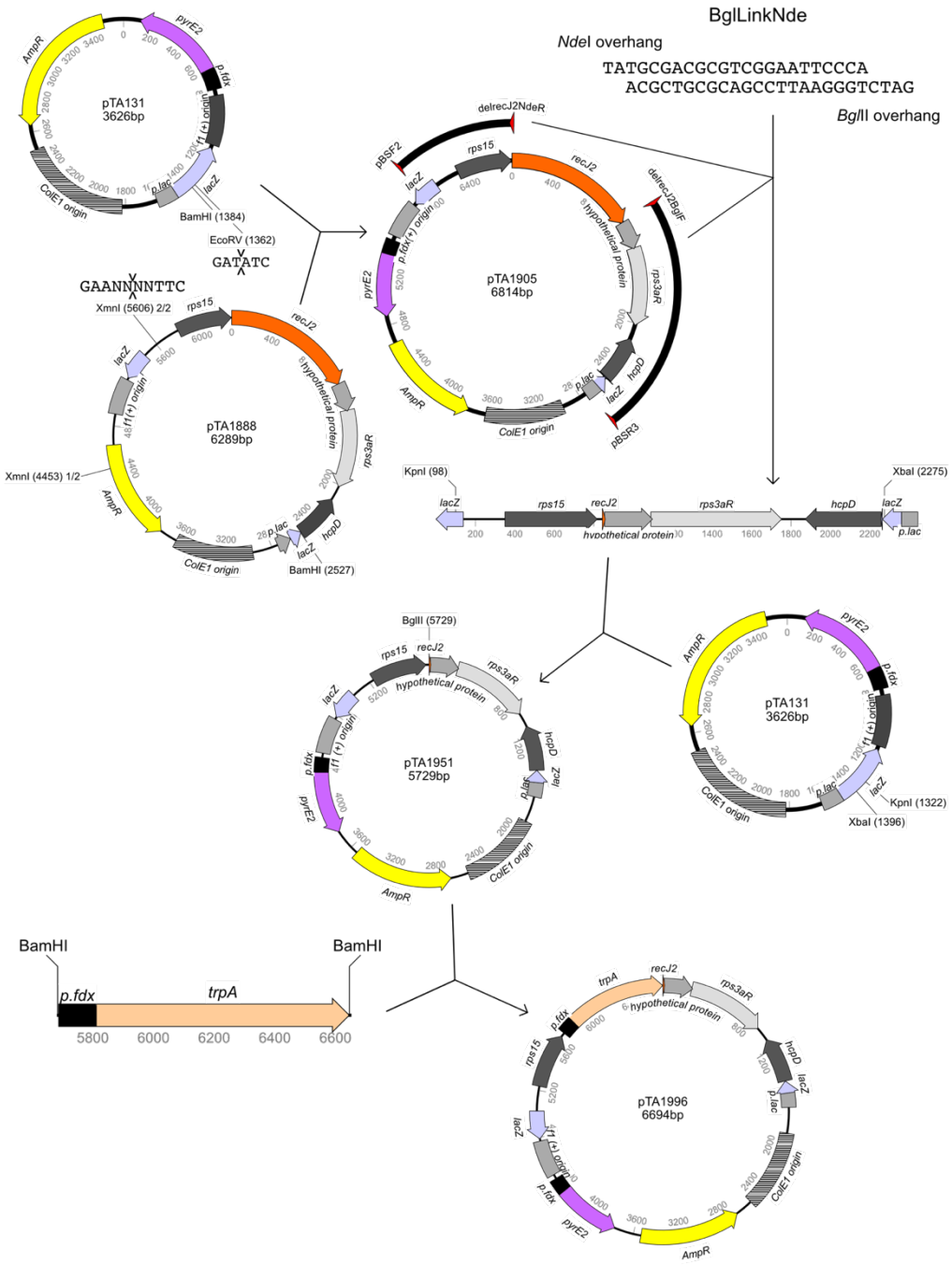


Figure 6.11 **Generation of *recJ2::trpA*+ deletion construct, pTA1996.** pTA1888 was digested with *Bam*HI and *Xmn*I and ligated into the *Bam*HI *Eco*RV site in pTA131 to generate the *recJ2, pyrE2*+ genomic clone, pTA1905. pTA1905 was used as a template to PCR amplify pBSF2 to delrecJ2NdeR and pBSR3 to delrecJ2BglF, which were digested with *Nde*I and *Bgl*II respectively and ligated via the BglLinkNde linker. The USlinkDS fragment was digested with *Kpn*I and *Xba*I and ligated into the *Kpn*I and *Xba*I sites of pTA131 to generate the *recJ2* deletion construct, pTA1951. The *trpA* marker digested from pTA298 using *Bam*HI was ligated into the *Bgl*II site of pTA1951 to generate the *recJ2::trpA*+ deletion construct, pTA1996.

pTA1997 (*recJ4::trpA*+ deletion construct)

Likewise, attempts to inverse PCR pTA1882 were unsuccessful, therefore the upstream and downstream regions were amplified separately, joined together by means of the BamLinkNde annealed oligo linker, and ligated into pTA131. pTA1882 was used as a template to PCR amplify the upstream region of *recJ4* using the primers pBSF2 and delrecJ4NdeR, generating a 1381 bp product that was then digested with *Nde*I to enable the ligation with the BamLinkNde linker. pTA1882 was again used as a template to PCR amplify the downstream region of *recJ4* using the primers pBSR3 and delrecJ4BamF, generating a 661 bp product that was then digested with *Bam*HI to enable the ligation with the BamLinkNde linker. Once the upstream and downstream regions were ligated to the BamLinkNde linker, the fragment was digested with *Not*I and *Hind*III and ligated into the *Not*I and *Hind*III sites of pTA131 to generate the *recJ4* deletion construct, pTA1991. pTA298 was digested with *Bam*HI to generate a 965 bp fragment containing the *trpA* CDS and the *ferredoxin* promoter, which was ligated into the *Bam*HI site of pTA1991. Restriction digest was used to screen for orientation of the *trpA* marker, the same orientation as *recJ4* was preferable, and was further confirmed via DNA sequencing to generate the *recJ4::trpA*+ deletion construct, pTA1997 (Figure 6.12).

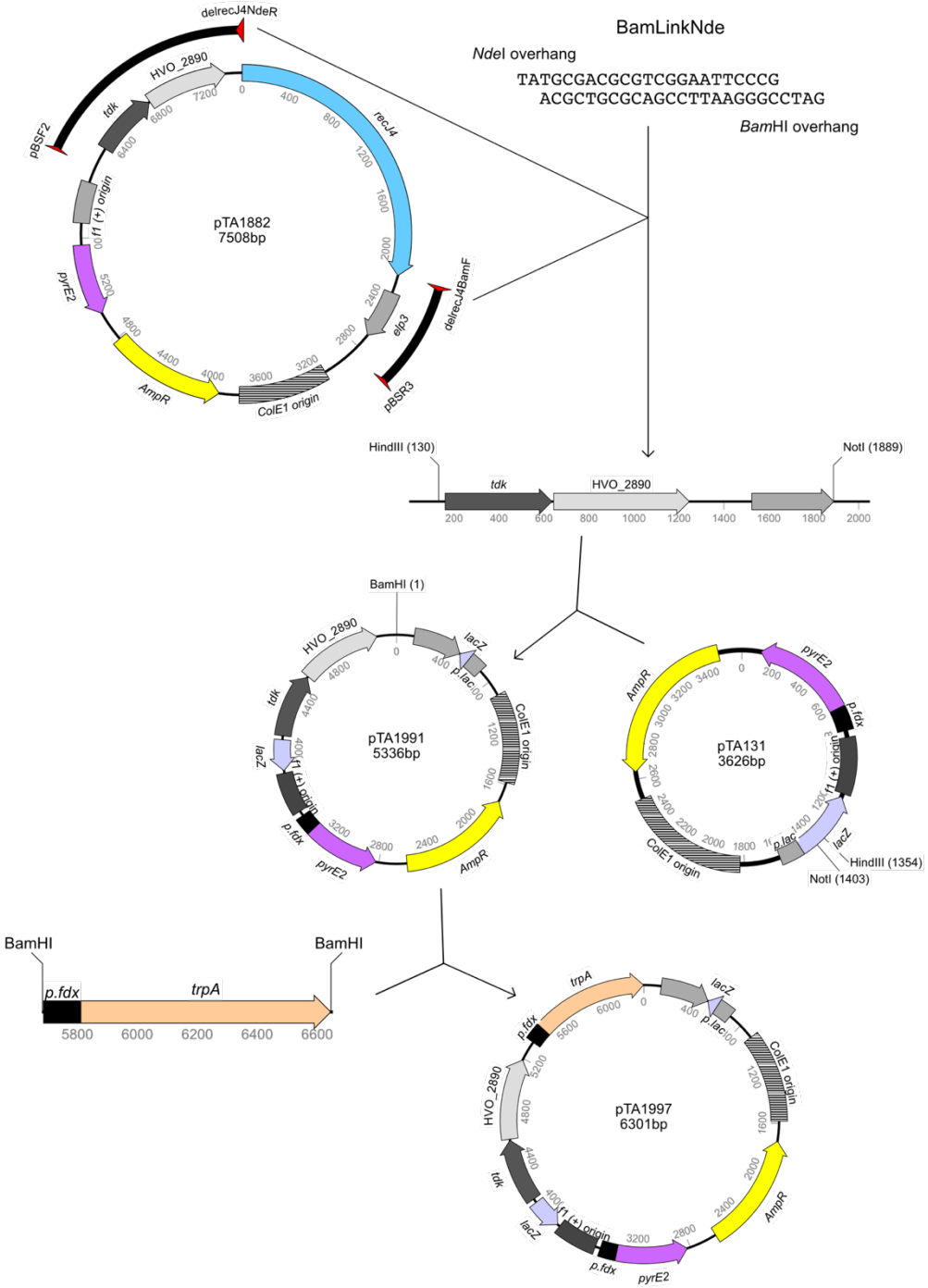


Figure 6.12 **Generation of *recJ4::trpA*+ deletion construct, pTA1997.** pTA1882 was used as a template to PCR amplify pBSF2 to delrecJ4NdeR and pBSR3 to delrecJ4BamF, which were digested with *Nde*I and *Bam*HI respectively and ligated via the BamLinkNde linker. The USlinkDS fragment was digested with *Not*I and *Hind*III and ligated into the *Not*I and *Hind*III sites of pTA131 to generate the *recJ4* deletion construct, pTA1991. The *trpA* marker digested from pTA298 using *Bam*HI was ligated into the *Bam*HI site of pTA1991 to generate the *recJ4::trpA*+ deletion construct, pTA1997.

Generation of *recJ* Deleted Strains

All of the *recJ* deletion constructs, unmarked and *trpA+* marked, were first introduced individually into H164 (*ΔpyrE2, bgaHa-Bb, leuB-Ag1, ΔtrpA).*

The *recJ1* and *recJ1::trpA+* deletion constructs, pTA1924 and pTA1958 respectively, were individually introduced into H164 (*ΔpyrE2, bgaHa-Bb, leuB-Ag1, ΔtrpA)* to generate the pop-in strains H3893 and H3901 respectively. Strains were grown over three consecutive overnights in YPC to allow the pop-out to occur. Strains were plated on 5-FOA in order to select for the pop-out of pTA1924/pTA1958. Colonies were patched and screened via colony hybridisation using a *recJ1+* radiolabelled probe to check for *recJ1* deletions.

The *recJ2* and *recJ2::trpA+* deletion constructs, pTA1951 and pTA1996 respectively, were individually introduced into H164 (*ΔpyrE2, bgaHa-Bb, leuB-Ag1, ΔtrpA)* to generate the pop-in strains H3895 and H3903 respectively. Strains were grown over three consecutive overnights in YPC to allow the pop-out to occur. Strains were plated on 5-FOA in order to select for the pop-out of pTA1951/pTA1996. Colonies were patched and screened via colony hybridisation using a *recJ2+* radiolabelled probe to check for *recJ2* deletions.

The *recJ3* and *recJ3::trpA+* deletion constructs, pTA1927 and pTA1960 respectively, were individually introduced into H164 (*ΔpyrE2, bgaHa-Bb, leuB-Ag1, ΔtrpA)* to generate the pop-in strains H3897 and H3905 respectively. Strains were grown over three consecutive overnights in YPC to allow the pop-out to occur. Strains were plated on 5-FOA in order to select for the pop-out of pTA1927/pTA1960. Colonies were patched and screened via colony hybridisation using a *recJ3+* radiolabelled probe to check for *recJ3* deletions.

The *recJ4* and *recJ4::trpA+* deletion constructs, pTA1991 and pTA1997 respectively, were individually introduced into H164 (*ΔpyrE2, bgaHa-Bb, leuB-Ag1, ΔtrpA)* to generate the pop-in strains H3899 and H3907 respectively. Strains were grown over three consecutive overnights in YPC to allow the pop-out to occur. Strains were plated on 5-FOA in order to select for the pop-out of pTA1991/pTA1997. Colonies were patched and screened via colony hybridisation using a *recJ4+* radiolabelled probe to check for *recJ4* deletions.

The strains H3929 (Δ*recJ1::trpA+*), H3930 (Δ*recJ2::trpA+*), H3931 (Δ*recJ3::trpA+*), and H3932 (Δ*recJ4::trpA+*) were verified via Southern blot. All unmarked *recJ* deletions maintained a wildtype genotype, thus were unsuccessful and not assigned strain numbers (Figure 6.13).

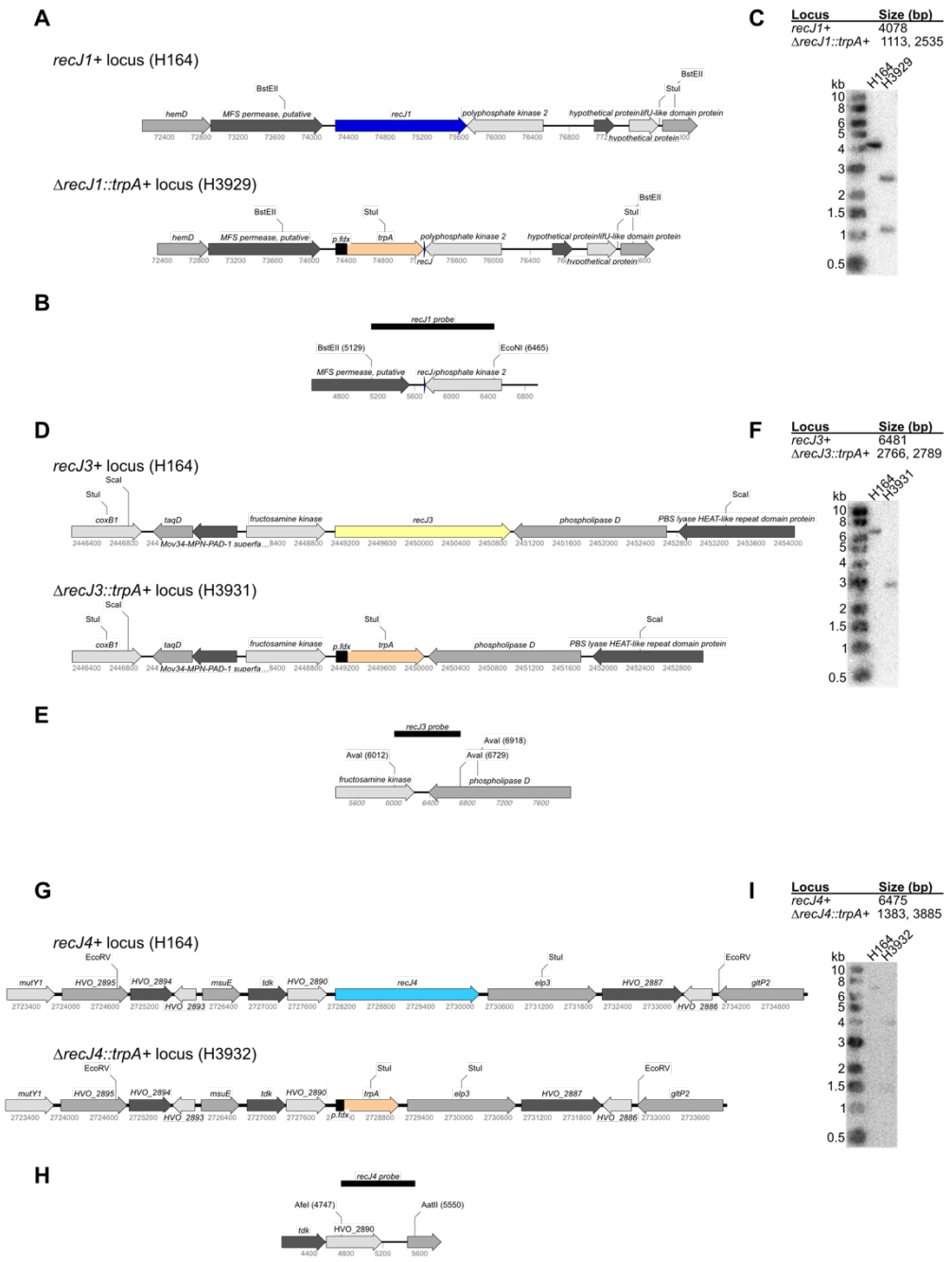


Figure 6.13 **Deletion of *recJ1, recJ3,* and *recJ4* from H164 (WT).** A) Gene locus for *recJ1*+ and Δ*recJ1::trpA*+. The *Bst*EII and *Stu*I sites used to digest genomic DNA are shown. B) A 1336bp Δ*recJ1* probe for southern hybridisation was made by digesting pTA1924 with *Bst*EII and *Eco*NI. C) Predicted locus size and Southern blot, showing H3929 is deleted for *recJ1*. D) Gene locus for *recJ3*+ and Δ*recJ3::trpA*+. The *Sca*I and *Stu*I sites used to digest genomic DNA are shown. E) A 717bp Δ*recJ3* probe for southern hybridisation was made by digesting pTA1927 with *Ava*I. F) Predicted locus size and Southern blot, showing H3931 is deleted for *recJ3*. G) Gene locus for *recJ4*+ and Δ*recJ4::trpA*+. The *Eco*RV and *Stu*I sites used to digest genomic DNA are shown. H) A 803bp Δ*recJ4* probe for southern hybridisation was made by digesting pTA1991 with *Afe*I and *Aat*II. I) Predicted locus size and Southern blot, showing H3932 is deleted for *recJ4*.

All unmarked deletions were unsuccessful therefore, all further *recJ* deletions were attempted using the *trpA+* marked constructs only. Typically the genotype of Δ*recJ2* candidate strains, generated using either construct, contained a mixed genotype of deleted and wildtype genes. This phenomena is known as merodiploid. Therefore, a Δ*recJ2* deletion strain was not obtained during this study.

Generation of Double and Triple Deletion Strains

In order to obtain the double and triple deletion strains, sequential deletions of *recJ1, recJ3,* and *recJ4* were made. First the *trpA+* marker was removed from the strain using the unmarked Δ*recJ* construct via the pop-in/pop-out method (Figure 6.14). A different Δ*recJ::trpA+* construct was then introduced to the resultant strain. The order in which genes are deleted in *H. volcanii* often affects the ease of the deletion process and the viability of the strain therefore, several sequential deletions were attempted varying the order in which the *recJ* genes were deleted. Strains were verified via colony lift and hybridisation (Figure 6.15) and by Southern blot (Figure 6.16. See Figure 6.13 for probe details).

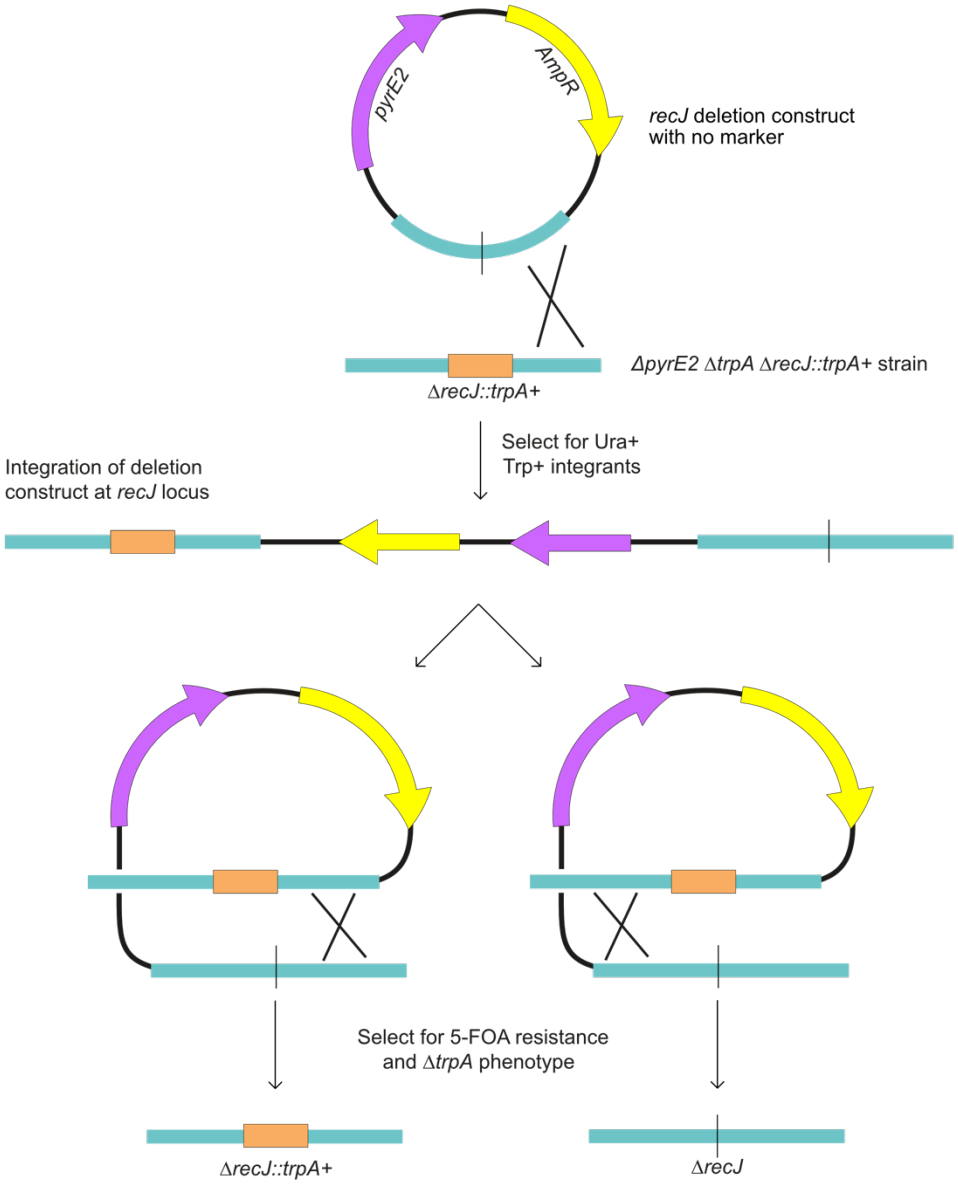


Figure 6.14 **Removal of the *trpA*+ gene marker via pop-in/pop-out**. Δ*pyrE2* Δ*recJ::trpA*+ strains are transformed with a *pyrE2*+ unmarked *recJ* deletion construct. Pop-ins are selected for by their ability to grow on media lacking uracil. Relieving uracil selection allows for pop-out event to occur, recombination between the regions of homology can either be upstream or downstream. Pop-out is selected for by plating on 5-FOA. The resulting gene locus will either be an unmarked deletion (Δ*trpA)* or a *trpA*+ marked deletion (*trpA*+). Pop-outs can be screened by *trpA*+ phenotype.

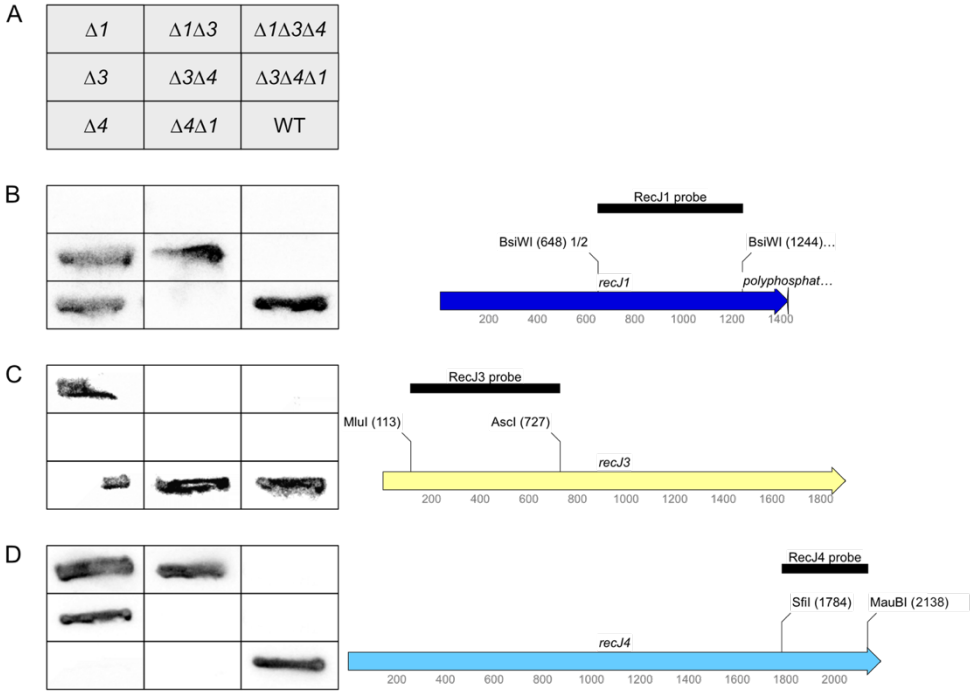


Figure 6.15 **Colony hybridisation verification of strains sequentially deleted for *recJ1*, *recJ3*, and *recJ4*.** A) Summary of the deleted *recJ* genes. B) The 596bp Δ*recJ1* probe for colony hybridisation was made by digesting pTA1887 with *Bsi*WI, and the corresponding hyb. C) The 614bp Δ*recJ3* probe for colony hybridisation was made by digesting pTA1889 with *Asc*I and *Mlu*I, and the corresponding hyb. D) The 354bp Δ*recJ4* probe for colony hybridisation was made by digesting pTA1882 with *Mau*BI and *Sfi*I, and the corresponding hyb.

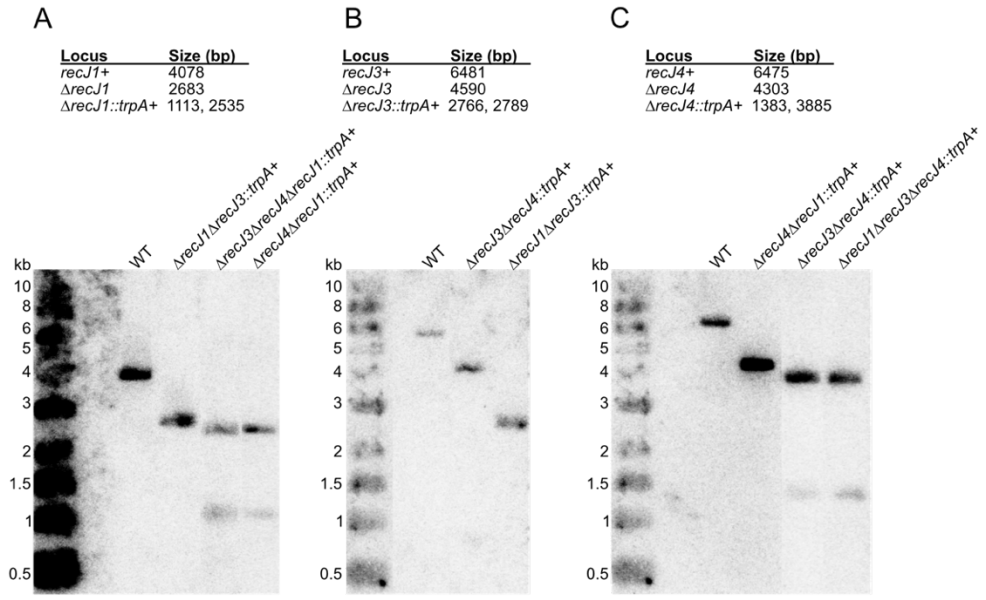


Figure 6.16 **Southern blot verification of strains sequentially deleted for *recJ1*, *recJ3*, and *recJ4*.** A) Predicted locus size and Southern blot using 1336bp Δ*recJ1* probe for southern hybridisation made by digesting pTA1924 with *Bst*EII and *Eco*NI as shown in figure 6.13. B) Predicted locus size and Southern blot using the 717bp Δ*recJ3* probe for southern hybridisation made by digesting pTA1927 with *Ava*I as shown in figure 6.13. C) Predicted locus size and Southern blot using the 803bp Δ*recJ4* probe for southern hybridisation made by digesting pTA1991 with *Afe*I and *Aat*II.

Attempt to generate recJ2 Deletion Strains

The length of the upstream and downstream flanking regions of *recJ2* are asymmetric in the *recJ2::trpA+* deletion construct, pTA1996, at 679 bp and 1403 bp respectively. Due to the longer length of homology, a recombination event is more likely to occur within the downstream flanking region. To avoid downstream recombination events occurring at both the pop-in and pop-out stages, causing the strain to revert to a *recJ2+* allele, colonies were screened via colony PCR to identify an upstream recombination event occurring at the pop-in stage.

$$P(US) \approx 0.33$$
$$P(DS) \approx 0.67$$

$$0.67^n \leq 0.05$$

$$nlog_{10}(0.67) = log_{10}(0.05)$$

$$n = \frac{log_{10}(0.05)}{log_{10}(0.67)}$$

$$n = 7.48$$

P(US) = the probability of an upstream recombination event
P(DS) = the probability of a downstream recombination event
n = the number of colonies

Therefore, screening 8 pop-in colonies would give >95% chance of obtaining at least one upstream recombination event.

Two colony PCR screens were employed both using the primer RecJ2_F, which binds outside of the upstream flanking region, with either RecJ2probeR or TRPAR which bind within the *recJ2* and *trpA* genes respectively (Figure 6.17). An upstream recombination event is identified by a PCR product of ~2.2 kb produced with the TRPAR primer and the absence of product with the RecJ2probeR primer. A downstream recombination event is identified by a PCR product of ~2.1 kb produced with the RecJ2probeR primer and the absence of product with the TRPAR primer. Merodiploid colonies yield products from both PCR reactions.

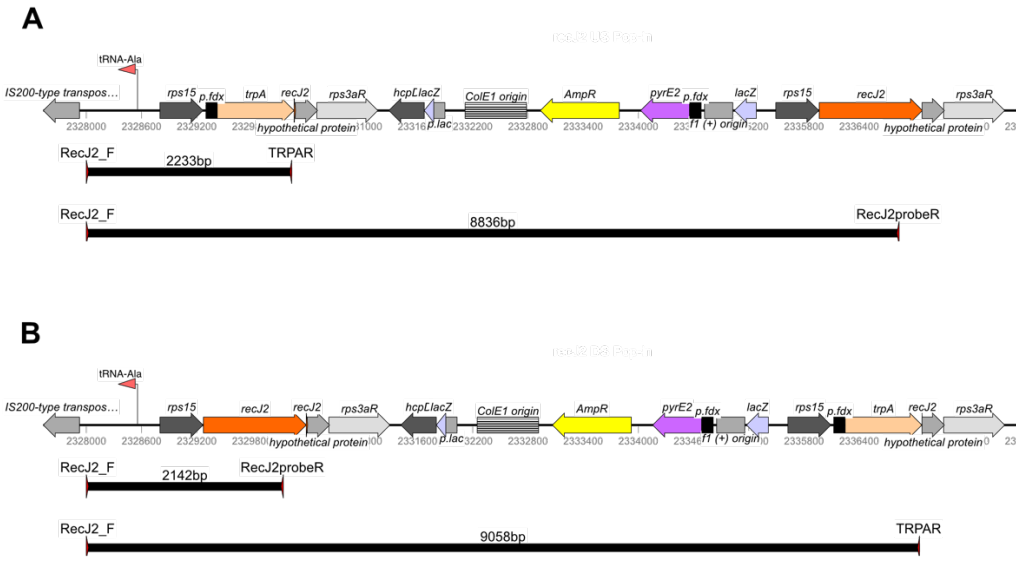


Figure 6.17 **PCR screen for upstream pop-in of pTA1996 at RecJ2 locus.** A) Upstream pop-in gene region. PCR from RecJ2_F to TRPAR and RecJ2probeR producing 2233bp and 8836bp respectively. B) Downstream pop-in gene region. PCR from RecJ2_F to TRPAR and RecJ2probeR producing 9058bp and 2142bp respectively. The 8836bp and 9058bp products are not observed due to the elongation time during the reaction.

Once an upstream pop-in strain had been identified, pop-out colonies where obtained and screened via colony hybridisation, 132 and 148 colonies for the wildtype and triple-delete (Δ*recJ1* Δ*recJ3* Δ*recJ4*) backgrounds respectively. All colonies screened were *recJ2*+. Around 20 colonies were further analysed via Southern blot revealing that all strains were merodiploid (Figure 6.18).

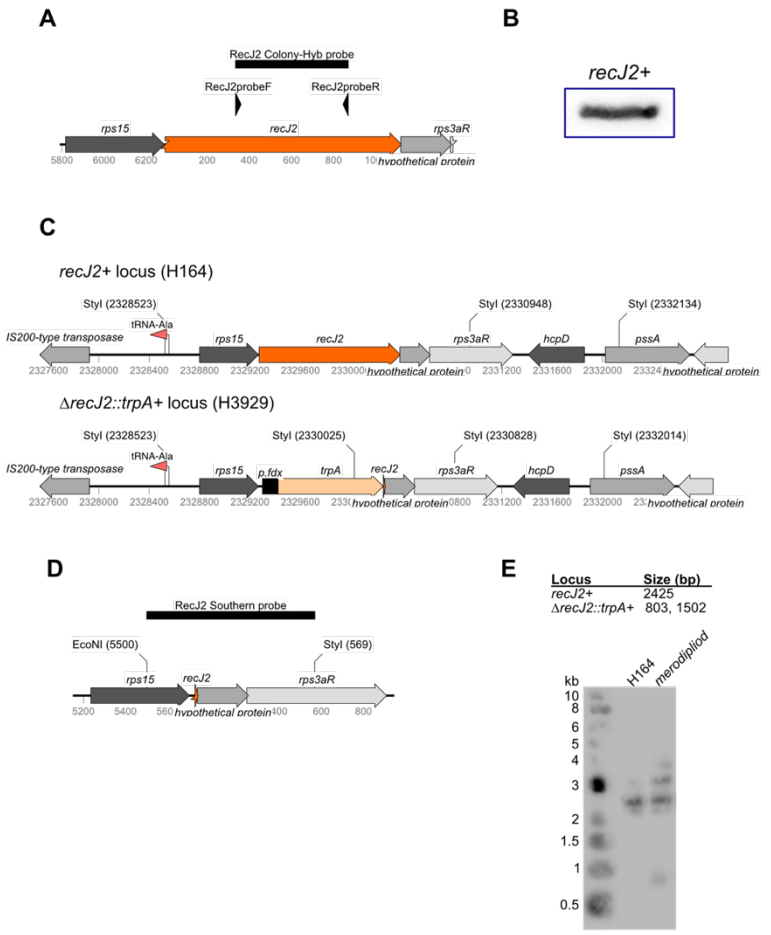


Figure 6.18 **Attempted deletion of *recJ2* from H164.** A) The 535bp Δ*recJ2* probe for colony hybridisation was made using pTA1888 as a template for PCR from RecJ2probeF to RecJ2probeR. B) Colony hybridisation showing the colony is *recJ2*+. C) Gene locus for *recJ2*+ and Δ*recJ2::trpA*+. The *Sty*I sites used to digest genomic DNA are shown. D) A 798bp Δ*recJ2* probe for southern hybridisation was made by digesting pTA1951 with *Eco*NI and *Sty*I. E) Predicted locus size and Southern blot, showing a strain that is merodiploid.

pTA2335 (*ginS* partial deletion construct maintaining promoter region for *bcp1, trpA*+)

Due to the close proximity of the downstream *bcp1* gene to the *ginS* gene, a construct that deletes the entirety of the *ginS* gene would also delete the promoter region for *bcp1*. Thus, a strain deleted for *ginS* would also be deleted for *p.bcp1*.

To ensure the *bcp1* gene was still functional whilst simultaneously eliminating *ginS* function, a partial *ginS* deletion construct maintaining ~100 bp at the 3' end of *ginS*, which is suspected to contain the promoter region for *bcp1*, was designed (Figure 6.19).

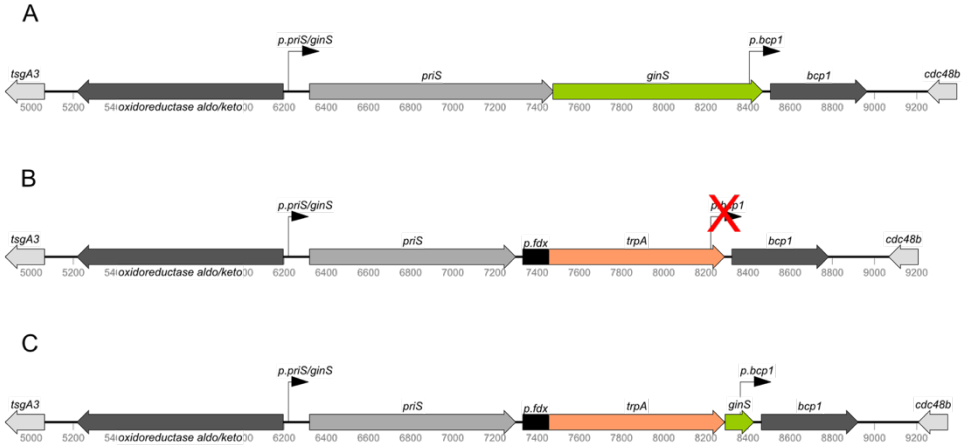


Figure 6.19 **Deletion of *ginS*.** A) Gene locus for *ginS*+ indicating the predicted promoter regions for *p.priS/ginS* and *p.bcp1*. B) Gene locus for the clean deletion: Δ*ginS::trpA*+ indicating the predicted promoter region for *p.priS/ginS* and the lost promoter *p.bcp1*. C) Gene locus for the partial deletion: Δ*ginS::trpA*+, maintaining ~100bp of 3'-*ginS* and therefore the predicted promoter for *bcp1*. The predicted promoter regions for *p.priS/ginS* and *p.bcp1* are shown.

pTA1791 was digested with *Nde*I and *Bam*HI and the BamLinkNde oligo linker was ligated in place of *ginS* creating the intermediate construct, pTA2184. pTA1716 was used as a template to PCR amplify the 3' region of *ginS* CDS using the primers ginCtermBamF and gintagDSXbaR introducing *Bam*HI and *Xba*I sites respectively. The product was digested with *Xba*I and *Bam*HI and ligated into the *Xba*I and *Bam*HI sites of pTA2184, to create the plasmid, pTA2315. pTA298 used as a template to PCR amplify *trpA* CDS and the *ferredoxin* promoter using the primers TrpABglF and TrpABglR introducing *Bgl*II sites either end on the *trpA* marker. The PCR product was digested with *Bgl*II and ligated into the *Bam*HI site of pTA2315. The ligated construct was digested with *Bam*HI to eliminate any self-ligated vector. Restriction digest was used to screen for orientation of the *trpA* marker, the same orientation as *ginS* was preferable, and was further confirmed via DNA sequencing to generate the *ginS::trpA*+ deletion construct, pTA2335 (Figure 6.20).

Due to time constraints, the deletion of *ginS* from *H. volcanii* using pTA2315 or pTA2335 was attempted by Victoria Smith (second year PhD student). As expected, deletion was not successful and all candidates remained positive for the *ginS* gene.

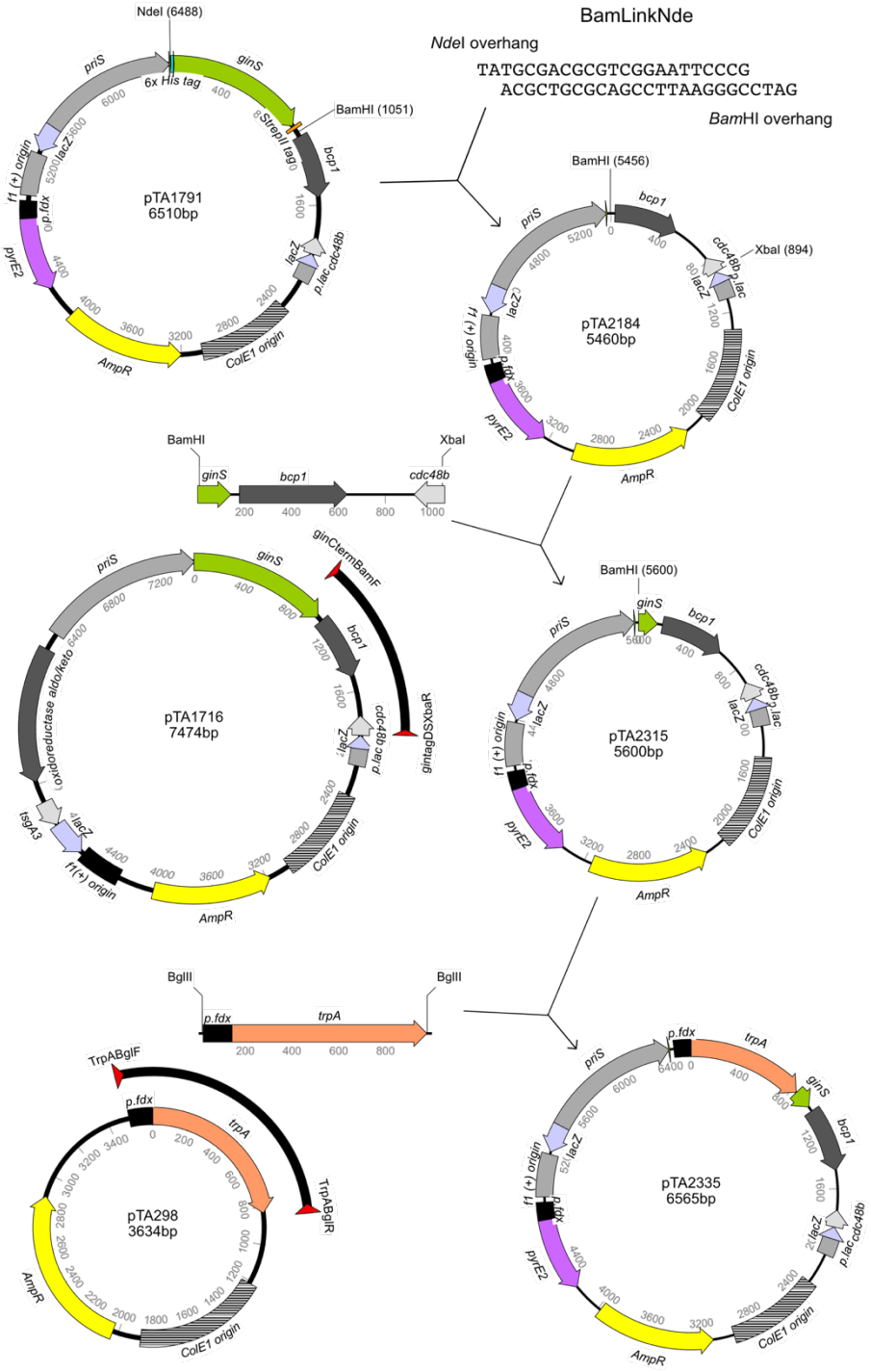


Figure 6.20 **Generation of *ginS::trpA+* deletion construct, pTA2335.** The BamLinkNde oligo linker was ligated into the *Nde*I and *Bam*HI sites of pTA1791 generating the intermediate construct, pTA2184. pTA1716 was used as a template to PCR amplify ginCtermBamF to gintagDSXbaR introducing *Bam*HI and *Xba*I sites respectively. The product was digested with *Bam*HI and *Xba*I and ligated into the *Bam*HI and *Xba*I sites of pTA2184, to create the Δ*ginS* deletion construct, pTA2315. pTA298 used as a template to PCR amplify TrpABglF to TrpABglR introducing *Bgl*II sites either end on the *trpA* CDS. The PCR product was digested with *Bgl*II and ligated into the *Bam*HI site of pTA2315, to generate the *ginS::trpA+* deletion construct, pTA2335.

6.4.3 Characterisation of *recJ* Deleted Strains

Growth Rate

The standard growth rates in liquid media were quantified for the viable *recJ* deleted strains. Strains were grown over two consecutive overnights to form a vigorously growing culture. The exponential phase of cell growth was calculated using an Epoch2 Microplate Spectrophotometer (BioTek) by measuring the A_{600} of the culture at 15-minute intervals over a 72-hour time course. A growth curve was plotted and the generation time during exponential phase for each strain was calculated (Figure 6.21).

The single *recJ*-deleted strains H3929 (Δ*recJ1::trpA*+), H3931 (Δ*recJ3::trpA*+), and H3932 (Δ*recJ4::trpA*+), and the double *recJ*-deleted strains H4170 (Δ*recJ1* Δ*recJ3::trpA*+), and H4173 (Δ*recJ3* Δ*recJ4::trpA*+), all had a slightly increased generation time compared with the wildtype strain, H164, which has a generation time of 3h. H3929 (Δ*recJ1::trpA*+), H3932 (Δ*recJ4::trpA*+), and H4170 (Δ*recJ1* Δ*recJ3::trpA*+) all had the generation time of 4.25h, whilst H3931 (Δ*recJ3::trpA*+), and H4173 (Δ*recJ3* Δ*recJ4::trpA*+) had a generation time of 3.75h. Conversely, the double delete strain H4174 (Δ*recJ4* Δ*recJ1::trpA*+), and the triple delete strains H4275 (Δ*recJ3* Δ*recJ4* Δ*recJ1::trpA*+), and H4376 (Δ*recJ1* Δ*recJ3* Δ*recJ4*) all had comparable generation times to H164 (WT) with generation times of 3.25h, 3h, and 3.25h respectively.

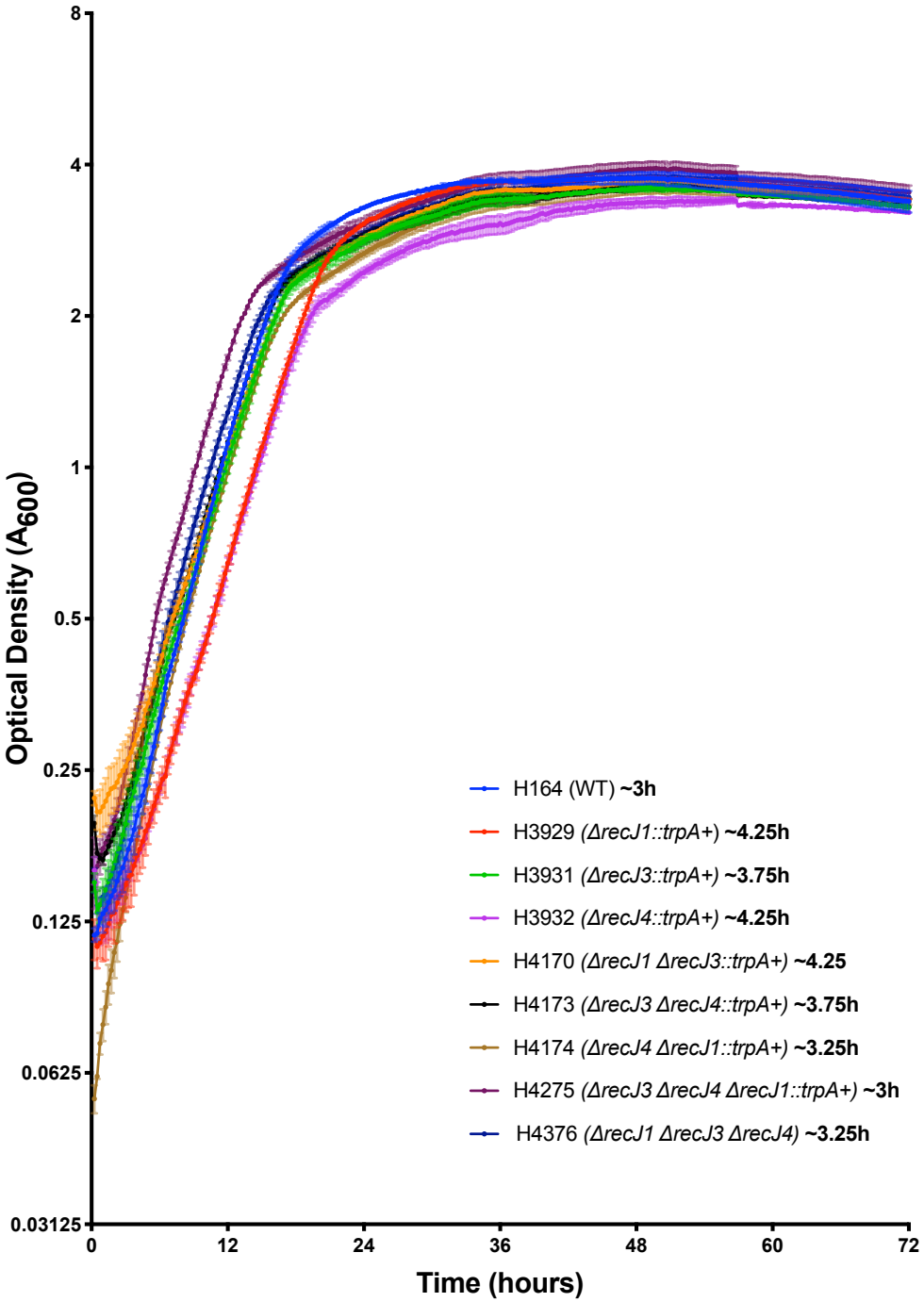


Figure 6.21 **Exponential growth rate of *recJ1, recJ3,* and *recJ4* delete strains.** H164 (*recJ*+), H3929 (Δ*recJ1::trpA*+), H3931 (Δ*recJ3::trpA*+), H3932 (Δ*recJ4::trpA*+), H4170 (Δ*recJ1* Δ*recJ3::trpA*+), H4173 (Δ*recJ3* Δ*recJ4::trpA*+), H4174 (Δ*recJ4* Δ*recJ1::trpA*+), H4275 (Δ*recJ3* Δ*recJ4* Δ*recJ1::trpA*+), and H4376 (Δ*recJ1* Δ*recJ3* Δ*recJ4*), were grown in liquid culture for two consecutive overnights. The strains were then diluted and grown for 72 hours. Growth was measured by A_{600}. Generation time is indicated next to each strain. All strains were incubated on the same 96-well plate and measured simultaneously using an Epoch2 Microplate Spectrophotometer (BioTek). Data points are a mean of two technical repeats within two biological repeats. Standard deviation is shown.

6.4.4 Development of Chromosomally-Tagged Expression Constructs

Chromosomally-tagged *recJ* expression constructs were generated in order to screen for *in vivo* protein:protein interactions of the four RecJ proteins expressed natively in *H. volcanii*. However, due to the time-consuming process of generating the *recJ* deletion strains, the appropriate *recJ* deletion strains for exploring *in vivo* protein:protein interactions were not made in this study. Thus, the chromosomally-tagged *recJ* expression constructs described below are yet to be introduced into the appropriate strain and screened for *in vivo* protein:protein interactions.

pTA2090 (*his$_6$-recJ1-strepII* gene replacement construct)

The *recJ1* genomic clone, pTA1887, was used as a template to PCR amplify the *recJ1* CDS using the primers RecJ1NcoI_F and RecJ1NheI_R introducing *Nco*I and *Nhe*I sites respectively. The product was digested with *Nco*I and *Nhe*I and ligated into the *Pci*I and *Nhe*I sites in pTA1667 to create the plasmid pTA2082. The *his$_6$-recJ1-strepII* CDS was digested from pTA2082 using *Nde*I and *Bam*HI and ligated into the *Nde*I and *Bam*HI sites in pTA1924 to create the *his$_6$-recJ1-strepII* gene replacement construct, pTA2090 (Figure 6.22).

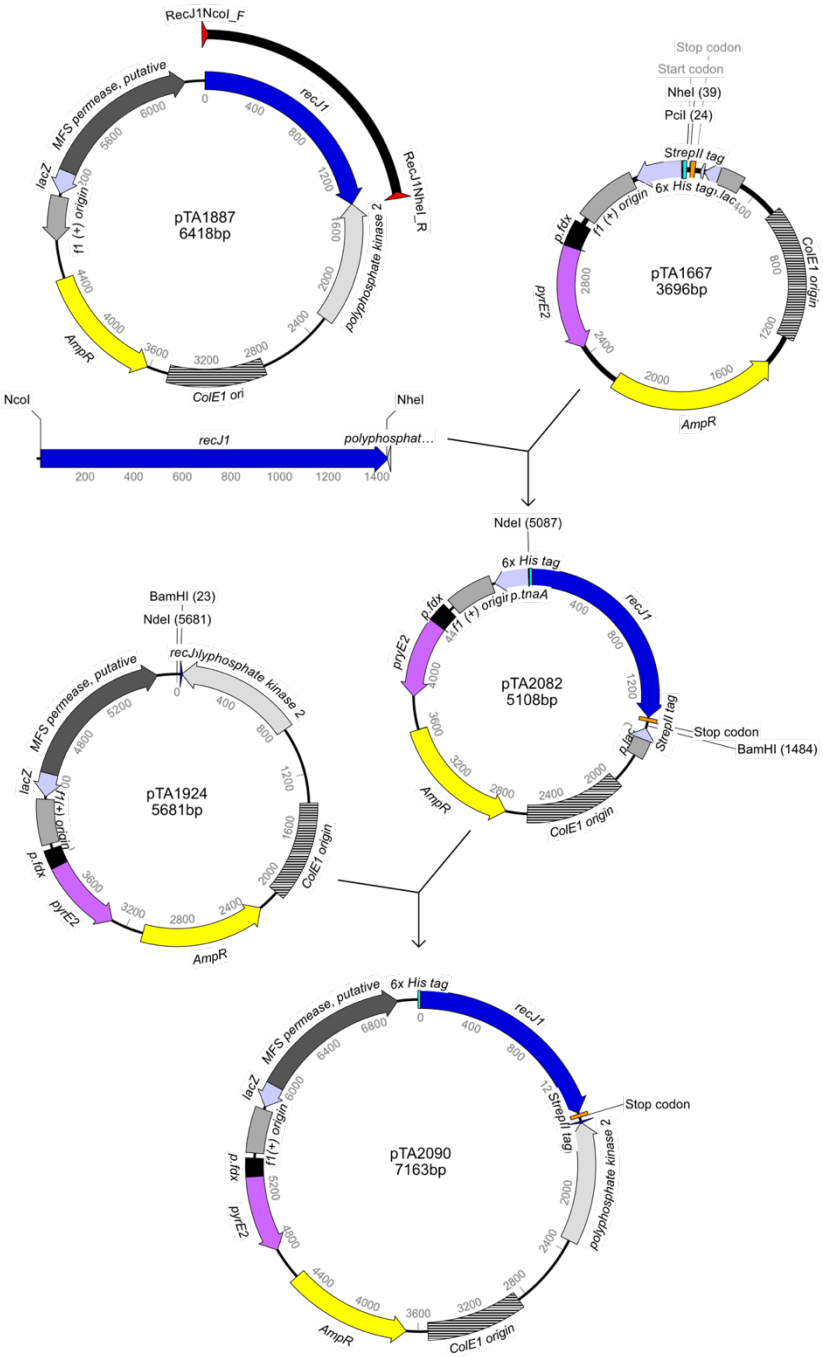


Figure 6.22 **Generation of *His$_6$-recJ1-StrepII* gene replacement construct, pTA2090.** pTA1887 was used as a template to PCR amplify RecJ1NcoI_F to RecJ1NheI_R introducing *Nco*I and *Nhe*I sites respectively. The product was digested with *Nco*I and *Nhe*I and ligated into the *Pci*I and *Nhe*I sites in pTA1667 to create the plasmid, pTA2082. pTA2082 was digested with *Nde*I and *Bam*HI and ligated into the *Nde*I and *Bam*HI sites in pTA1924 to create the *his$_6$-recJ1-strepII* gene replacement construct, pTA2090.

pTA2091 (*his$_6$-recJ2-strepII* gene replacement construct)

The *recJ2* genomic clone, pTA1888, was used as a template to PCR amplify the *recJ2* CDS using the primers RecJ2PciI_F and RecJ2NheI_R introducing *Pci*I and *Nhe*I sites respectively. The product was digested with *Pci*I and *Nhe*I and ligated into the *Pci*I and *Nhe*I sites in pTA1667 to create the plasmid, pTA2083. The *his$_6$-recJ2-strepII* CDS was digested from pTA2083 using *Nde*I and *Bam*HI and ligated into the *Nde*I and *Bgl*II sites in pTA1951 to create the *his$_6$-recJ2-strepII* gene replacement construct, pTA2091 (Figure 6.23).

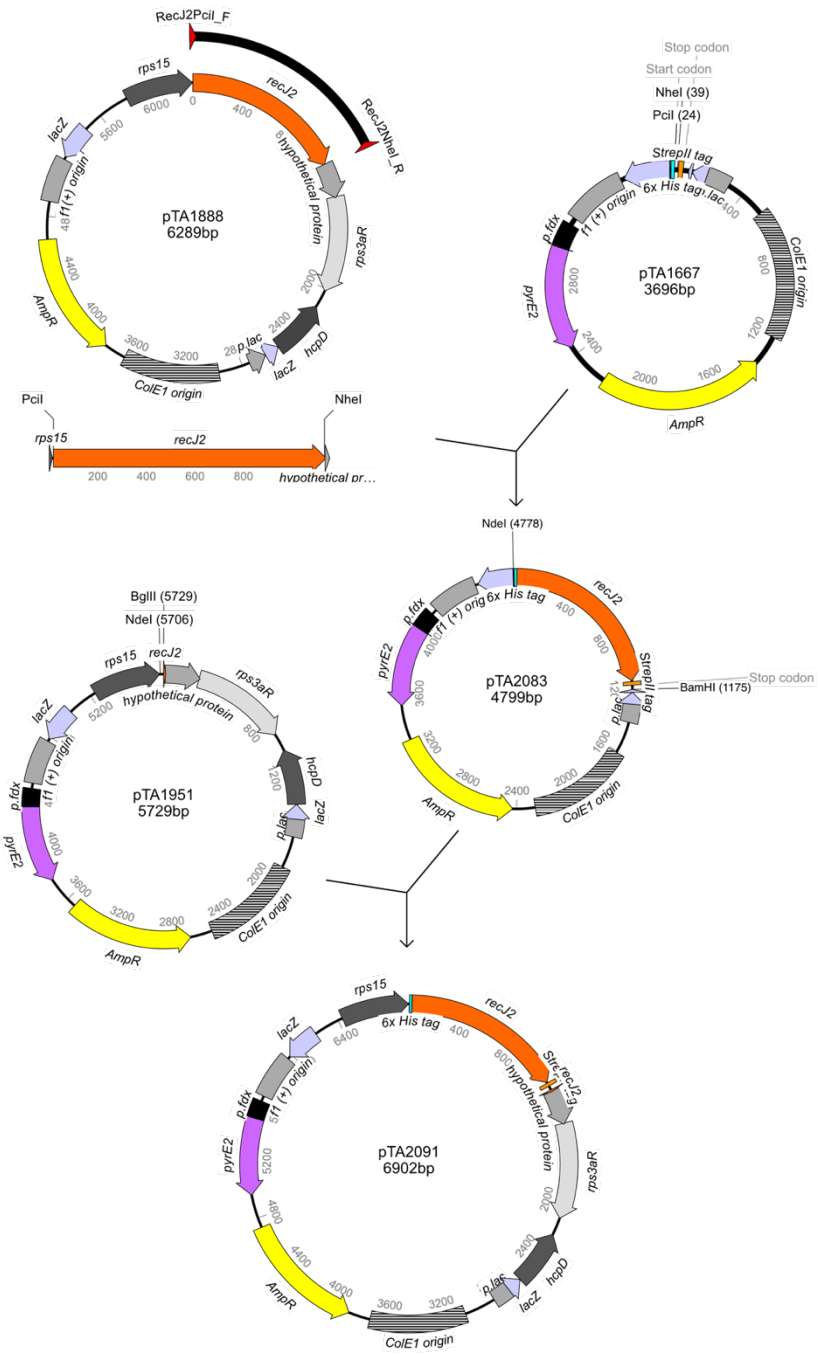


Figure 6.23 **Generation of *His$_6$-recJ2-StrepII* gene replacement construct, pTA2091.** pTA1888 was used as a template to PCR amplify RecJ2PciI_F to RecJ2NheI_R introducing *Pci*I and *Nhe*I sites respectively. The product was digested with *Pci*I and *Nhe*I and ligated into the *Pci*I and *Nhe*I sites in pTA1667 to create the plasmid, pTA2083. pTA2083 was digested with *Nde*I and *Bam*HI and ligated into the *Nde*I and *Bgl*II sites in pTA1951 to create the *his$_6$-recJ2-strepII* gene replacement construct, pTA2091.

pTA2095 (*his$_6$-recJ4-strepII* gene replacement construct)

The *recJ4* genomic clone, pTA1882, was used as a template to PCR amplify the *recJ4* CDS using the primers RecJ4NcoI_F and RecJ4NheI_R introducing *Nco*I and *Nhe*I sites respectively. The product was digested with *Nco*I and *Nhe*I and ligated into the *Pci*I and *Nhe*I sites in pTA1667 to create the plasmid, pTA2085. The *his$_6$-recJ4-strepII* CDS was digested from pTA2085 using *Nde*I and *Bam*HI and ligated into the *Nde*I and *Bam*HI sites in pTA1991 to create the *his$_6$-recJ4-strepII* gene replacement construct, pTA2095 (Figure 6.24).

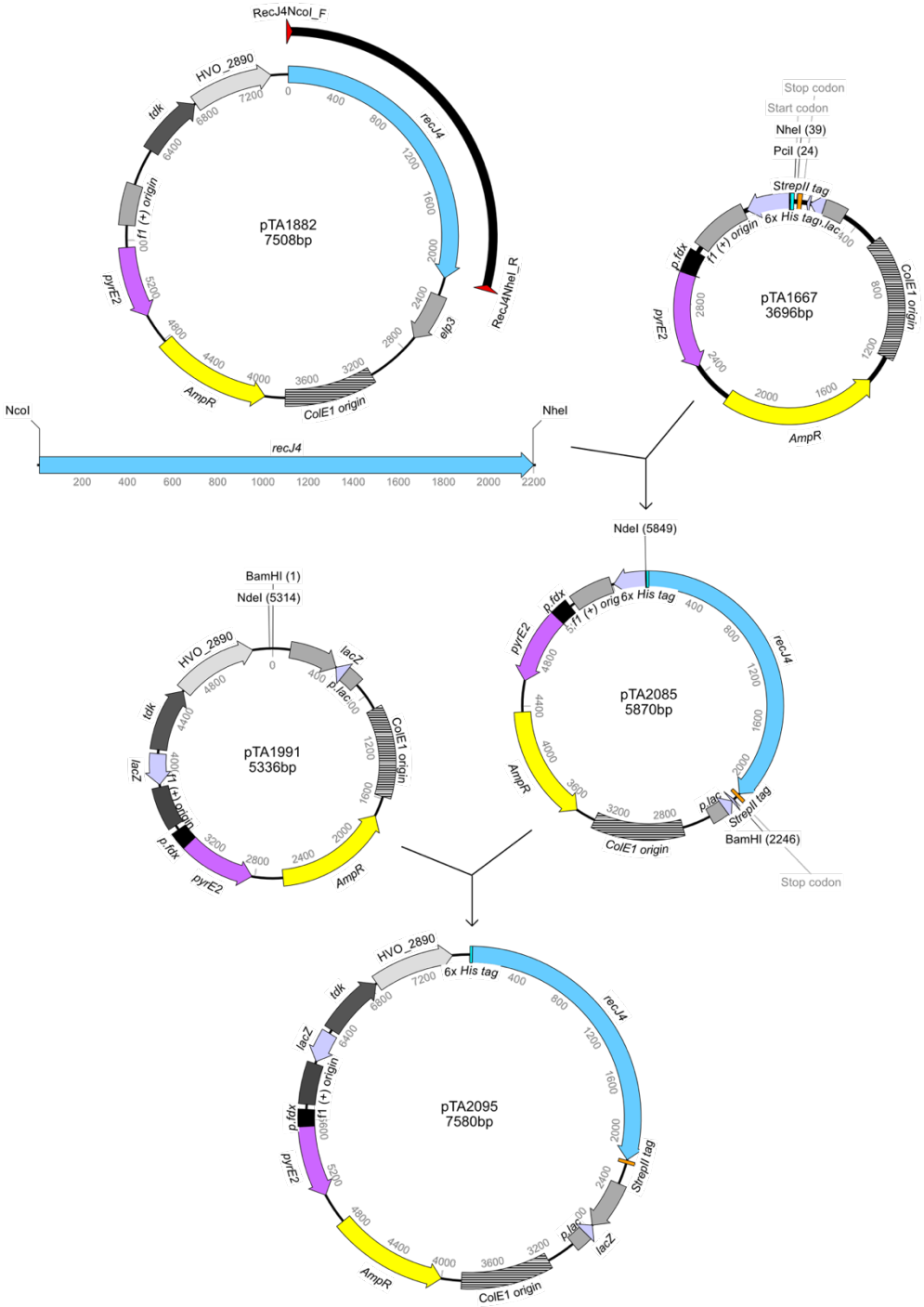


Figure 6.24 **Generation of *His$_6$-recJ4-StrepII* gene replacement construct, pTA2095.** pTA1882 was used as a template to PCR amplify RecJ4NcoI_F to RecJ4NheI_R introducing *Nco*I and *Nhe*I sites respectively. The product was digested with *Nco*I and *Nhe*I and ligated into the *Pci*I and *Nhe*I sites in pTA1667 to create the plasmid, pTA2085. pTA2085 was digested with *Nde*I and *Bam*HI and ligated into the *Nde*I and *Bam*HI sites in pTA1991 to create the *his$_6$-recJ4-strepII* gene replacement construct, pTA2095.

pTA2130 (Intermediate to generate the *his$_6$-recJ3-strepII* gene replacement construct)

The *recJ3* genomic clone, pTA1889, was used as a template to PCR amplify the *recJ3* CDS using the primers RecJ3BspHI_F and RecJ3NheI_R introducing *Bsp*HI and *Nhe*I sites respectively. The product was digested with *Bsp*HI and *Nhe*I and ligated into the *Pci*I and *Nhe*I sites in pTA1667 to create the plasmid, pTA2084. The downstream flanking region of *recJ3* was digested from pTA1927 using *Bgl*II and *Not*I and was ligated into the *Bam*HI and *Not*I sites in pTA2084, to create the *his$_6$-recJ3-strepII* gene replacement intermediate construct, pTA2130. This study was unsuccessful in cloning the upstream flanking region into pTA2130 to generate the complete *his$_6$-recJ3-strepII* gene replacement construct (Figure 6.25).

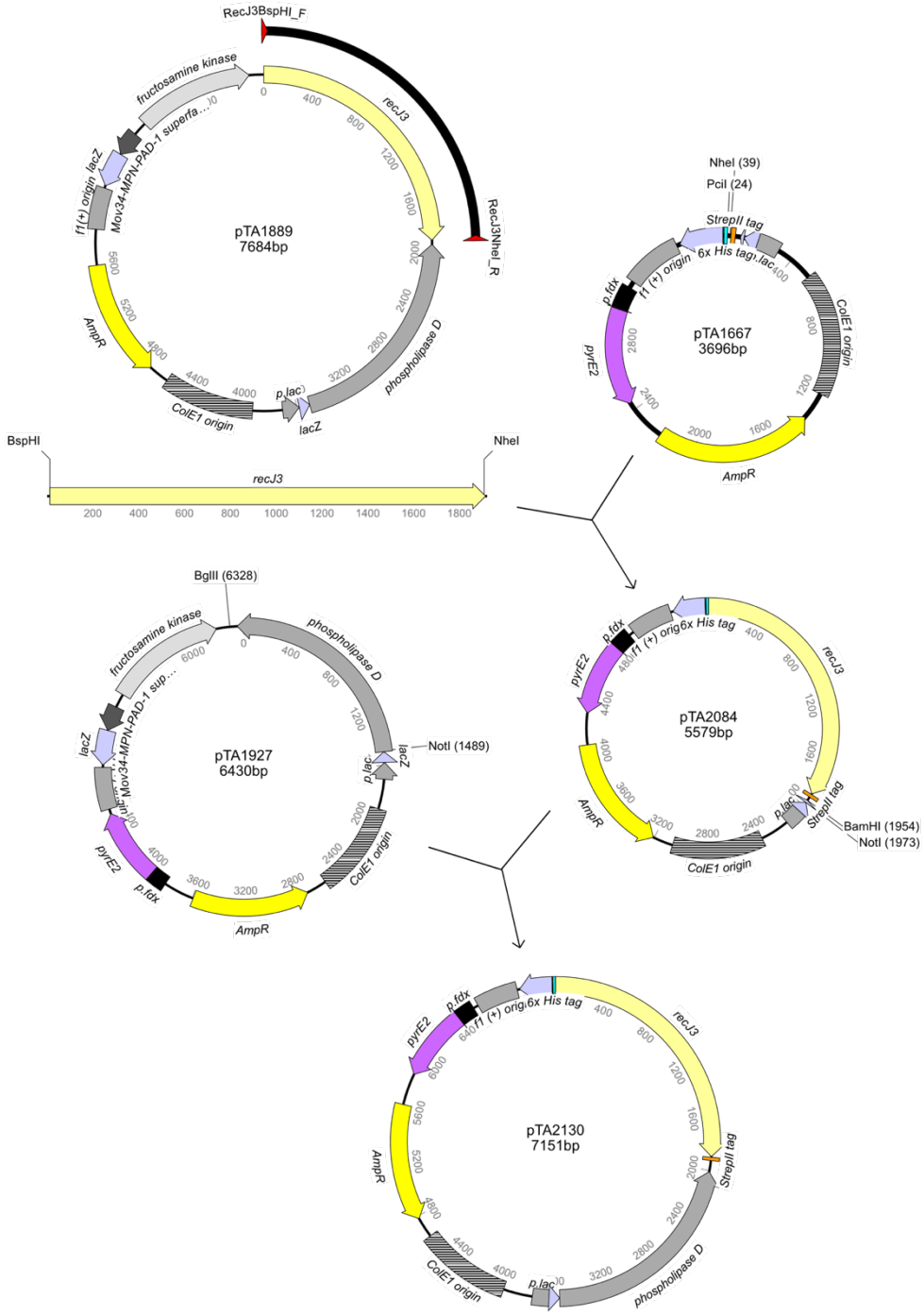


Figure 6.25 **Generation of the *His$_6$-recJ3-StrepII* gene replacement intermediate construct, pTA2130.** pTA1889,was used as a template to PCR amplify RecJ3BspHI_F to RecJ3NheI_R introducing *Bsp*HI and *Nhe*I sites respectively. The product was ligated into the *Pci*I and *Nhe*I sites in pTA1667 to create the plasmid, pTA2084. pTA1927 was digested using *Bgl*II and *Not*I and was ligated into the *Bam*HI and *Not*I sites in pTA2084, to create the *his$_6$-recJ3-strepII* gene replacement intermediate construct, pTA2130.

6.5 Discussion

RecJ2 is essential and therefore most likely the Cdc45 homologue?

Genetic analysis has revealed that not only is it possible to delete *H. volcanii recJ1, recJ3, and recJ4* individually, it is also possible to make the double and triple deletions with no fitness defect. However, this study failed to delete *recJ2* from *H. volcanii* in either a wildtype (H164) or the triple-*recJ*-delete (H4376) genetic backgrounds; multiple attempts were made, screening 132 and 148 colonies respectively. Around 20 colonies were further analysed via Southern blot revealing that all strains were merodiploid. These data suggest that *recJ2* is essential with a certainty of ~85% for the wildtype and triple-*recJ*-delete genetic backgrounds. Certainty was calculated based on the difficulty to delete *radA* in the Δ*oriC1* Δ*oriC2* Δ*oriC3* background (Table 6.5). To date, this is the most difficult gene/background combination attempted for deletion.

Table 6.4 **Difficulty to generate a Δ*radA* in strains deleted for origins**

Genotype	**WT**	**Δ*oriC1***	**Δ*oriC2***	**Δ*oriC3***	**Δ*oriC1* Δ*oriC2* Δ*oriC3***	**Δ*oriC1* Δ*oriC2* Δ*oriC3* Δ*ori-pHV4***
Δ*radA* fraction	62/66	66/66	19/44	9/73	1/70	0/455

radA fraction indication the number of successful transformations to generate a Δ*radA* phenotype over the number of assayed colonies. Deleting the *radA* gene becomes increasingly difficult the more origins that are deleted. When all origins are deleted (Δ*oriC1* Δ*oriC2* Δ*oriC3* Δ*ori-pHV4*), *radA* is essential.

$$P(\Delta) = 1/70$$
$$P(+) = 69/70$$

$$\frac{69}{70} = 98.6\%$$

$$98.6^{132} = 0.150$$
$$98.6^{148} = 0.119$$

$$1 - 0.150 = 0.850$$
$$1 - 0.119 = 0.881$$

$$C \approx 85\%$$

$P(\Delta)$ = the probability of a *radA* delete
$P(+)$ = the probability of *radA+*
C = the certainty of a gene being essential

S. MacNeill states that *recJ2* is non-essential in the following publication (Giroux and Macneill, 2015) however, there is no published primary data to support this claim. Moreover, the MacNeill group verifies gene deletions and replacements via PCR only, which is not a robust method for polyploid organisms. PCR involves amplification, which gives a non-linear readout and fails to reflect the presence of minority alleles in a merodiploid polyploid genome. Therefore, the PCR method is non-quantitative and often fails to identify merodiploid cells, which can be mistaken for true gene deletions. Methods such colony hybridisation and Southern blotting do not involve amplification and can therefore reliably detect the presence of one wildtype allele in the presence of 19 deletion alleles. These methods are better suited at detecting minority alleles in a merodiploid polyploid genome however, caution should be taken and more than one method of strain verification should be used. The data in this study further reinforces the fact that any singular method is not sufficient for *Haloferax volcanii* strain verification.

Archaeal replication resembles its eukaryotic counterpart, for which the CMG complex is essential. The CMG complex is highly conserved throughout eukaryotes and archaea, comprising Cdc45, the replicative helicase MCM, and GINS (Makarova, Koonin and Kelman, 2012). Previous and current studies have shown that the replicative helicase, MCM, and GINS are essential for both origin-dependent and recombination-dependent replication in *H. volcanii* (Marriott, 2017). Therefore it would follow that the Cdc45 homologue is also essential. However, known Cdc45 homologues in a number of species, such as the *Thermococcus kodakaraensis gan* gene, are typically non-essential (Nagata,

Ishino, *et al.*, 2017). This suggests that the *H. volcanii* Cdc45 homologue would also be nonessential, providing evidence against RecJ2 as a candidate. It is possible that *recJ2* is nonessential and the difficulty to delete the gene can be explained by the chromosomal location. A substantial genetic screen (>60 halophiles) revealed that *recJ2* is part of a highly conserved, co-transcribed operon including the 30S ribosomal genes (Personal correspondence, Kira Makarova), which provides an alternative explanation to the essentiality of RecJ2.

The *H. volcanii* Cdc45 homologue will most likely be identified not by genetics but by completion of *in vivo* co-purification assays of GINS and the RecJ proteins under native expression levels.

RecJ1, 3, and 4 have a DNA repair role?

Domain analysis of the four *H. volcanii* (HVO) RecJ proteins yielded mixed results. Protein sequence alignment failed to reach a consensus, Pfam domain analysis identified HVO RecJ1 as most similar to TKO GAN, and Phyre2 predicted 3D structures for HVO RecJ1, RecJ3, and RecJ4 that closely resembled the known crystal structures of TKO GAN, TTH RecJ, and human Cdc45. However, the structure and domain composition of HVO RecJ2 remains a mystery. Phylogenetic analysis also indicates that RecJ2 is very different from not only the other three HVO RecJ proteins, but Euryarchaeota RecJ proteins in general. These results suggest that HVO RecJ2 is not the Cdc45 homologue and one of the other three are more likely candidates.

Furthermore, RecJ4 has been shown to interact *in vivo* with Hel308 (See chapter 4: Biochemical Analysis of Hel308 for further details) and Orc1 (Darya Ausannikava, University of Nottingham, personal communication). Orc1 is the replication initiation protein associated with ori-C1 suggesting a DNA replication role for RecJ4. Due to sequence and structural similarities between RecJ3 and RecJ4, it is probable that RecJ3 has a similar role to RecJ4 or that either protein is redundant. RecJ1 also shares a similar predicted 3D structure again suggesting a common function. However, RecJ4 differs from RecJ1 and RecJ3 in one importance aspect: it is predicted that RecJ1 and RecJ3 have maintained the exonuclease function whereas RecJ4 has lost this ability.

Hel308 is an anti-recombinase (See Chapter 3: Separation of function in Hel308 for further details), which is thought to unwind D-loops in order to promote an alternative repair pathway; it is possible that RecJ4 plays a role in the alternative repair pathway. Alternatively, if RecJ4 is the Cdc45 homologue, perhaps Hel308 does not fully unwind the D-loop but rather remodel the DNA structure into a replication fork, it would follow that Hel308 would recruit RecJ4 and by

extension the CMG complex, in order to carry out recombination-dependent replication.

Development of Chromosomally-Tagged Expression Constructs

Chromosomally-tagged *recJ1* and *recJ4* expression constructs to replace gene-deletions were successfully generated in order to screen for *in vivo* protein:protein interactions. A chromosomally-tagged *recJ2* expression gene replacement construct was also generated, the design of which may need to be reconsidered due to the lack of a Δ*recJ2* strain. This study failed to generate a complete *His_6-recJ3-StrepII* expression gene replacement construct. There is evidence that the C-terminus is important for RecJ3 function (Feng *et al.*, 2018), therefore the construct need to be redesigned without a tag at the 3' end of the gene. Due to the time-consuming process of generating the *recJ* deletion strains, the appropriate *recJ* deletion strains for exploring *in vivo* protein:protein interactions were not made in this study. Thus, the chromosomally-tagged *recJ* expression constructs are yet to be introduced into the appropriate Δ*recJ* strains and screened for *in vivo* protein:protein interactions.

6.6 Future Perspectives

Due to time constraints, full genetic and biochemical characterisation of the RecJ proteins was not possible. It is still unclear what the function and roles of the four RecJ proteins are and which, if any, is the Cdc45 homologue. Future study should aid in the deeper understanding of the RecJ proteins.

Which RecJ is the Cdc45 homologue?

In order to determine which of the four RecJ proteins is the Cdc45 homologue, it is of interest to screen for *in vivo* protein:protein interactions of GINS by expressing StrepII-tagged GINS natively in *H. volcanii* and analysing co-purifying proteins. This should identify which of the RecJ proteins associates with GINS and, by extension, helps form the CMG complex. Furthermore, it is of interest to complete the complementing *in vivo* co-purification assays by expressing the individual His_6-tagged and StrepII-tagged RecJ proteins natively in *H. volcanii* and analysing co-purifying proteins. This may give further insight into the cellular role of the individual RecJ proteins and confirm which RecJ interacts with GINS at a cellular level. Protein:protein interactions as determined by protein pull-down assays and mass spectrometry are described in Chapter 4: Biochemical analysis of Hel308.

Do RecJ proteins have a DNA repair role?

To determine the role of the other three RecJ proteins, interaction partners revealed using pull-down assays would be candidates for deletion in combination with the associated RecJ protein. Likewise, to further explore the interaction of RecJ4 with the DNA repair proteins Hel308 and Orc1, it would be of interest to generate mutants deleted for *recJ4, hel308,* and *orc1* in combination.

Further insight on whether the RecJ proteins have a DNA repair role would be gained by determining sensitivity to DNA damaging agents such as UV, which causes DNA photolesions and single strand breaks, and MMC, a DNA interstrand crosslinking agent. Characterisation of strains via DNA damaging assays are described in Chapter 3: Separation of function in Hel308 and Chapter 5: Nucleases and Helicases.

Additionally, the *recJ* gene deletions featured in this study were introduced into the H164 (*ΔpyrE2, bgaHa-Bb, leuB-Ag1, ΔtrpA)* genetic background in order to determine the role, if any, that the RecJ proteins have in homologous recombination. Due to the lengthy process of generating the gene deletion strains, the 'recombination assays' used to determine levels of recombination (Delmas *et al.*, 2009; Lestini, Duan and Allers, 2010) were not completed in this study and further study should explore this. The 'Recombination assay' is discussed in further detail in Section 2.2.7. Complementation of biochemical findings with genetic characterisation will improve the understanding of the role of the RecJ proteins in *H. volcanii.* Characterisation of strains via recombination assays are described in Chapter 3: Separation of function in Hel308.

Is RecJ2 essential?

This study failed to delete *recJ2* from *H. volcanii* in either a wildtype (H164) or the triple-*recJ*-delete (H4376) genetic backgrounds, suggesting that the *recJ2* gene may be essential. The Cdc45 homologue in *Thermococcus kodakaraensis*, GAN, is nonessential. This suggests that RecJ2 is essential because it acts in a different pathway, and not as part of the CMG complex. Revealing the protein:protein interactions of RecJ2 will give an insight into which pathway it acts. Additionally, it is of interest to determine whether *recJ2* is also essential in an originless strain or if this phenotype is related to the presence of origins.

The *recJ2* gene is found within a highly conserved, co-transcribed operon including the 30S ribosomal genes (Personal correspondence, Kira Makarova), which provides an alternative explanation to the essentiality of RecJ2.

Little is known about the 3D structure and domain composition of RecJ2, therefore it would be interesting to determine the biochemical function and solve the 3D structure by overexpression of RecJ2 and completion of DNA binding assays, nuclease assays, and crystallisation for X-ray crystallography. Complementation of biochemical findings with genetic characterisation will improve the understanding of the role of the RecJ2 in *H. volcanii.*

6.7 Conclusion

The CMG complex is highly conserved throughout eukaryotes and archaea, comprising Cdc45, the replicative helicase MCM, and GINS. MCM has previously been identified as essential for both origin-dependent and recombination-dependent replication in *Haloferax volcanii*. GINS is also essential for both replicative pathways. These data suggest that the CMG complex is likely to be essential for driving both origin-dependent and recombination-dependent replication. It is still unclear which of the RecJ proteins is the Cdc45 homologue but the more likely candidates of the four RecJ proteins are RecJ2 and RecJ4.

Chapter 7: Conclusion and Future Perspectives

Hel308 is a monomeric 3' to 5' DNA helicase conserved throughout archaea and metazoans, but is absent from bacteria and fungi. Hel308 acts analogously to RecQ-family helicases, which serve as guardians of genome stability. Hel308 has been implicated in a variety of overlapping DNA repair pathways, including homologous recombination. Hel308 localises at damaged DNA replication forks in both humans and archaea. However, the exact role of Hel308 remains elusive. *Haloferax volcanii* was used as a model archaeon to study cellular role of Hel308. Deletion mutants of *hel308*, or point mutants that affect the ATP-dependent helicase activity, are slow-growing and are sensitive to DNA cross-linking agents, which are potent blocks to DNA replication. By contrast, point mutants that exhibit hyper-recombinant phenotypes are not deficient in DNA repair or normal growth.

Hel308 is Involved in multiple DNA Repair Pathways

Multiple DNA repair proteins from various DNA repair pathways were identified as interacting partners of Hel308. These interactions implicate Hel308 in homologous recombination (HR), mismatch repair (MMR), nucleotide excision repair (NER), and transcription coupled repair (TCR).

Most interestingly, the overexpression of Hel308 (WT) and Hel308-D145N yields little of the target protein but co-purifies a large quantity of a 30kDa protein identified by mass spectrometry as DlnA (HVO_2382). DlnA is found in the highly conserved RadB neighbourhood and belongs to the YqgF family of endonucleases. In *E. coli*, YqgF has been proposed to function as an alternative to RuvC (Aravind, Makarova and Koonin, 2000), which acts to resolve Holliday junctions (Rocha, Cornet and Michel, 2005). In *H. volcanii*, the role of DlnA is yet to be determined but this interaction with Hel308 may give further insight to how both proteins function. For this reason, it would be of interest to repeat the overexpression of Hel308 in a strain deleted for DlnA.

Hel308 was shown to interact with the RNA polymerase subunit Rpo1 and the NER protein UvrA *in vivo.* Transcription-coupled repair (TCR) is a sub-pathway of nucleotide excision repair (NER) in which RNA polymerase-arresting lesions located in the transcribed strand of active genes are preferentially targeted by the UvrABC complex. TCR is not dependent on *uvrA* in *H. volcanii* (Stantial *et al.*, 2016) suggesting that Hel308 could be acting as the coupling factor in TCR. Hel308 also acts as an anti-recombinase, therefore it is possible that Hel308 promotes DNA repair via NER instead of homologous recombination, either by acting with UvrA directly at DNA lesions or, acting during TCR at the sites of arrested RNA polymerases. Confirmation of these interactions and deletion of

hel308 in combination with TCR and NER proteins could give further insights into the role of Hel308 in these DNA repair pathways.

Hel308 is an anti-recombinase

Previous studies have shown that Hel308 homologues can disassemble recombinase filament from ssDNA and D-loop structures. Furthermore, mutants deleted for the RadA paralogue *radB* in *H. volcanii* exhibit a decrease in the level of homologous recombination to 5% of that in wild type cells. However, a double deletion of *hel308* with *radB* restores, in part, homologous recombination. Therefore, it was proposed that Hel308 functions as an anti-recombinase. This study provides further evidence to support this proposal.

The identification of RadA and RadB as interacting partners of Hel308 supports the proposal that Hel308 is involved in homologous recombination. Bacterial RecQ interacts with RecJ to generate ssDNA tails during homologous recombination, this is known as the end resection step. (Morimatsu and Kowalczykowski, 2014). These ssDNA tails are then coated with the RecA-family recombinase nucleoprotein filament. Hel308 is a RecQ family helicase therefore, the interaction with RecJ3/4 and RadA (the RecA-family recombinase homologue in *H. volcanii*) suggests that Hel308 could act analogously to bacterial RecQ.

Topoisomerases of both type I (TopA, HVO_0681) and type II (Topo VI, HVO_1570/1571 and GyrA HVO_1573) were identified as interacting partners of Hel308a *in vivo.* Topoisomerases are able to interconvert supercoiled and relaxed DNA states by the introduction of positive and negative supercoiling. Subsequently, topoisomerases can also introduce and remove DNA catenanes and knots (Collin, Karkare and Maxwell, 2011). Topoisomerases are critical at resolving hemi-catenanes, which form as a result of the dissolution of Holliday junctions, suggesting that Hel308 may have a role in the latter stages of homologous recombination.

Deleting the *hel308* gene increases recombination but decreases DNA repair and growth. This suggests that Hel308 actively targets and unwinds D-loops, supporting the proposal that Hel308 functions as an anti-recombinase. The point mutants F316A and R743A resulted in ~166,000-fold and ~35,500-fold increase in recombination frequency respectively, without affecting DNA repair or growth. This result is due exclusively to increased non-crossover recombination, since not a single crossover event was observed for these strains. The synthesis dependent strand annealing (SDSA) pathway produces non-crossover products, which reduces the likelihood of a genome rearrangement (San Filippo, Sung and Klein, 2008). Therefore, Hel308 regulates the levels of recombination and may

also influence the pathway choice by which the recombination intermediates are resolved. It would be of interest to determine how Hel308 is involved in this pathway choice.

Separation of functions in recombination and DNA repair

Strains deleted for *hel308* exhibit significant fitness defects and are sensitive to MMC treatment and sometimes UV irradiation, depending on the genotypic background. Mutating residues F316 and R743 increases recombination without affecting DNA repair or growth. By contrast, ATPase-null (*hel308-D145N*) strains exhibit a decrease in recombination levels to half that of wild type and exhibit Δ*hel308*-like phenotype regarding DNA repair and growth. The hyper-recombinant point mutations identify a separation of function of Hel308, where presumably the ability to bind D-loop intermediates of recombination is affected, but the helicase activity involved in DNA repair is not. To further explore this separation of function, it would also be of interest to couple the hyper-recombinant and ATPase-null point mutations in Hel308 with the deletion of the recombination genes *radA* and *radB,* as well as the 5-gene *radB* neighbourhood deletion (Δ*rcrA*Δ*radB*Δ*ndnR*Δ*hrp*Δ*cdc48a*). The 5-gene *radB* neighbourhood is highly conserved throughout archaea (Figure 4.32) and comprises the homologous recombination mediator protein *radB*, ubiquitin-like modification regulatory proteins *rcrA* and *cdc48a*, and a protein with unknown function *ndnR,* which has been re-annotated as *dlnA.* RadB and Cdc48a were both identified as interacting partners of Hel308. Most interestingly, high levels of DlnA co-purified with the over-expression of wild-type Hel308 and the Hel308-D145N mutant (See Chapter 4: Biochemical analysis of Hel308 for further details). Therefore, combining the 5-gene *radB* neighbourhood deletion with the hyper-recombinant and ATPase-null point mutations in Hel308 should give further insight into the separation of function of Hel308.

Hel308 binds DNA prior to ATP binding

Replacing *hel308* with the ATPase-null allele *hel308-D145N* decreases recombination levels to half that of wild type. The combined ATPase-null and hyper-recombinant mutations, *hel308-D145N-F316A* and *hel308-D145N-R743A*, also decrease recombination levels to half that of wild type. Therefore, the ATPase-null (D145N) mutation is dominant to the hyper-recombinant (F316A and R743A) mutations. The decrease in recombination when the D145N mutation is present implies that helicase activity is important for homologous recombination; conversely, the increase in recombination in the Δ*hel308* strain contradicts this. Walker A mutations (K53A/R) were not viable and possibly lethal to the cell, whereas the Walker B mutation (D145N) was viable and showed a Δ*hel308*-like phenotype in terms of DNA repair and growth. This

suggests that Hel308 first binds DNA tightly. On binding ATP, Hel308 changes conformation and loosens its grip on the DNA but remains static. Hydrolysis of the ATP causes a second conformational change and a release of energy, resulting in a paddling motion and translocation along the DNA (Figure 3.39). Unable to bind ATP or translocate, the Hel308-K53A/R effectively creates a "road-block" by binding tightly to DNA and stalling DNA replication and repair. This results in catastrophic genome arrest and hence the cells are not viable. Conversely, the Hel308-D145N protein is able to bind DNA successfully followed by ATP, but remains static on the ssDNA, thereby blocking access for DNA replication and repair proteins. Due to the relatively loose grip that Hel308-D145N has on the DNA, the helicase can be dissociated granting access to DNA replication, recombination, and repair proteins. Therefore, the Walker B mutation acts as a 'speed-bump' slowing down recombination by blocking access to the ssDNA for RadA nucleofilament formation.

Hel308 is a DNA repair regulator

To further explore the role of Hel308 in homologous recombination, *hel308* was successfully deleted in combination with the homologous recombination gene *radA*. Strains deleted for *radA* usually have a severe growth defect (Woods and Dyall-Smith, 1997). As expected, the Δ*radA* strain (H298) had a severe fitness defect. Strains deleted for *hel308* also exhibit significant fitness defects but not to the same severity as the Δ*radA* strains. Most interestingly, the strain deleted for both *hel308* and *radA* (H3515) was healthier than either single deletion, with a generation time more comparable to wild-type. A similar phenotype was observed in survival fraction following UV irradiation and MMC treatment. Strains deleted for *radA* are very sensitive to both UV irradiation and MMC treatment, as was seen in this study. Similarly, strains deleted for *hel308* are sensitive to MMC treatment and sometimes UV irradiation, depending on the genotypic background. Again, the strain deleted for both *hel308* and *radA* (H3515) was healthier than either single deletion, exhibiting a wildtype-like response to both UV irradiation and MMC treatment. These data suggest that in the absence of Hel308-mediated repair and homologous recombination, a third distinct DNA-repair pathway is activated. The fitness defects observed in the single *radA* and *hel308* deleted strains suggest that this third DNA repair pathway is suppressed by both Hel308-mediated repair and homologous recombination. An alternative explanation is that in deleting both the *hel308* and *radA* genes, a suppressor mutation has arisen, not unlike the Δ*recBC* suppressor mutations in *E. coli* (Clark and Margulies, 1965). Only a singular clone with the Δ*hel308*Δ*radA* phenotype was generated, which supports the suppressor-mutation theory. However, a suppressor mutation or mutations would need the ability to bypass two distinct repair pathways in order to restore wildtype-like fitness. This seems unlikely considering that the *E. coli recBC* mutants, which

contain the *sbcB* or *sbcCD* suppressor mutations, are never as healthy or as proficient in homologous recombination as the wild-type. In order to determine whether a suppressor mutation or mutations have arisen during the deletion of *hel308* and *radA*, analysis of the H3515 strain via whole genome sequencing needs to be completed. Complementation assays, reintroducing *hel308* and *radA* at the respective chromosomal loci, would also aid in the understanding of the complex interplay between these pathways. Furthermore, it would be of interest to reattempt the triple deletion of *hel308, radB,* and *radA.*

Hel308 may interact with the CMG complex and replication factors in *H. volcanii*

Archaeal replication resembles its eukaryotic counterpart, for which the CMG complex is essential. The eukaryotic CMG complex comprises Cdc45, MCM, and GINS (Makarova, Koonin and Kelman, 2012). Whilst *H. volcanii* MCM and GINS homologues are known, a Cdc45 homologue has yet to be identified. Bioinformatic and structural studies have revealed that human Cdc45 and archaeal RecJ are evolutionarily related to bacterial RecJ (Makarova, Koonin and Kelman, 2012; Pellegrini, 2017) and therefore, have led to the proposal that archaeal RecJ is the Cdc45 homologue. Both the RecJ3, RecJ4 and MCM were identified in as potential interacting partners of Hel308, suggesting that either RecJ3 or RecJ4 could be the Cdc45 homologue. Due to sequence and structural similarities between RecJ3 and RecJ4, it is probable that RecJ3 has a similar role to RecJ4, or that either protein is redundant. Furthermore, RecJ4 co-purified with the replication initiation protein Orc1 (Darya Ausannikava, University of Nottingham, personal communication), providing further evidence that RecJ4 is the Cdc45 homologue.

Orc5 and PriS were also identified as potential interacting partners of Hel308. Orc5 is one of several Cdc6/Orc1 homologues present in *H. volcanii.* The main chromosome of the wild isolate, DS2, has three origins of replication, *oriC1,2,3*, each with a distinct associated Orc protein. Orc5 is associated with the *oriC2* replication origin (Norais *et al.*, 2007; Hawkins *et al.*, 2013). PriS is the small catalytic subunit of the eukaryotic-type DNA primase. DNA primases synthesise RNA primers on both the leading and lagging strands in order to initiate DNA synthesis (Le Breton *et al.*, 2007).

The potential interaction of Hel308 with RecJ, MCM, Orc5, and PriS implicates Hel308 in DNA replication, which is consistent with the proposal that Hel308 acts in the restart of stalled replication forks. The data presented in this study are preliminary therefore, in order to confirm these interactions, repetition of the protein:protein interaction assays alongside an empty vector control is required. Genuine interactions may also be determined by the reverse protein:protein

interaction assays using tagged versions of the potential interaction partners. Identification of the presence of Hel308 via mass spectrometry would provide further evidence that the interaction *in vivo* is genuine. Where possible, it would be of interest to delete the genes encoding these interaction partners of Hel308 in combination with *hel308*, give further insight and aid understanding of the role of Hel308 in DNA replication.

Summary

This study has emphasised that Hel308 is a multifaceted helicase that plays several roles in many overlapping DNA repair pathways in *H. volcanii.* Hel308 is implicated in interstrand crosslink repair, homologous recombination, mismatch repair, nucleotide excision repair, transcription coupled repair, and DNA replication. Hel308 has also shown to be a regulator of DNA repair pathways, including homologous recombination. In order to perform these many diverse roles, Hel308 has demonstrated a separation of function between helicase-independent recombination and helicase-dependent DNA repair. The information presented in this study provides an insight into the understanding of Hel308 mediated repair and genome stability in *H. volcanii.*

References

Adams, M. D., McVey, M. and Sekelsky, J. J. (2003) 'Drosophila BLM in Double-Strand Break Repair by Synthesis- Dependent Strand Annealing', Science, 4295(1997).

Adelman, C. et al. (2013) 'HELQ promotes RAD51 paralogue-dependent repair to avert germ cell loss and tumorigenesis.', Nature, 502(7471), pp. 381–4.

Allers, T. et al. (2004) 'Development of Additional Selectable Markers for the Halophilic Archaeon Haloferax volcanii Based on the leuB and trpA Genes', Applied and Environmental Microbiology, 70(2), pp. 943–953.

Allers, T. et al. (2010) 'Improved strains and plasmid vectors for conditional overexpression of His-tagged proteins in Haloferax volcanii', Applied and Environmental Microbiology, 76(6), pp. 1759–1769.

Allers, T. and Mevarech, M. (2005) 'Archaeal genetics - the third way.', Nature reviews. Genetics, 6(January), pp. 58–73.

Allers, T. (2010) 'Overexpression and purification of halophilic proteins in Haloferax volcanii', Bioengineered Bugs, 1(4), pp. 288–290.

Altman-Price, N. and Mevarech, M. (2009) 'Genetic evidence for the importance of protein acetylation and protein deacetylation in the halophilic archaeon Haloferax volcanii', Journal of Bacteriology, 191(5), pp. 1610–1617.

Alvarez, D. F. et al. (2010) 'Publishing flow cytometry data', American Journal of Physiology - Lung Cellular and Molecular Physiology, 298(2), pp. 1–7.

Arahal, D. R. et al. (1996) 'Phylogenetic analyses of some extremely halophilic Archaea from Dead Sea water, determined on the basis of their 16S rRNA sequences.', Appl. Environ. Microbiol., 62(10), pp. 3779–3786.

Aravind, L. and Koonin, E. V. (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 Research, 11(8), pp. 1365–1374.

Aravind, L., Makarova, K. S. and Koonin, E. V (2000) 'Holliday junction resolvases and related nucleases : identification of new families , phyletic distribution and evolutionary trajectories', 28(18), pp. 3417–3432.

Arezi, B. and Kuchta, R. D. (2000) 'Eukaryotic DNA primase', Trends in Biochemical Sciences, 25(11), pp. 572–576.

Arias-Palomo, E. et al. (2013) 'The bacterial DnaC helicase loader is a DnaB ring breaker', Cell. Elsevier Inc., 153(2), pp. 438–448.

Atkinson, J. and McGlynn, P. (2009) 'Replication fork reversal and the maintenance of genome stability', Nucleic Acids Research, 37(11), pp. 3475–3492.

Auchtung, T. A., Takacs-Vesbach, C. D. and Cavanaugh, C. M. (2006) '16S rRNA phylogenetic investigation of the candidate division "Korarchaeota"', Applied and Environmental Microbiology, 72(7), pp. 5077–5082.

Ausiannikava, D. and Allers, T. (2017) 'Diversity of DNA replication in the archaea', Genes, 8(2), pp. 1–14.

Azeroglu, B. et al. (2016) 'RecG Directs DNA Synthesis during Double-Strand Break Repair', PLoS Genetics, 12(2), pp. 1–23.

Bab-Dinitz, E. et al. (2006) 'Haloferax volcanii PitA: An example of functional interaction between the Pfam chlorite dismutase and antibiotic biosynthesis monooxygenase families?', Bioinformatics, 22(6), pp. 671–675.

Babron, M. 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(7), pp. 1523–1527.

Bachrati, C. Z. and Hickson, I. D. (2009) 'Dissolution of Double Holliday Junctions by the Concerted Action of BLM and Topoisomerase IIIa', Annual review of biochemistry, 582(2), pp. 91–102.

Baek, G. H. et al. (2013) 'Cdc48: A Swiss Army Knife of Cell Biology', Journal of Amino Acids, 2013, pp. 1–12.

Barry, E. R. and Bell, S. D. (2006) 'DNA replication in the archaea.', Microbiology and molecular biology reviews : MMBR, 70(4), pp. 876–887.

Bell, S. P. and Dutta, A. (2002) 'DNA Replication in Eukaryotic Cells', Annual Review of Biochemistry, 71(1), pp. 333–374.

Benson, F. E., Baumann, P. and West, S. C. (1998) 'Synergistic actions of Rad51 and Rad52 in recombination and DNA repair', Nature, 391(6665), pp. 401–404.

Bin, R. et al. (2009) 'Structure and function of a novel endonuclease acting on branched DNA substrates', Biochemical Society Transactions, 39(1), pp. 145–149.

Bitan-Banin, G. et al. (2003) 'Development of a Gene Knockout System for the Halophilic Archaeon', Society, 185(3), pp. 772–778.

Bizard, A. H. and Hickson, I. D. (2014) 'The dissolution of double Holliday junctions', Cold Spring Harbor Perspectives in Biology, 6(7), pp. 1-14.

Blastyák, A. et al. (2007) 'Yeast Rad5 Protein Required for Postreplication Repair Has a DNA Helicase Activity Specific for Replication Fork Regression', Molecular Cell, 28(1), pp. 167–175.

Bocquet, N. et al. (2014) 'Structural and mechanistic insight into Holliday junction dissolution by Topoisomerase IIIα and RMI1', 21(3), pp. 261–268.

De Boer, J. and Hoeijmakers, J. H. J. (2000) 'Nucleotide excision repair and human syndromes', Carcinogenesis, 21(3), pp. 453–460.

Bolt, E. L., Lloyd, R. G. and Sharples, G. J. (2001) ‘Genetic analysis of an archaeal Holliday junction resolvase in Escherichia coli’, Journal of Molecular Biology, 310(3), pp. 577–589.

Bowater, R. and Doherty, A. J. (2006) ‘Making ends meet: Repairing breaks in bacterial DNA by non-homologous end-joining’, PLoS Genetics, 2(2), pp. 93–99.

Boyd, J. B. et al. (1981) ‘Third-chromosome mutagen-sensitive mutants of Drosophila melanogaster.’, Genetics, 97(3–4), pp. 607–623.

Boyd, J. B., Sakaguchi, K. and Harris, P. V. (1990) ‘mus308 Mutants of Drosophila exhibit hypersensitivity to DNA cross-linking agents and are defective in a deoxyribonuclease’, Genetics, 125(4), pp. 813–819.

Braun, F. et al. (2019) ‘Cyclic nucleotides in archaea: Cyclic di-AMP in the archaeon Haloferax volcanii and its putative role’, Microbiology Open, 8(9), pp. 1–23.

Brendel, J. 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’, Journal of Biological Chemistry, 289(10), pp. 7164–7177.

Le Breton, M. et al. (2007) ‘The Heterodimeric Primase from the Euryarchaeon Pyrococcus abyssi: A Multifunctional Enzyme for Initiation and Repair?’, Journal of Molecular Biology. Elsevier Ltd, 374(5), pp. 1172–1185.

Breuert, S. et al. (2006) ‘Regulated polyploidy in halophilic archaea’, PLoS ONE, 1(1), pp. 1-10.

Brochier-Armanet, C. et al. (2008) ‘Mesophilic Crenarchaeota: proposal for a third archaeal phylum, the Thaumarchaeota.’, Nature reviews. Microbiology, 6(3), pp. 245–252.

Brosh, R. M. (2013) ‘DNA helicases involved in DNA repair and their roles in cancer.’, Nature reviews. Cancer, 13(8), pp. 542–58.

Bryson, V. and Szybalski, W. (1952) ‘Microbial selection’, Science, 116(3003), pp. 45–51.

Bugreev, D. V. et al. (2007) ‘Novel pro- and anti-recombination activities of the Bloom’s syndrome helicase’, Genes and Development, 21, p. 3085.

Büttner, K., Nehring, S. and Hopfner, K.-P. (2007) ‘Structural basis for DNA duplex separation by a superfamily-2 helicase.’, Nature structural & molecular biology, 14(7), pp. 647–652.

Cadet, J., Sage, E. and Douki, T. (2005) ‘Ultraviolet radiation-mediated damage to cellular DNA’, Mutation Research - Fundamental and Molecular Mechanisms of Mutagenesis, 571(1-2 SPEC. ISS.), pp. 3–17.

Cameron, B. et al. (1991) ‘Genetic and sequence analyses of a Pseudomonas denitrificans DNA fragment containing two cob genes’, Journal of Bacteriology, 173(19), pp. 6058–6065.

Cann, I. K. O. et al. (1998) ‘A heterodimeric DNA polymerase: Evidence that members of Euryarchaeota possess a distinct DNA polymerase’, Proceedings of the National Academy of Sciences of the United States of America, 95(24), pp. 14250–14255.

Cann, I. K. O. et al. (1999) ‘Functional interactions of a homolog of proliferating cell nuclear antigen with DNA polymerases in Archaea’, Journal of Bacteriology, 181(21), pp. 6591–6599.

Cao, S., Hepowit, N. and Maupin-Furlow, J. A. (2015) ‘Ubiquitin-like protein SAMP1 and JAMM/MPN+ metalloprotease HvJAMM1 constitute a system for reversible regulation of metabolic enzyme activity in archaea’, PLoS ONE, 10(5), pp. 1–11.

Carolina B. Gabbai and Marians, K. J. (2010) ‘Recruitment to Stalled Replication Forks of the PriA DNA Helicase and Replisome-loading Activities is Essential for Survival’, DNA Repair (Amst), 23(1), pp.1-7.

Caruthers, J. M. and McKay, D. B. (2002) ‘Helicase structure and mechanism’, Current Opinion in Structural Biology, 12(1), pp. 123–133.

Castelle, C. J. et al. (2015) ‘Genomic expansion of domain archaea highlights roles for organisms from new phyla in anaerobic carbon cycling’, Current Biology. Elsevier Ltd, 25(6), pp. 690–701.

Ceccaldi, R. et al. (2015) ‘Homologous recombination-deficient tumors are hyper- dependent on POLQ-mediated repair’, Nature, 518(7538), pp. 258–262.

Chaban, B., Ng, S. Y. M. and Jarrell, K. F. (2006) ‘Archaeal habitats--from the extreme to the ordinary.’, Canadian journal of microbiology, 52(September 2016), pp. 73–116.

Chamieh, H., Ibrahim, H. and Kozah, J. (2015) ‘Genome-wide identification of SF1 and SF2 helicases from archaea’, Gene. Elsevier B.V., 576(1), pp. 214–228.

Chan, S. H., Yu, A. M. and McVey, M. (2010) ‘Dual roles for DNA polymerase theta in alternative end-joining repair of double-strand breaks in Drosophila’, PLoS Genetics, 6(7), pp. 1–16.

Changela, A., DiGate, R. J. and Mondragón, A. (2001) ‘Crystal structure of a complex of a type IA DNA topoisomerase with a single-stranded DNA molecule’, Nature, 411(6841), pp. 1077–1081.

Charlebois, R. L. et al. (1991) ‘Detailed physical map and set of overlapping clones covering the genome of the archaebacterium Haloferax volcanii DS2’, Journal of Molecular Biology, 222(3), pp. 509–524.

Chen, Z., Yang, H. and Pavletich, N. P. (2008) ‘Mechanism of homologous recombination from the RecA-ssDNA/dsDNA structures’, Nature, 453(7194), pp. 489–494.

Chow, K.-H. and Courcelle, J. (2007) ‘RecBCD and RecJ/RecQ Initiate DNA Degradation on Distinct Substrates in UV-Irradiated Escherichia coli’, Radiation Research, 168(4), pp. 499–506.

Christian, J. H. B. and Waltho, J. A. (1962) ‘Solute concentrations within cells of halophilic and non-halophilic bacteria’, Biochimica et Biophysica Acta, 65(3), pp. 506–508.

Ciccarelli, D. et al. (2006) ‘of a Highly Resolved Tree of Life’, Science, 311(March), pp. 1283–1288.

Clare L. Fasching et al. (2015) ‘Top3-Rmi1 dissolve Rad51-mediated D-loops by a topoisomerase-based mechanism’, Bone, 23(1), pp. 1–7.

Clark, A. J. and Margulies, A. D. (1965) ‘Isolation and Characterization of Recombination-deficient Mutants of Escherichia coli K12*’, PNAS, 53(1), pp. 451–459.

Cline, S., Schalkwyk, L. and Doolittle, W. (1989) ‘Transformation of the Archaebacterium Halobacterium-Volcanii With Genomic Dna’, Journal of Bacteriology, 171(9), pp. 4987–4991.

Coin, F. et al. (1999) ‘Mutations in XPB and XPD helicases found in xeroderma pigmentosum patients impair the transcription function of TFIIH.’, The EMBO journal, 18(5), pp. 1357–66.

Collin, F., Karkare, S. and Maxwell, A. (2011) ‘Exploiting bacterial DNA gyrase as a drug target: Current state and perspectives’, Applied Microbiology and Biotechnology, 92(3), pp. 479–497.

Connaris, H. et al. (1999) ‘Expression, reactivation, and purification of enzymes from Haloferax volcanii in Escherichia coli’, Biotechnology and Bioengineering, 64(1), pp. 38–45.

Constantinesco, F. et al. (2004) ‘A bipolar DNA helicase gene, herA, clusters with rad50, mre11 and nurA genes in thermophilic archaea’, Nucleic Acids Research, 32(4), pp. 1439–1447.

Constantinesco, F., Forterre, P. and Elie, C. (2002) ‘NurA, a novel 5′–3′ nuclease gene linked to rad50 and mre11 homologs of thermophilic Archaea’, 3(6), pp. 537–542.

Constantinou, A. et al. (2000) ‘Werner’s syndrome protein (WRN) migrates Holliday junctions and co-localizes with RPA upon replication arrest’, EMBO Reports, 1(1), pp. 80–84.

Corbett, K. D., Shultzaberger, R. K. and Berger, J. M. (2004) ‘The C-terminal domain of DNA gyrase A adopts a DNA-bending -pinwheel fold’, Proceedings of the National Academy of Sciences, 101(19), pp. 7293–7298.

Costa, R. M. A. et al. (2003) ‘The eukaryotic nucleotide excision repair pathway’, Biochimie, 85(11), pp. 1083–1099.

Crameri, a et al. (1996) ‘Improved green fluorescent protein by molecular evolution using DNA shuffling.’, Nature biotechnology, 14(3), pp. 315–319.

Crowley, D. J. 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(1), p. 11.

Cubeddu, L. and White, M. F. (2005) ‘DNA damage detection by an archaeal single-stranded DNA-binding protein’, Journal of Molecular Biology, 353(3), pp. 507–516.

Davis, A. J. and Chen, D. J. (2013) ‘DNA double strand break repair via non-homologous end-joining’, Translational Cancer Research, 2(3), pp. 130–143.

Debussche, L. et al. (1992) ‘Assay, purification, and characterization of cobaltochelatase, a unique complex enzyme catalyzing cobalt insertion in hydrogenobyrinic acid a,c- diamide during coenzyme B12 biosynthesis in Pseudomonas denitrificans’, Journal of Bacteriology, 174(22), pp. 7445–7451.

Della, M. et al. (2004) ‘Mycobacterial Ku and ligase proteins constitute a two-component NHEJ repair machine’, Science, 306(5696), pp. 683–685.

Delmas, S. et al. (2009) ‘Mre11-Rad50 promotes rapid repair of DNA damage in the polyploid archaeon Haloferax volcanii by restraining homologous recombination’, PLoS Genetics, 5(7), pp. 1-12.

Delmas, S., Duggin, I. G. and Allers, T. (2013) ‘DNA damage induces nucleoid compaction via the Mre11-Rad50 complex in the archaeon Haloferax volcanii’, Molecular Microbiology, 87(1), pp. 168–179.

DeLong, E. F. and Pace, N. R. (2001) ‘Environmental diversity of bacteria and archaea.’, Systematic biology, 50(4), pp. 470–478.

Dixon, D. A. and Kowalczykowski, S. C. (1993) ‘The recombination hotspot χ is a regulatory sequence that acts by attenuating the nuclease activity of the E. coli RecBCD enzyme’, Cell, 73(1), pp. 87–96.

Dogliotti, E. et al. (2001) ‘The mechanism of switching among multiple BER pathways’, Progress in Nucleic Acid Research and Molecular Biology, 68, pp. 3–27.

Doherty, A. J., Jackson, S. P. and Weller, G. R. (2001) ‘Identification of bacterial homologues of the Ku DNA’, 500, pp. 186–188.

Duggin, I. G. et al. (2008) ‘The replication fork trap and termination of chromosome replication’, Molecular Microbiology, 70(6), pp. 1323–1333.

Duggin, I. et al. (2015) ‘CetZ tubulin-like proteins control archaeal cell shape’, Nature, 519(7543), pp. 362–365.

Eichler, J. and Maupin-Furlow, J. (2013) ‘Post-translation modification in Archaea: Lessons from Haloferax volcanii and other haloarchaea’, FEMS Microbiology Reviews, 37(4), pp. 583–606.

Elkins, J. G. et al. (2008) ‘A korarchaeal genome reveals insights into the evolution of the Archaea.’, Proceedings of the National Academy of Sciences of the United States of America, 105(23), pp. 8102–8107.

Embley, T. M. et al. (2008) ‘The archaebacterial origin of eukaryotes’, Proceedings of the National Academy of Sciences, 105(51), pp. 20356–20361.

Embley, T. M. and Williams, T. A. (2015) 'Evolution: Steps on the road to eukaryotes', Nature, 521(7551), pp. 169–170.

Eme, L. et al. (2017) 'Archaea and the origin of eukaryotes', Nature Reviews Microbiology. Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved., 15, p. 711.

Enemark, E. J. and Joshua-Tor, L. (2008) 'On helicases and other motor proteins', Current Opinion in Structural Biology, 18(2), pp. 243–257.

Erzberger, J. P., Mott, M. L. and Berger, J. M. (2006) 'Structural basis for ATP-dependent DnaA assembly and replication-origin remodeling', Nature Structural and Molecular Biology, 13(8), pp. 676–683.

Evans, M. D., Dizdaroglu, M. and Cooke, M. S. (2004) 'Oxidative DNA damage and disease: Induction, repair and significance', Mutation Research - Reviews in Mutation Research, 567(1), pp. 1-61.

Eydmann, T. et al. (2008) 'Rtf1-mediated eukaryotic site-specific replication termination', Genetics, 180(1), pp. 27–39.

Fairman-Williams, M. E., Guenther, U. P. and Jankowsky, E. (2010) 'SF1 and SF2 helicases: Family matters', Current Opinion in Structural Biology, 20(3), pp. 313–324.

Fang, L., Davey, M. J. and O'Donnell, M. (1999) 'Replisome assembly at oriC, the replication origin of E. coli, reveals an explanation for initiation sites outside an origin', Molecular Cell, 4(4), pp. 541–553.

Feng, L. et al. (2018) 'The trimeric Hef-associated nuclease HAN is a 3→5 exonuclease and is probably involved in DNA repair', Nucleic Acids Research. Oxford University Press, 46(17), pp. 9027–9043.

Forouzan, D. et al. (2012) 'The archaeal proteasome is regulated by a network of AAA ATPases', Journal of Biological Chemistry, 287(46), pp. 39255–39262.

Foster, P. G., Cox, C. J. and Martin Embley, T. (2009) 'The primary divisions of life: A phylogenomic approach employing composition-heterogeneous methods', Philosophical Transactions of the Royal Society B: Biological Sciences, 364(1527), pp. 2197–2207.

Frick, D. N. and Richardson, C. C. (2001) 'Key Words replication, RNA, oligoribonucleotides, polymerase, helicase', Annu. Rev. Biochem, 70(1) pp. 39–80.

Fujii, R. et al. (2002) 'Identification of a neuropeptide modified with bromine as an endogenous ligand for GPR7', Journal of Biological Chemistry, 277(37), pp. 34010–34016.

Fujikane, R. et al. (2005) 'Identification of a novel helicase activity unwinding branched DNAs from the hyperthermophilic archaeon, Pyrococcus furiosus', Journal of Biological Chemistry, 280(13), pp. 12351–12358.

Fujikane, R., Shinagawa, H. and Ishino, Y. (2006) 'The archaeal Hjm helicase has recQ-like functions, and may be involved in repair of stalled replication fork', Genes to Cells, 11(2), pp. 99–110.

Fukui, K. (2010) ‘DNA mismatch repair in eukaryotes and bacteria’, Journal of Nucleic Acids, 2010(c).

Gamble-Milner, R. J. (2016) ‘Genetic Analysis of the Hel308 Helicase in the Archaeon Haloferax volcanii’, (May).

Gari, K. et al. (2008) ‘The Fanconi Anemia Protein FANCM Can Promote Branch Migration of Holliday Junctions and Replication Forks’, Molecular Cell, 29(1), pp. 141–148.

Ge, X. Q., Jackson, D. A. and Blow, J. J. (2007) ‘Dormant origins licensed by excess Mcm2-7 are required for human cells to survive replicative stress’, Genes and Development, 21(24), pp. 3331–3341.

Genschel, J. and Modrich, P. (2009) ‘Functions of MutLα, replication protein A (RPA), and HMGB1 in 5′-directed mismatch repair’, Journal of Biological Chemistry, 284(32), pp. 21536–21544.

German, J. (1993) ‘Bloom Syndrome A Mendelian Prototype of Somatic Mutational Disease’, Medicine, 72(6), pp. 393–406.

Giroux, X. and Macneill, S. A. (2015) ‘Molecular Genetic Methods to Study DNA Replication Protein Function in Haloferax volcanii, A Model Archaeal Organism’, DNA Replication: Methods and Protocols, 1300, pp. 187-218

Gong, F., et al. (2001) ‘The mechanism of tryptophan induction of tryptophanase operon expression: Tryptophan inhibits release factor-mediated cleavage of TnaC-peptidyl-tRNAPro’, PNAS, 98(16), pp. 8997–9001.

Gorbalenya, A. E. and Koonin, E. V (1993) ‘Helicases: amino acid sequence comparisons and structure-function relationships’, Current Opinion in Structural Biology, 3(3), pp. 419–429.

Graziano, G. and Merlino, A. (2014) ‘Molecular bases of protein halotolerance’, Biochimica et Biophysica Acta - Proteins and Proteomics. Elsevier B.V., 1844(4), pp. 850–858.

Grogan, D. W. (2000) ‘The question of DNA repair in hyperthermophilic archaea’, Trends in Microbiology, 8(4), pp. 180–185.

Grogan, D. W. (2004) ‘Stability and repair of DNA in hyperthermophilic archaea’, Current Issues in Molecular Biology, 6(2), pp. 137–144.

Van Der Gulik, P. T. S., Hoff, W. D. and Speijer, D. (2017) ‘In defence of the three-domains of life paradigm’, BMC Evolutionary Biology. BMC Evolutionary Biology, 17(1), pp. 1–5.

Gupta, S., Yeeles, J. T. P. and Marians, K. J. (2014) ‘Regression of replication forks stalled by leading-strand template damage: I. Both RecG and RuvAB catalyze regression, but RuvC cleaves the holliday junctions formed by RecG preferentially’, Journal of Biological Chemistry, 289(41), pp. 28376–28387.

Guy, C. P. et al. (2006) ‘Interactions of RadB, a DNA repair protein in archaea, with DNA and ATP’, Journal of Molecular Biology, 358(1), pp. 46–56.

Guy, C. P. and Bolt, E. L. (2005) 'Archaeal Hel308 helicase targets replication forks in vivo and in vitro and unwinds lagging strands', Nucleic Acids Research, 33(11), pp. 3678–3690.

Hagmann, M. (1997) 'RAGged repair: What's New in V(D)J Recombination', Biol. Chem., 378(August), pp. 815–819.

Haldenby, S., White, M. F. and Allers, T. (2009) 'RecA family proteins in archaea: RadA and its cousins', Biochemical Society Transactions, 37(1), pp. 102–107.

Hall, M. C. and Matson, S. W. (1999) 'Helicase motifs: the engine that powers DNA unwinding.', Molecular microbiology, 34(5), pp. 867–877.

Hanawalt, P. C. (1994) 'Evolution of concepts in DNA repair', Environmental and Molecular Mutagenesis, 23(2 S), pp. 78–85.

Haque, R. U., Paradisi, F. and Allers, T. (2019) 'Haloferax volcanii as immobilised whole cell biocatalyst : new applications for halophilic systems'. Applied Microbiology and Biotechnology.

Harris, J. K. et al. (2003) 'The genetic core of the universal ancestor', Genome Research, 13(3), pp. 407–412.

Hartman, A. L. et al. (2010) 'The complete genome sequence of Haloferax volcanii DS2, a model archaeon', PLoS ONE, 5(3).

Hawkins, M. et al. (2013) 'Accelerated growth in the absence of DNA replication origins', Nature. Nature Publishing Group, 503(7477), pp. 544–547.

Hefferin, M. L. and Tomkinson, A. E. (2005) 'Mechanism of DNA double-strand break repair by non-homologous end joining', DNA Repair, 4(6), pp. 639–648.

Helleday, T., Eshtad, S. and Nik-Zainal, S. (2014) 'Mechanisms underlying mutational signatures in human cancers', Nature Reviews Genetics. Nature Publishing Group, 15(9), pp. 585–598.

Heller, R. C. and Marians, K. J. (2005a) 'The disposition of nascent strands at stalled replication forks dictates the pathway of replisome loading during restart', Molecular Cell, 17(5), pp. 733–743.

Heller, R. C. and Marians, K. J. (2005b) 'Unwinding of the nascent lagging strand by Rep and PriA enables the direct restart of stalled replication forks', Journal of Biological Chemistry, 280(40), pp. 34143–34151.

Hishida, T. et al. (2002) 'Role of Walker Motif A of RuvB Protein in Promoting Branch Migration of Holliday Junctions', Journal of Biological Chemistry, 274(36), pp. 25335–25342.

Holmes, M. L. and Dyall-Smith, M. L. (2000) 'Sequence and expression of a halobacterial beta-galactosidase gene.', Molecular microbiology, 36, pp. 114–122.

Holmes, M. L., Nuttall, S. D. and Dyall-smith, M. L. (1991) 'Construction and use of halobacterial shuttle vectors and further studies on Haloferax DNA gyrase . Construction and Use of Halobacterial Shuttle Vectors

and Further Studies on Haloferax DNA Gyrase', Plasmid, 173(12), pp. 3807–3813.

Hong, Y. et al. (2012) 'Dissection of the functional domains of an archaeal Holliday junction helicase', DNA Repair. Elsevier B.V., 11(2), pp. 102–111.

Hopfner, K. P. et al. (2000) 'Mre11 and Rad50 from Pyrococcus furiosus: Cloning and biochemical characterization reveal an evolutionarily conserved multiprotein machine', Journal of Bacteriology, 182(21), pp. 6036–6041.

Hopfner, K. P. et al. (2002) 'The Rad50 zinc-hook is a structure joining Mre11 complexes in DNA recombination and repair', Nature, 418(6897), pp. 562–566.

Horiuchi, T. et al. (2004) 'Cooperation of the N-terminal Helicase and C-terminal Endonuclease Activities of Archaeal Hef Protein in Processing Stalled Replication Forks', Journal of Biological Chemistry, 279(51), pp. 53175–53185.

Hsieh, P. and Yamane, K. (2009) 'DNA mismatch repair: Molecular mechanism, cancer, and ageing', 129, pp. 391–407.

Hu, Y. et al. (2007) 'RECQL5/Recql5 helicase regulates homologous recombination and suppresses tumor formation via disruption of Rad51 presynaptic filaments', Genes and Development, 21(23), pp. 3073–3084.

Humbard, M. A., Zhou, G. and Maupin-Furlow, J. A. (2009) 'The N-terminal penultimate residue of 20S proteasome α1 influences its Nα acetylation and protein levels as well as growth rate and stress responses of Haloferax volcanii', Journal of Bacteriology, 191(12), pp. 3794–3803.

Ibarra, A., Schwob, E. and Méndez, J. (2008) 'Excess MCM proteins protect human cells from replicative stress by licensing backup origins of replication', Proceedings of the National Academy of Sciences of the United States of America, 105(26), pp. 8956–8961.

Ira, G. et al. (2003) 'Srs2 and Sgs1–Top3 Suppress Crossovers during Double-Strand Break Repair in Yeast', Physiology & behavior, 176(4), pp. 139–148.

Ira, G. et al. (2004) 'DNA end resection, homologous recombination and DNA damage checkpoint activation require CDK1', 344(6188), pp. 1173–1178.

Ishibashi, M. et al. (2003) 'Activation of Halophilic Nucleoside Diphosphate Kinase by a Non-ionic Osmolyte, Trimethylamine N-Oxide', Journal of protein chemistry, 22, pp. 345–351.

Ishino, S. et al. (2016) 'Identification of a mismatch-specific endonuclease in hyperthermophilic Archaea', Nucleic Acids Research, 44(7), pp. 2977–2986.

Jaakkola, S. T. et al. (2016) 'Buried Alive: Microbes from Ancient Halite', Trends in Microbiology. Elsevier Ltd, 24(2), pp. 148–160.

Jaenicke, R. (2000) ‘Stability and stabilization of globular proteins in solution’, Journal of Biotechnology, 79(3), pp. 193–203.

De Jager, M. et al. (2001) ‘Human Rad50/Mre11 is a flexible complex that can tether DNA ends’, Molecular Cell, 8(5), pp. 1129–1135.

Jasin, M. and Rothstein, R. (2013) ‘Repair of strand breaks by homologous recombination’, Cold Spring Harbor Perspectives in Biology, 5(11), pp. 1–18.

Jensen, R. B., Carreira, A. and Kowalczykowski, S. C. (2010) ‘Purified human BRCA2 stimulates RAD51-mediated recombination’, Physiology & behavior, 176(7316), pp. 139–148.

Jiricny, J. (2006) ‘The multifaceted mismatch-repair system’, Nature Reviews Molecular Cell Biology, 7(5), pp. 335–346.

Jolley, K. A. et al. (1996) ‘Dihydrolipoamide dehydrogenase from the halophilic archaeon Haloferax volcanii: Homologous overexpression of the cloned gene’, Journal of Bacteriology, 178(11), pp. 3044–3048.

Jun, S. H. et al. (2011) ‘Archaeal RNA polymerase and transcription regulation’, Crit Rev Biochem Mol Biol. 46(30), pp. 3921–3932.

Kadyrov, F. A. et al. (2006) ‘Endonucleolytic Function of MutLα in Human Mismatch Repair’, Cell, 126(2), pp. 297–308.

Kaplan, D. L. and O’Donnell, M. (2003) ‘Rho factor: Transcription termination in four steps’, Current Biology, 13(18), pp. 714–716.

Karanja, K. K. et al. (2012) ‘DNA2 and EXO1 in replication-coupled, homology-directed repair and in the interplay between HDR and the FA/BRCA network’, Cell Cycle, 11(21), pp. 3983–3996.

Karow, J. K. et al. (2000) ‘The Bloom’s syndrome gene product promotes branch migration of Holliday junctions’, Proceedings of the National Academy of Sciences of the United States of America, 97(12), pp. 6504–6508.

Kawate, H. et al. (1998) ‘Separation of killing and tumorigenic effects of an alkylating agent in mice defective in two of the DNA repair genes’, Proceedings of the National Academy of Sciences of the United States of America, 95(9), pp. 5116–5120.

Kee, Y. and D’Andrea, A. D. (2010) ‘Expanded roles of the Fanconi anemia pathway in preserving genomic stability’, Genes and Development, 24(16), pp. 1680–1694.

Kelley, L. A. et al. (2015) ‘Europe PMC Funders Group Phyre2 web portal for protein modelling, prediction and analysis’, Nature protocols, 10(6), pp. 845–858.

Kelman, L. M. and Kelman, Z. (2004) ‘Multiple origins of replication in archaea’, Trends in Microbiology, 12(9), pp. 399–401.

Kelman, Z. and O’donnell, M. (1995) ‘Structural and functional similaritites of prokaryotic and eukaryotic DNA polymerase sliding clamps’, Nucleic Acids Research, 23(23), p. 4938.

Kelman, Z. and White, M. F. (2005) ‘Archaeal DNA replication and repair’, Current Opinion in Microbiology, 8(6), pp. 669–676.
Kil, Y. V, Baitin, D. M. and Lanzov, V. A. (2000) ‘Efficient Strand Transfer by the RadA Recombinase from the Hyperthermophilic’, Journal of Bacteriology, 182(1), p. 130.
Klapstein, K., Chou, T. and Bruinsma, R. (2004) ‘Physics of RecA-mediated homologous recognition’, Biophysical Journal. Elsevier, 87(3), pp. 1466–1477.
Kobayashi, J. et al. (2002) ‘NBS1 localizes to γ-H2AX foci through interaction with the FHA/BRCT domain’, Current Biology, 12(21), pp. 1846–1851.
Komori, K. et al. (2000) ‘Both RadA and RadB are involved in homologous recombination in Pyrococcus furiosus’, Journal of Biological Chemistry, 275(43), pp. 33782–33790.
Komori, K. et al. (2002) ‘Novel endonuclease in Archaea cleaving DNA with various branched structure’, Genes & Genetic Systems, 77(4), pp. 227–227.
Komori, K. et al. (2004) ‘Cooperation of the N-terminal helicase and C-terminal endonuclease activities of archaeal Hef protein in processing stalled replication forks’, Journal of Biological Chemistry, 279(51), pp. 53175–53185.
Koonin, E. V. (2015) ‘Archaeal ancestors of eukaryotes: Not so elusive any more’, BMC Biology. BMC Biology, 13(1), pp. 1–7.
Koonin, E. V. and Wolf, Y. I. (2008) ‘Genomics of bacteria and archaea: The emerging dynamic view of the prokaryotic world’, Nucleic Acids Research, 36(21), pp. 6688–6719.
Koonin, E. V. and Yutin, N. (2014) ‘The dispersed archaeal eukaryome and the complex archaeal ancestor of eukaryotes’, Cold Spring Harbor Perspectives in Biology, 6(4), pp. 1–16.
Korolev, S. et al. (1997) ‘Major domain swiveling revealed by the crystal structures of complexes of E. coli Rep helicase bound to single-stranded DNA and ADP’, Cell, 90(4), pp. 635–647.
Kowalczykowski, S. C. (2000) ‘Initiation of genetic recombination and recombination-dependent replication’, Trends in Biochemical Sciences, 25(4), pp. 156–165.
Krejci, L. et al. (2002) ‘Interaction with Rad51 is indispensable for recombination mediator function of Rad52’, Journal of Biological Chemistry, 277(42), pp. 40132–40141.
Krejci, L. et al. (2003) ‘DNA helicase Srs2 disrupts the Rad51 presynaptic filament’, Nature, 423(6937), pp. 255–260.
Krejci, L. et al. (2012) ‘Homologous recombination and its regulation’, Nucleic Acids Research, 40(13), pp. 5795–5818.

Kuchta, R. D. and Stengel, G. (2011) 'Mechanism and Evolution of DNA Primases Primer removal and its consequences', Biochimica et Biophysica Act, 1804(5), pp. 1180–1189.

Kuriyan, J. and O'Donnell, M. (1993) 'Sliding Clamps of DNA Polymerases'. J. Mol. Biol, 234, pp. 915-925.

Kurokawa, Y. et al. (2008) 'Reconstitution of DNA strand exchange mediated by Rhp51 recombinase and two mediators', PLoS Biology, 6(4), pp. 836–848.

Kvaratskhelia, M., Wardleworth, B. N. and White, M. F. (2001) 'Multiple Holliday junction resolving enzyme activities in the Crenarchaeota and Euryarchaeota', FEBS Letters, 491(3), pp. 243–246.

de Laat, W. L., Jaspers, N. G. J. and Hoeijmakers, J. H. J. (1999) 'Molecular mechanism of global genome nucleotide excision repair', Genes & Development, 6(20), pp. 23–34.

Lam, W. and Doolittle, W. (1989) 'Shuttle Vectors for the Archaebacterium Halobacterium-Volcanii', Proceedings of the National Academy of Sciences of the United States of America, 86(14), pp. 5478–5482.

Langer, D. et al. (1995) 'Transcription in archaea: similarity to that in eucarya.', Proceedings of the National Academy of Sciences, 92(13), pp. 5768–5772.

Lankinen, M. H., Vilpo, L. M. and Vilpo, J. A. (1996) 'UV- and γ-irradiation-induced DNA single-strand breaks and their repair in human blood granulocytes and lymphocytes', Mutation Research - Fundamental and Molecular Mechanisms of Mutagenesis, 352(1–2), pp. 31–38.

Lanyi, J. K. (1974) 'Salt-dependent properties of proteins from extremely halophilic bacteria.', Bacteriological reviews, 38(3), pp. 272–90.

Large, A. 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', Molecular Microbiology, 66(5), pp. 1092–1106.

Lestini, R. et al. (2013) 'Intracellular dynamics of archaeal FANCM homologue Hef in response to halted DNA replication', Nucleic Acids Research, 41(22), pp. 10358–10370.

Lestini, R., Duan, Z. and 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. Elsevier B.V., 9(9), pp. 994–1002.

Li, G. M. (2008) 'Mechanisms and functions of DNA mismatch repair', Cell Research, 18(1), pp. 85–98.

Li, J. et al. (2006) 'Similarities and differences between cyclobutane pyrimidine dimer photolyase and (6-4) photolyase as revealed by resonance raman spectroscopy: Electron transfer from the FAD cofactor to ultraviolet-damaged DNA', Journal of Biological Chemistry, 281(35), pp. 25551–25559.

Li, W. et al. (2013) ‘Genetic variants in epidermal growth factor receptor pathway genes and risk of esophageal squamous cell carcinoma and gastric cancer in a Chinese population.’, PloS one, 8(7), p. 68999.

Li, Y. F., Sancar, A. and Heelis, P. F. (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(25), pp. 6322–6329.

Li, Z. et al. (2008) ‘Hjm/Hel308a DNA Helicase from Sulfolobus tokodaii promotes replication fork regression and interacts with Hjc endonuclease in vitro’, Journal of Bacteriology, 190(8), pp. 3006–3017.

Liang, C. et al. (2013) ‘Gene-Environment Interactions of Novel Variants Associated with Head and Neck Cancer’, 34(8), pp. 1111–1118.

Lin, Z., Nei, M. and Ma, H. (2007) ‘The origins and early evolution of DNA mismatch repair genes - Multiple horizontal gene transfers and co-evolution’, Nucleic Acids Research, 35(22), pp. 7591–7603.

Lindahl, T. (1993) ‘Instability and decay of the primary structure of DNA’, Nature, 362, pp. 709–715.

Lindås, A. C. and Bernander, R. (2013) ‘The cell cycle of archaea’, Nature Reviews Microbiology. Nature Publishing Group, 11(9), pp. 627–638.

Lindgren, M. T. et al. (2008) ‘A unique cell division machinery in the Archaea’, Proceedings of the National Academy of Sciences, 105(48), pp. 18942–18946.

Liu, J. et al. (2011) ‘Rad51 paralogues Rad55-Rad57 balance the antirecombinase Srs2 in Rad51 filament formation’, Nature. Nature Publishing Group, 479(7372), pp. 245–248.

Lohman, T. M. and Bjornson, K. P. (1996) ‘Mechanisms of helicase-catalyzed DNA unwinding.’, Annual review of biochemistry, 65(1), pp. 169–214.

Luebben, S. W. et al. (2013) ‘Helq acts in parallel to Fancc to suppress replication-associated genome instability’, Nucleic Acids Research, 41(22), pp. 10283–10297.

Lundgren, M. et al. (2004) ‘Three replication origins in Sulfolobus species: Synchronous initiation of chromosome replication and asynchronous termination’, Proceedings of the National Academy of Sciences of the United States of America, 101(18), pp. 7046–7051.

Lutsenko, E. and Bhagwat, A. S. (1999) ‘Principal causes of hot spots for cytosine to thymine mutations at sites of cytosine methylation in growing cells’, Mutation Research/Reviews in Mutation Research, 437(1), pp. 11–20.

Machwe, A. et al. (2007) ‘Replication fork regression in vitro by the Werner syndrome protein (WRN): Holliday junction formation, the effect of leading arm structure and a potential role for WRN exonuclease activity’, Nucleic Acids Research, 35(17), pp. 5729–5747.

Mackintosh, S. G. and Raney, K. D. (2006) 'DNA unwinding and protein displacement by superfamily 1 and superfamily 2 helicases', Nucleic Acids Research, 34(15), pp. 4154–4159.

Maier, L.-K., Dyall-Smith, M. and Marchfelder, A. (2015) 'The Adaptive Immune System of Haloferax volcanii', Life, 5(1), pp. 521–537.

Maier, L. K. et al. (2019) 'The nuts and bolts of the Haloferax CRISPR-Cas system I-B', RNA Biology. Taylor & Francis, 16(4), pp. 469–480.

Makarova, K. S. et al. (2009) 'Prokaryotic homologs of Argonaute proteins are predicted to function as key components of a novel system of defense against mobile genetic elements', Biology Direct, 4, p. 29.

Makarova, K. S. et al. (2010) 'Evolution of diverse cell division and vesicle formation systems in Archaea', Nature Reviews Microbiology. Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved., 8, p. 731.

Makarova, K. S., Koonin, E. V. and Kelman, Z. (2012) 'The CMG (CDC45/RecJ, MCM, GINS) complex is a conserved component of the DNA replication system in all archaea and eukaryotes', Biology Direct. BioMed Central Ltd, 7(1), p. 7.

Maniscalco, M. et al. (2014) 'Light-dependent expression of four cryptic archaeal circadian gene homologs', Frontiers in Microbiology, 5(MAR), pp. 1–10.

Mankouri, H. W., Craig, T. J. and Morgan, A. (2002) 'SGS1 is a multicopy suppressor of srs2: Functional overlap between DNA helicases', Nucleic Acids Research, 30(5), pp. 1103–1113.

Mankouri, H. W., Ngo, H.-P. and Hickson, I. D. (2007) 'Shu Proteins Promote the Formation of Homologous Recombination Intermediates That Are Processed by Sgs1-Rmi1-Top3', Molecular Biology of the Cell, 19(June), pp. 2609–2619.

Manosas, M. et al. (2013) 'RecG and UvsW catalyse robust DNA rewinding critical for stalled DNA replication fork rescue', Nature Communications. Nature Publishing Group, 4, pp. 1–11.

Maréchal, A. and Zou, L. (2013) 'DNA damage sensing by the ATM and ATR kinases', Cold Spring Harbor Perspectives in Biology, 5(9), pp. 1–17.

Marini, F. and Wood, R. D. (2002) 'A human DNA helicase homologous to the DNA cross-link sensitivity protein Mus308', Journal of Biological Chemistry, 277(10), pp. 8716–8723.

Marriott, H. and Allers, T. (2016) 'Archaea and the meaning of life'.

Marriott, H. (2017) 'Genome Architecture and DNA Replication in Haloferax volcanii', (September).

Martens-Habbena, W. et al. (2011) 'Cdv-based cell division and cell cycle organization in the thaumarchaeon Nitrosopumilus maritimus', Molecular Microbiology, 82(3), pp. 555–566.

Matos, J. et al. (2011) ‘Regulatory control of the resolution of DNA recombination intermediates during meiosis and mitosis’, Cell. Elsevier Inc., 147(1), pp. 158–172.

Matson, S. W., Bean, D. W. and George, J. W. (1994) ‘DNA helicases: Enzymes with essential roles in all aspects of DNA metabolism’, BioEssays, 16(1), pp. 13–22.

Matsumoto, Y. and Kim, K. (1995) ‘Excision of deoxyribose phosphate residues by DNA polymerase β during DNA repair’, Science, 269(5224), pp. 699–702.

Matsuoka, S. et al. (2007) ‘ATM and ATR substrate analysis reveals extensive protein networks responsive to DNA damage’, Science, 316(5828), pp. 1160–1166.

Maupin-Furlow, J. A. (2013) ‘Ubiquitin-like proteins and their roles in archaea’, Trends in Microbiology. Elsevier Ltd, 21(1), pp. 31–38.

Mayle, R. et al. (2015) ‘Mus81 and converging forks limit the mutagenicity of replication fork breakage’, 344(6190).

Mazin, A. et al. (2010) ‘Rad54, the motor of HR’, DNA Repair, 9(3), pp. 286–302.

McClendon, T. B. et al. (2016) ‘Promotion of Homologous Recombination by SWS-1 in Complex with RAD-51 Paralogs in Caenorhabditis elegans.’, Genetics, 203(May), pp. 133–145.

McCready, S. et al. (2005) ‘UV irradiation induces homologous recombination genes in the model archaeon, Halobacterium sp. NRC-1.’, Saline systems, 1, p. 3.

McGlynn, P. and Lloyd, R. G. (2001) ‘Rescue of stalled replication forks by RecG: Simultaneous translocation on the leading and lagging strand templates supports an active DNA unwinding model of fork reversal and holliday junction formation’, Proceedings of the National Academy of Sciences of the United States of America, 98(15), pp. 8227–8234.

McGlynn, P. and Lloyd, R. G. (2002) ‘Recombinational repair and restart of damaged replication forks’, Nature Reviews Molecular Cell Biology, 3(11), pp. 859–870.

McIlwraith, M. J. and West, S. C. (2008) ‘DNA Repair Synthesis Facilitates RAD52-Mediated Second-End Capture during DSB Repair’, Molecular Cell, 29(4), pp. 510–516.

Meetei, A. R. et al. (2005) ‘A human ortholog of archaeal DNA repair protein Hef is defective in Fanconi anemia complementation group M.’, Nature genetics, 37(9), pp. 958–63.

Michel, B. et al. (2001) ‘Rescue of arrested replication forks by homologous recombination’, Proceedings of the National Academy of Sciences of the United States of America, 98(15), pp. 8181–8188.

Michel, B. and Bernander, R. (2014) ‘Chromosome replication origins: Do we really need them?’, BioEssays, 36(6), pp. 585–590.

Mimitou, E. P. and Symington, L. S. (2008) 'Sae2, Exo1 and Sgs1 collaborate in DNA double-strand break processing', 455(7214), pp. 1–13.

Miranda, H. V. et al. (2011) 'E1- and ubiquitin-like proteins provide a direct link between protein conjugation and sulfur transfer in archaea', Proceedings of the National Academy of Sciences of the United States of America, 108(11), pp. 4417–4422.

Mitchell, A. L. et al. (2019) 'InterPro in 2019: Improving coverage, classification and access to protein sequence annotations', Nucleic Acids Research, 47(D1), pp. 351–360.

Modrich, P. (2006) 'Mechanisms in Eukaryotic Mismatch Repair', Biochemistry, 281(41), pp. 30305–30309.

Mohaghegh, P. et al. (2001) 'The Bloom's and Werner's syndrome proteins are DNA structure-specific helicases.', Nucleic acids research, 29(13), pp. 2843–9.

Moldovan, G. L., Pfander, B. and Jentsch, S. (2007) 'PCNA, the Maestro of the Replication Fork', Cell, 129(4), pp. 665–679.

Morimatsu, K. and Kowalczykowski, S. C. (2003) 'RecFOR proteins load RecA protein onto gapped DNA to accelerate DNA strand exchange: A universal step of recombinational repair', Molecular Cell, 11(5), pp. 1337–1347.

Morimatsu, K. and Kowalczykowski, S. C. (2014) 'RecQ helicase and RecJ nuclease provide complementary functions to resect DNA for homologous recombination', Proceedings of the National Academy of Sciences, 111(48), pp. E5133–E5142.

Mullakhanbhai, M. F. and Larsen, H. (1975) 'Halobacterium volcanii spec. nov., a Dead Sea halobacterium with a moderate salt requirement', Archives of Microbiology, 104(1), pp. 207–214.

Muzzini, D. M. et al. (2008) 'Caenorhabditis elegans POLQ-1 and HEL-308 function in two distinct DNA interstrand cross-link repair pathways', DNA Repair, 7(6), pp. 941–950.

Nagata, M. et al. (2017) 'Possible function of the second RecJ-like protein in stalled replication fork repair by interacting with Hef', Scientific Reports. Springer US, 7(1), pp. 1–12.

Nagata, M., Ishino, S., et al. (2017) 'The Cdc45/RecJ-like protein forms a complex with GINS and MCM, and is important for DNA replication in Thermococcus kodakarensis', Nucleic Acids Research. Oxford University Press, 45(18), pp. 10693–10705.

Nair, P. A., Smith, P. and Shuman, S. (2010) 'Structure of bacterial LigD 3′-phosphoesterase unveils a DNA repair superfamily', Proceedings of the National Academy of Sciences of the United States of America, 107(29), pp. 12822–12827.

Nakae, S. et al. (2016) 'Structure of the EndoMS-DNA Complex as Mismatch Restriction Endonuclease', Structure. Elsevier Ltd., 24(11), pp. 1960–1971.

Neale, M. J. and Keeney, S. (2006) 'Clarifying the mechanics of DNA strand exchange in meiotic recombination', Nature, 442(7099), pp. 153–158.

Neuwald, A. F. et al. (1999) 'Assembly , Operation , and Disassembly of Protein Complexes AAA + : A Class of Chaperone-Like ATPases Associated with the Assembly , Operation , and Disassembly of Protein Complexes', (516), pp. 27–43.

Newman, T. J. et al. (2013) 'Replisome stall events have shaped the distribution of replication origins in the genomes of yeasts', Nucleic Acids Research, 41(21), pp. 9705–9718.

Neylon, C. et al. (2005) 'Replication Termination in Escherichia coli: Structure and Antihelicase Activity of the Tus-Ter Complex', 69(3), pp. 501–526.

Nguyen, B. et al. (2014) 'Diffusion of human replication protein a along single-stranded DNA', Journal of Molecular Biology. Elsevier Ltd, 426(19), pp. 3246–3261.

Nilsson, P. et al. (2004) 'Three replication origins in Sulfolobus species: Synchronous initiation of chromosome replication and asynchronous termination', Proceedings of the National Academy of Sciences, 101(18), pp. 7046–7051.

Nishino, T., Komori, K., Tsuchiya, D., et al. (2005) 'Crystal structure and functional implications of Pyrococcus furiosus Hef helicase domain involved in branched DNA processing', Structure, 13(1), pp. 143–153.

Nishino, T., Komori, K., Ishino, Y., et al. (2005) 'Structural and functional analyses of an archaeal XPF/Rad1/Mus81 nuclease: Asymmetric DNA binding and cleavage mechanisms', Structure, 13(8), pp. 1183–1192.

Norais, C. et al. (2007) 'Genetic and physical mapping of DNA replication origins in Haloferax volcanii', PLoS Genetics, 3(5), pp. 729–743.

Northall, S. J. et al. (2017) 'DNA binding and unwinding by Hel308 helicase requires dual functions of a winged helix domain', DNA Repair. Elsevier, 57(July), pp. 125–132.

Nunoura, T. 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 Research, 39(8), pp. 3204–3223.

O'Donnell, M., Langston, L. and Stillman, B. (2013) 'Principles and concepts of DNA replication in bacteria, archaea, and eukarya', Cold Spring Harbor Perspectives in Biology, 5(7).

Ogino, H. et al. (2017) 'The RecJ2 protein in the thermophilic archaeon Thermoplasma acidophilum is a 3′-5′ exonuclease that associates with a DNA replication complex', Journal of Biological Chemistry, 292(19), pp. 7921–7931.

Oh, S. D. et al. (2007) 'BLM Ortholog, Sgs1, Prevents Aberrant Crossing-over by Suppressing Formation of Multichromatid Joint Molecules', Cell, 130(2), pp. 259–272.

Oksenych, V. and Coin, F. (2010) ‘The long unwinding road: XPB and XPD helicases in damaged DNA opening’, Cell Cycle, 9(1), pp. 90–96.

Olsen, G. J. and Woese, C. R. (1997) ‘Archaeal genomics: an overview.’, Cell, 89(7), pp. 991–994.

Olsson, M. and Lindahl, T. (1980) ‘Repair of alkylated DNA in Escherichia coli. Methyl group transfer from O6-methylguanine to a protein cysteine residue’, Journal of Biological Chemistry, 255(22), pp. 10569–10571.

Ordonez, H. and Shuman, S. (2013) ‘Mycobacterium smegmatis Lhr is a DNA-dependent ATPase and a 3′-to-5′ DNA translocase and helicase that prefers to unwind 3′-tailed RNA:DNA hybrids’, Journal of Biological Chemistry, 288(20), pp. 14125–14134.

Oren, A. (1999) ‘Bioenergetic aspects of halophilism.’, Microbiology and molecular biology reviews : MMBR, 63(2), pp. 334–48.

Oren, A. et al. (2002) ‘Intracellular ion and organic solute concentrations of the extremely halophilic bacterium Salinibacter ruber’, Extremophiles, 6(6), pp. 491–498.

Oren, A. (2008) ‘Microbial life at high salt concentrations: phylogenetic and metabolic diversity.’, Saline systems, 4(1), p. 2.

Ortenberg, R., Rozenblatt-Rosen, O. and Mevarech, M. (2000) ‘The extremely halophilic archaeon Haloferax volcanii has two very different dihydrofolate reductases’, Molecular Microbiology, 35(6), pp. 1493–1505.

Oyama, T. et al. (2009) ‘Atomic structures and functional implications of the archaeal RecQ-like helicase Hjm.’, BMC structural biology, 9, p. 2.

Ozaki, S. and Katayama, T. (2009) ‘DnaA structure, function, and dynamics in the initiation at the chromosomal origin’, Plasmid. Elsevier Inc., 62(2), pp. 71–82.

Pace, N. R. (1997) ‘A molecular view of microbial diversity and the biosphere.’, Science (New York, N.Y.), 276(5313), pp. 734–740.

Pages, V. and Fuchs, R. P. (2003) ‘Uncoupling of leading- and lagging-strand DNA replication during lesion bypass in vivo’, Science, 300(5623), pp. 1300–1303.

Pâques, F. and Haber, J. E. (1999) ‘Multiple pathways of recombination induced by double-strand breaks in Saccharomyces cerevisiae.’, Microbiology and molecular biology reviews : MMBR, 63(2), pp. 349–404.

Parmar, K., D’Andrea, A. and Niedernhofer, L. J. (2009) ‘Mouse models of Fanconi anemia.’, Mutation research, 668(1–2), pp. 133–40.

Patel, S. . and Picha, K. M. (2000) ‘Structure and function of hexameric helicases’, Annual review of biochemistry, 69, pp. 651–697.

Paull, T. T. and Gellert, M. (2000) ‘A mechanistic basis for Mre11-directed DNA joining at microhomologies’, Proceedings of the National Academy of Sciences, 97(12), pp. 6409–6414.

Pellegrini, L. (2017) ‘Structural insights into Cdc45 function: was there a nuclease at the heart of the ancestral replisome?’, Biophysical Chemistry. The Author, 225, pp. 10–14.

Pester, M., Schleper, C. and Wagner, M. (2011) ‘The Thaumarchaeota: An emerging view of their phylogeny and ecophysiology’, Current Opinion in Microbiology, 14(3), pp. 300–306.

Petermann, E. and Helleday, T. (2010) ‘Pathways of mammalian replication fork restart’, Nature Reviews Molecular Cell Biology. Nature Publishing Group, 11(10), pp. 683–687.

Petit, C. and Sancar, A. (1999) ‘Nucleotide excision repair: From E. coli to man’, Biochimie, 81(1–2), pp. 15–25.

Phung, D. K. et al. (2013) ‘Archaeal β-CASP ribonucleases of the aCPSF1 family are orthologs of the eukaryal CPSF-73 factor’, Nucleic Acids Research, 41(2), pp. 1091–1103.

Piersen, C. E., McCullough, A. K. and Lloyd, R. S. (2000) ‘AP lyases and dRPases: Commonality of mechanism’, Mutation Research - DNA Repair, 459(1), pp. 43–53.

Pisani, D., Cotton, J. A. and McInerney, J. O. (2007) ‘Supertrees disentangle the chimerical origin of eukaryotic genomes’, Molecular Biology and Evolution, 24(8), pp. 1752–1760.

Pitcher, R. S. et al. (2005) ‘Domain structure of a NHEJ DNA repair ligase from Mycobacterium tuberculosis’, Journal of Molecular Biology, 351(3), pp. 531–544.

Pitcher, R. S., Brissett, N. C. and Doherty, A. J. (2007) ‘Nonhomologous End-Joining in Bacteria: A Microbial Perspective’, Annual Review of Microbiology, 61(1), pp. 259–282.

Pluciennik, A. et al. (2010) ‘PCNA function in the activation and strand direction of MutL endonuclease in mismatch repair’, Proceedings of the National Academy of Sciences, 107(37), pp. 16066–16071.

Prakash, R. et al. (2009) ‘Yeast Mphl helicase dissociates Rad51-made D-loops: Implications for crossover control in mitotic recombination’, Genes and Development, 23(1), pp. 67–79.

Ragunathan, K., Liu, C. and Ha, T. (2012) ‘RecA filament sliding on DNA facilitates homology search’, eLife, 2012(1), pp. 1–14.

Ralf, C., Hickson, I. D. and Wu, L. (2006) ‘The Bloom’s syndrome helicase can promote the regression of a model replication fork’, Journal of Biological Chemistry, 281(32), pp. 22839–22846.

Ramakrishnan, C., Dani, V. S. and Ramasarma, T. (2002) ‘A conformational analysis of Walker motif A [GXXXXGKT (S)] in nucleotide-binding and other proteins The sequence GXXXXGKT / S , popularly known as Walker motif A , is widely believed to be the site for binding nucleotides in many proteins . Examination’, Protein Engineering, 15(10), pp. 783–798.

Ravanat, J. L., Douki, T. and Cadet, J. (2001) 'Direct and indirect effects of UV radiation on DNA and its components', Journal of Photochemistry and Photobiology B: Biology, 63(1–3), pp. 88–102.

Reuter, C. J. and Maupin-furlow, J. a (2004) 'Analysis of Proteasome-Dependent Proteolysis in Haloferax volcanii Cells , Using Short-Lived Green Fluorescent Proteins Analysis of Proteasome-Dependent Proteolysis in Haloferax volcanii Cells , Using Short-Lived Green Fluorescent Proteins †', Society, 70(12), pp. 7530–7538.

Reyes-Lamothe, R., Nicolas, E. and Sherratt, D. J. (2012) 'Chromosome Replication and Segregation in Bacteria', Annual Review of Genetics, 46(1), pp. 121–143.

Richards, J. D., Johnson, K. A., et al. (2008) 'Structure of the DNA repair helicase Hel308 reveals DNA binding and autoinhibitory domains', Journal of Biological Chemistry, 283(8), pp. 5118–5126.

Richards, J. D., Cubeddu, L., et al. (2008) 'The Archaeal XPB Protein is a ssDNA-Dependent ATPase with a Novel Partner', Journal of Molecular Biology, 376(3), pp. 634–644.

Rinke, C. et al. (2013) 'Insights into the phylogeny and coding potential of microbial dark matter', Nature. Nature Publishing Group, 499(7459), pp. 431–437.

Rivera, M. C. and Lake, J. A. (2004) 'The ring of life provides evidence for a genome fusion origin of eukaryotes', Nature, 431(7005), pp. 152–155.

Roberts, J. A., Bell, S. D. and White, M. F. (2003) 'An archaeal XPF repair endonuclease dependent on a heterotrimeric PCNA', Molecular Microbiology, 48(2), pp. 361–371.

Rocha, E. P. C., Cornet, E. and Michel, B. (2005) 'Comparative and evolutionary analysis of the bacterial homologous recombination systems', PLoS Genetics, 1(2), pp. 0247–0259.

Saini, N. et al. (2013) 'Migrating bubble during break-induced replication drives conservative DNA synthesis', Nature. Nature Publishing Group, 502(7471), pp. 389–392.

Saintigny, Y. et al. (2002) 'Homologous Recombination Resolution Defect in Werner Syndrome', Molecular and Cellular Biology, 22(20), pp. 6971–6978.

Sakofsky, C. J., Ayyar, S. and Malkova, A. (2012) 'Break-induced replication and genome stability', Biomolecules, 2(4), pp. 483–504.

Samson, R. Y. et al. (2008) 'Europe PMC Funders Group Europe PMC Funders Author Manuscripts A Role for the ESCRT System in Cell Division in Archaea', Science, 322(5908), pp. 1710–1714.

San Filippo, J., Sung, P. and Klein, H. (2008) 'Mechanism of Eukaryotic Homologous Recombination', Annual Review of Biochemistry, 77(1), pp. 229–257.

Sancar, A. et al. (2004) ‘Molecular Mechanisms of Mammalian DNA Repair and the DNA Damage Checkpoints’, Annual Review of Biochemistry, 73(1), pp. 39–85.

Sartori, A. A. and Jiricny, J. (2003) ‘Enzymology of Base Excision Repair in the Hyperthermophilic Archaeon Pyrobaculum aerophilum’, Journal of Biological Chemistry, 278(27), pp. 24563–24576.

Schaaper, R. M. (1993) ‘Base selection, proofreading, and mismatch repair during DNA replication in Escherichia coli’, Journal of Biological Chemistry, 268(32), pp. 23762–23765.

Scharer, O. D. (2013) ‘Nucleotide Excision Repair in Eukaryotes’, Encyclopedia of Biological Chemistry: Second Edition, pp. 341–344.

Schröder, E., Jönsson, T. and Poole, L. (2003) ‘Hydroxyapatite chromatography: altering the phosphate-dependent elution profile of protein as a function of pH’, Analytical Biochemistry, 313(1), pp. 176–178.

Schwartz, E. K. and Heyer, W. D. (2011) ‘Processing of joint molecule intermediates by structure-selective endonucleases during homologous recombination in eukaryotes’, Chromosoma, 120(2), pp. 109–127.

Sclafani, R. A. and Holzen, T. M. (2007) ‘Cell Cycle Regulation of DNA Replication’, Annual Review of Genetics, 41(1), pp. 237–280.

Sehorn, M. G. et al. (2004) ‘Human meiotic recombinase Dmc1 promotes ATP-dependent homologous DNA strand exchange’, Nature, 429(6990), pp. 433–437.

Seigneur, M. et al. (1998) ‘RuvAB acts at arrested replication forks’, Cell, 95(3), pp. 419–430.

Sfeir, A. and Symington, L. S. (2015) ‘Microhomology-mediated end joining: a back-up survival mechanism or dedicated pathway?’, 344(6188), pp. 1173–1178.

Sharples, G. J., Ingleston, S. M. and Lloyd, R. G. (1999) ‘Holliday junction processing in bacteria: Insights from the evolutionary conservation of RuvABC, RecG, and RusA’, Journal of Bacteriology, 181(18), pp. 5543–5550.

Shen, J. C. and Loeb, L. A. (2000) ‘The Werner syndrome gene: The molecular basis of RecQ helicase-deficiency diseases’, Trends in Genetics, 16(5), pp. 213–220.

Shereda, R. D. et al. (2008) ‘SSB as an organizer/mobilizer of genome maintenance complexes’, Critical Reviews in Biochemistry and Molecular Biology, 43(5), pp. 289–318.

Sherwood, K. E., Cano, D. J. and Maupin-Furlow, J. A. (2009) ‘Glycerol-mediated repression of glucose metabolism and glycerol kinase as the sole route of glycerol catabolism in the haloarchaeon Haloferax volcanii’, Journal of Bacteriology, 191(13), pp. 4307–4315.

Simmons, E. (2018) Genetic and Biochemical Analysis of the Archaeal Helicase Hel308.

Singleton, M. R. et al. (2004) ‘Crystal structure of RecBCD enzyme reveals a machine for processing DNA breaks’, Nature, 432(7014), pp. 187–193.

Singleton, M. R., Dillingham, M. S. and Wigley, D. B. (2007) ‘Structure and mechanism of helicases and nucleic acid translocases.’, Annual review of biochemistry, 76, pp. 23–50

Singleton, M. R., Scaife, S. and Wigley, D. B. (2001) ‘Structural analysis of DNA replication fork reversal by RecG’, Cell, 107(1), pp. 79–89.

Singleton, M. R. and Wigley, D. B. (2002) ‘Modularity and specialization in superfamily 1 and 2 helicases’, Journal of Bacteriology, 184(7), pp. 1819–1826.

Sinha, R. P. and Häder, D. P. (2002) ‘UV-induced DNA damage and repair: A review’, Photochemical and Photobiological Sciences, 1(4), pp. 225–236..

Smith, J. and Modrich, P. (1996) ‘Mutation detection with MutH, MutL, and MutS mismatch repair proteins’, Proceedings of the National Academy of Sciences of the United States of America, 93(9), pp. 4374–4379.

Smith, P. et al. (2011) ‘Structures and activities of archaeal members of the LigD 3’-phosphoesterase DNA repair enzyme superfamily’, Nucleic Acids Research, 39(8), pp. 3310–3320.

Snider, J., Thibault, G. and Houry, W. A. (2008) ‘The AAA+ superfamily of functionally diverse proteins’, Genome Biology, 9(4), p. 216.

Sommers, J. A. et al. (2009) ‘FANCJ uses its motor ATPase to destabilize protein-DNA complexes, unwind triplexes, and inhibit RAD51 strand exchange’, Journal of Biological Chemistry, 284(12), pp. 7505–7517.

Soultanas, P. and Wigley, D. B. (2001) ‘Unwinding the “Gordian knot” of helicase action’, Trends in Biochemical Sciences, 26(1), pp. 47–54.

Spang, A. et al. (2015) ‘Complex archaea that bridge the gap between prokaryotes and eukaryotes’, Nature, 521(7551), pp. 173–179.

Spang, A., Caceres, E. F. and Ettema, T. J. G. (2017) ‘Genomic exploration of the diversity, ecology, and evolution of the archaeal domain of life’, Science, 357(6351).

Stantial, N. et al. (2016) ‘Transcription-coupled repair of UV damage in the halophilic archaea.’, DNA repair. Elsevier B.V., 41, pp. 63–68.

Stark, J. M. et al. (2004) ‘Genetic Steps of Mammalian Homologous Repair with Distinct Mutagenic Consequences’, Molecular and Cellular Biology, 24(21), pp. 9305–9316.

Stroud, A., Liddell, S. and Allers, T. (2012) ‘Genetic and biochemical identification of a novel single-stranded DNA-binding complex in Haloferax volcanii’, Frontiers in Microbiology, 3(JUN), pp. 1–14.

Stubbe, J. and Riggs-Gelasco, P. (1998) ‘Harnessing free radicals: Formation and function of the tyrosyl radical in ribonucleotide reductase’, Trends in Biochemical Sciences, 23(11), pp. 438–443.

Stuchinskaya, T. et al. (2009) ‘Public Access NIH Public Access’, J Mol Biol, 32(7), pp. 736–740.

Sugasawa, K. et al. (1996) ‘HHR23B, a human Rad23 homolog, stimulates XPC protein in nucleotide excision repair in vitro.’, Molecular and Cellular Biology, 16(9), pp. 4852–4861.

Sugasawa, K. et al. (2001) ‘Journal of Industrial and Engineering Chemistry Phenyl VOCs catalytic combustion on supported CoMn / AC oxide catalyst’, Journal of Industrial and Engineering Chemistry, 21, pp. 932–941.

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 and Development, 11(9), pp. 1111–1121.

Sung, P. et al. (2003) ‘Rad51 Recombinase and Recombination Mediators’, Journal of Biological Chemistry, 278(44), pp. 42729–42732.

Svendsen, J. M. et al. (2009) ‘Mammalian BTBD12/SLX4 assembles a Holliday junction resolvase and is required for DNA repair’, 138(1), pp. 63–77.

Symington, L. S. and Gautier, J. (2011) ‘Double-Strand Break End Resection and Repair Pathway Choice’, Annual Review of Genetics, 45(1), pp. 247–271.

Tafel, A. A., Wu, L. and McHugh, P. J. (2011) ‘Human HEL308 localizes to damaged replication forks and unwinds lagging strand structures’, Journal of Biological Chemistry, 286(18), pp. 15832–15840.

Takata, K. et al. (2013) ‘Human DNA helicase HELQ participates in DNA interstrand crosslink tolerance with ATR and RAD51 paralogs’, Nat Commun. Nature Publishing Group, 4, p. 2338.

Taniguchi, T. et al. (2002) ‘S-phase-specific interaction of the Fanconi anemia protein, FANCD2, with BRCA1 and RAD51’, Blood, 100(7), pp. 2414–2420.

Tapias, A. et al. (2004) ‘Ordered Conformational Changes in Damaged DNA Induced by Nucleotide Excision Repair Factors’, Journal of Biological Chemistry, 279(18), pp. 19074–19083.

Taylor, M. R. G. et al. (2015) ‘Rad51 Paralogs Remodel Pre-synaptic Rad51 Filaments to Stimulate Homologous Recombination’, Cell, 162(2), pp. 271–286.

Teng, G. and Papavasiliou, F. N. (2007) ‘Immunoglobulin Somatic Hypermutation’, Annual Review of Genetics, 41(1), pp. 107–120.

Thangavel, S. et al. (2015) ‘DNA2 drives processing and restart of reversed replication forks in human cells’, Journal of Cell Biology, 208(5), pp. 545–562.

Tomasz, M. (1995) ‘Mitomycin C: small, fast and deadly (but very selective)’, Chemistry and Biology, 2(9), pp. 575–579.

Torres, R., Serrano, E. and Alonso, J. C. (2019) 'Bacillus subtilis RecA interacts with and loads RadA/Sms to unwind recombination intermediates during natural chromosomal transformation', Nucleic Acids Research, pp. 1–18.

Turner, J. et al. (1999) 'The internal workings of a DNA polymerase clamp-loading machine', EMBO Journal, 18(3), pp. 771–783.

Tuteja, N. and Tuteja, R. (2004) 'Unraveling DNA helicases Motif, structure, mechanism and function', European Journal of Biochemistry, 271(10), pp. 1849–1863.

Veaute, X. et al. (2003) 'The Srs2 helicase prevents recombination by disrupting Rad51 nucleoprotein filaments', Nature, 423(6937), pp. 309–312.

Venkitaraman, A. R. (2002) 'Cancer Susceptibility and the Functions of BRCA1 and BRCA2', A Companion to the Literature and Culture of the American West, 108, pp. 409–424.

Wadsworth, R. I. M. (2002) 'Identification and properties of the crenarchaeal single-stranded DNA binding protein from Sulfolobus solfataricus', Nucleic Acids Research, 29(4), pp. 914–920.

Walker, J. R., Corpina, R. A. and Goldberg, J. (2001) 'Structure of the Ku heterodimer bound to dna and its implications for double-strand break repair', Nature, 412(6847), pp. 607–614.

Walters, A. D. and Chong, J. P. J. (2017) 'Non-essential MCM-related proteins mediate a response to DNA damage in the archaeon Methanococcus maripaludis', Microbiology (United Kingdom), 163(5), pp. 745–753.

Ward, J. D. et al. (2010) 'Overlapping Mechanisms Promote Postsynaptic RAD-51 Filament Disassembly during Meiotic Double-Strand Break Repair', Molecular Cell. Elsevier Ltd, 37(2), pp. 259–272.

Wardell, K. (2013) 'Regulation of Homologous Recombination in the Archaeon Haloferax volcanii', (September).

Wardell, K. et al. (2017) 'RadB acts in homologous recombination in the archaeon Haloferax volcanii , consistent with a role as recombination mediator', DNA Repair. Elsevier, 55(March), pp. 7–16.

Weller, G. R. et al. (2002) 'Identification of a DNA nonhomologous end-joining complex in bacteria', Science, 297(5587), pp. 1686–1689.

Wendel, B. M. et al. (2017) 'SbcC-SbcD and ExoI process convergent forks to complete chromosome replication', Proceedings of the National Academy of Sciences of the United States of America, 115(2), pp. 349–354.

Wendoloski, D., Ferrer, C. and Dyall-Smith, M. L. (2001) 'A new simvastatin (mevinolin)-resistance marker from Haloarcula hispanica and a new Haloferax volcanii strain cured of plasmid pHV2', Microbiology, 147(4), pp. 959–964.

Wendorff, T. J. and Berger, J. M. (2018) ‘Topoisomerase VI senses and exploits both DNA crossings and bends to facilitate strand passage’, eLife, 7, pp. 1–35.
West, S. C. (1997) ‘ Processing of Recombination Intermediates by the Ruv ABC Proteins’, Annual Review of Genetics, 31(1), pp. 213–244.
Whitby, M. C. (2010) ‘The FANCM family of DNA helicases/translocases’, DNA Repair. Elsevier B.V., 9(3), pp. 224–236.
White, M. F. (2003) ‘Archaeal DNA repair: Paradigms and puzzles’, Biochemical Society Transactions, 31(3), pp. 690–693.
White, M. F. (2011) ‘Homologous recombination in the archaea: the means justify the ends’, Biochemical Society Transactions, 39(1), pp. 15–19.
White, M. F. and Allers, T. (2018) ‘DNA repair in the archaea-an emerging picture’, FEMS microbiology reviews, 42(4), pp. 514–526.
Williams, T. A. et al. (2013) ‘An archaeal origin of eukaryotes supports only two primary domains of life’, Nature. Nature Publishing Group, 504(7479), pp. 231–236.
Wilson, D. M. and Bohr, V. A. (2007) ‘The mechanics of base excision repair, and its relationship to aging and disease’, DNA Repair, 6(4), pp. 544–559.
Wilson, M. A. et al. (2013) ‘Pif1 helicase and Polδ promote recombination-coupled DNA synthesis via bubble migration’, Nature, 502(7471), pp. 393–396.
Winker, S. and Woese, C. R. (1991) ‘A Definition of the Domains Archaea, Bacteria and Eucarya in Terms of Small Subunit Ribosomal RNA Characteristics’, Systematic and Applied Microbiology, 14(4), pp. 305–310.
Woese, C. R. and Fox, G. E. (1977) ‘Phylogenetic structure of the prokaryotic domain: the primary kingdoms.’, Proceedings of the National Academy of Sciences of the United States of America, 74(11), pp. 5088–5090.
Woese, C. R., Kandler, O. and Wheelis, M. L. (1990) ‘Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya.’, Proceedings of the National Academy of Sciences of the United States of America, 87(12), pp. 4576–4579.
Woodman, I. L. and Bolt, E. L. (2009) ‘Molecular biology of Hel308 helicase in archaea.’, Biochemical Society transactions, 37(Pt 1), pp. 74–78.
Woodman, I. L. and Bolt, E. L. (2011) ‘Winged helix domains with unknown function in Hel308 and related helicases.’, Biochemical Society transactions, 39(1), pp. 140–144.
Woodman, I. L., Brammer, K. and Bolt, E. L. (2011) ‘Physical interaction between archaeal DNA repair helicase Hel308 and Replication Protein A (RPA)’, DNA Repair. Elsevier B.V., 10(3), pp. 306–313.

Woodman, I. L., Briggs, G. S. and Bolt, E. L. (2007) 'Archaeal Hel308 Domain V Couples DNA Binding to ATP Hydrolysis and Positions DNA for Unwinding Over the Helicase Ratchet', Journal of Molecular Biology. Elsevier Ltd, 374(5), pp. 1139–1144.

Woods, W. G. and Dyall-Smith, M. L. (1997) 'Construction and analysis of a recombination-deficient (radA) mutant of Haloferax volcanii', Molecular Microbiology, 23(4), pp. 791–797.

Wu, L. and Hickson, I. O. (2003) 'The Bloom's syndrome helicase suppresses crossing over during homologous recombination', Nature, 426(6968), pp. 870–874.

Wu, Y. et al. (2015) 'EEPD1 Rescues Stressed Replication Forks and Maintains Genome Stability by Promoting End Resection and Homologous Recombination Repair', PLoS Genetics, 11(12), pp. 1–29.

Yi, G. S. et al. (2017) 'Two archaeal RecJ nucleases from methanocaldococcus jannaschii show reverse hydrolysis polarity: Implication to their unique function in archaea', Genes, 8(9), pp. 1–15.

Yin, J. et al. (2006) 'Crystal structure of the vitamin B12 biosynthetic cobaltochelatase, CbiXS, from Archaeoglobus fulgidus', Journal of Structural and Functional Genomics, 7(1), pp. 37–50.

Zhai, B. et al. (2017) 'Structure and Function of a Novel ATPase that Interacts with Holliday Junction Resolvase Hjc and Promotes Branch Migration', Journal of Molecular Biology. Elsevier Ltd, 429(7), pp. 1009–1029.

Zhai, B. et al. (2018) 'The archaeal ATPase PINA interacts with the helicase Hjm via its carboxyl terminal KH domain remodeling and processing replication fork and Holliday junction', Nucleic Acids Research. Oxford University Press, 46(13), pp. 6627–6641.

Zhang, C. et al. (2013) 'Genetic manipulation in Sulfolobus islandicus and functional analysis of DNA repair genes.', Biochemical Society transactions, 41, pp. 405–10.

Zhao, A., Gray, F. C. and MacNeill, S. A. (2006) 'ATP- and NAD+-dependent DNA ligases share an essential function in the halophilic archaeon Haloferax volcanii', Molecular Microbiology, 59(3), pp. 743–752.

Zheng, L. and Shen, B. (2011) 'Okazaki fragment maturation: Nucleases take centre stage', Journal of Molecular Cell Biology, 3(1), pp. 23–30.

Zhou, G. et al. (2010) 'Ubiquitin-like small archaeal modifier proteins (SAMPs) in Haloferax volcanii', Nature. Nature Publishing Group, 463(7277), pp. 54–60.

Zhou, Z. et al. (2018) 'Two or three domains: a new view of tree of life in the genomics era', Applied Microbiology and Biotechnology. Applied Microbiology and Biotechnology, 102(7), pp. 3049–3058.

Zou, Y. and Van Houten, B. (1999) 'Strand opening by the UvrA2B complex allows dynamic recognition of DNA damage', EMBO Journal, 18(17), pp. 4889–4901.

Zuker, M. (2003) ‘Mfold web server for nucleic acid folding and hybridization prediction’, Nucleic Acids Research, 31(13), pp. 3406–3415.

Appendix 1: Screening for Interacting partners of Hel308 Mass Spectrometry Results

Band a (~100 kDa)

Protein accesion	Identified proteins	HVO number	Predicted MW (Da)	MASCOT score	No. of peptides	Peptide sequences
D4GYK9	**Hel308a**	0014	90355	4664	415(253)	66(49)
D4GWM8	**Cdc48a**	2380	81989	2385	69(43)	42(28)
D4GZY6/ D4GWR0	Elongation factor 1 α1/2	0359/ 2413	45759/ 45628	2170	87(47)	37(25)
D4GXM0	**RecJ4**	2889	79159	2168	60(42)	33(28)
Q1XBW2	DnaK	1590	67250	1661	57(30)	25(17)
D4GY19	Dsa1	2960	54912	1301	40(19)	23(11)
D4GPB9	Smc-like protein Sph2	B0118	73801	1301	23(13)	23(13)
D4GTB5	BccA	2486	65346	1269	23(19)	21(17)
D4GZ00	**Top6B**	1571	87438	1248	34(15)	22(12)
D4GXP9	GatE	2902	67597	1179	26(19)	20(14)
D4GTI4	CitB1	1955	70406	1120	30(21)	16(12)
D4GUG0	Phosphoenolpyruvate carboxylase	2621	102399	1058	22(13)	20(11)
O30561	cct1	0133	58925	1049	24(16)	14(11)
D4GUC3	metG	0809	81936	1034	20(15)	17(13)
D4GUD0	Phosphoenolpyruvate synthase	0812	83133	995	20(10)	19(10)
D4GVH7	acdA	1000	74554	986	24(15)	18(11)
D4GR47	hyuA2	A0379	72968	838	16(10)	14(9)
D4GU82	Hypothetical	2069	89505	784	12(10)	11(10)
D4GT20	**topA**	0681	93686	763	14(8)	14(8)
D4GYP9	glyS	0054	64952	744	12(11)	10(10)
D4GYQ7	dppA1	0062	67307	685	39(19)	11(9)
D4GSA2	katG	1778	80004	671	18(10)	13(8)
O30560	cct2	0455	59298	665	13(6)	12(6)
D4GXK6	DUF814 domain protein	2883	78933	662	12(6)	12(6)
D4GXF1	porA	1305	68470	612	15(8)	10(6)
D4GT75	pccB2	2471	56772	565	10(6)	10(6)
D4GT33	**nrdA ribonucleotide reductase**	2452	114376	560	12(7)	11(6)
P25062	csg	2072	85189	515	10(8)	5(5)
D4GZX6	**rpoA1**	0349	108913	404	8(5)	8(5)
D4GTT9	lon	0783	75482	344	8(2)	8(2)
D4GRT2	cadA	A0624	92544	332	7(3)	6(3)
D4GP53	chllD	B0051	73309	329	9(2)	8(2)
Q48332	atpA	0316	64515	313	7(4)	6(3)
D4GXK9	rpap1	1337	64868	295	4(2)	4(2)
D4GUV5	korA	0888	64026	261	6(3)	6(3)
D4GYH4	aglB	1530	113675	258	6(2)	6(2)
D4GV93	Muc19 precursor	2160	232813	256	5(3)	5(3)

D4GX08	hcpE	1228	83702	251	5(2)	5(2)
D4H037	Putative iron-sulphur protein	1692	80206	244	6(2)	6(2)
D4GZ02	**gyrA**	1573	94853	206	5(1)	5(1)
Q9V2V6/ Q9V2V5	Protease subunit α1/2	1091/29 23	27613/267 27	201	4(3)	4(3)
D4GUW1	acs1	0894	74351	186	4(2)	4(2)
D4GTG8/ D4GSE8	**MutLa/b**	1939/05 51	59769/787 29	180	4(2)	4(2)

Band b (~100 kDa)

Protein accesion	**Identified proteins**	**HVO number**	**Predicted MW (Da)**	**MASCOT score**	**No. of peptides**	**Peptide sequences**
D4GYK9	**Hel308a**	0014	90355	4986	493(306)	68(51)
D4GZY6/ D4GWR0	Elongation factor 1 α1/2	0359/ 2413	45759/ 45628	2864	159(87)	47(30)
D4GXM0	**RecJ4**	2889	79159	2563	78(64)	35(31)
D4GWM8	**Cdc48a**	2380	81989	1802	40(27)	33(22)
D4GPB9	Smc-like protein Sph2	B0118	73801	1722	36(27)	27(19)
D4GUD0	Phosphoenolpyruvate synthase	0812	83133	1611	35(23)	28(22)
D4GZ00	**Top6B**	1571	87438	1593	50(27)	26(18)
D4GUG0	Phosphoenolpyruvate carboxylase	2621	102399	1573	31(21)	27(18)
Q1XBW2	DnaK	1590	67250	1536	53(30)	24(15)
D4GT33	**nrdA**	2452	114376	1369	27(18)	23(15)
D4GTI4	CitB1	1955	70406	1336	47(31)	19(15)
D4GXP9	GatE	2902	67597	1227	29(18)	20(13)
D4GVH7	acdA	1000	74554	1213	36(25)	21(14)
D4GUC3	metG	0809	81936	1129	22(19)	17(15)
D4GR47	hyuA2	A0379	72968	1114	20(17)	16(15)
D4GT20	**topA**	0681	93686	1114	21(14)	20(13)
D4GSA2	katG	1778	80004	1052	29(18)	20(13)
D4GXK6	DUF814 domain protein	2883	78933	974	19(12)	16(11)
O30561	cct1	0133	58925	823	17(15)	11(10)
D4GZX6	**rpoA1**	0349	108913	762	16(10)	14(9)
D4GYP9	glyS	0054	64952	762	13(13)	10(10)
D4GTG8/ D4GSE8	**MutLa/b**	1939/ 0551	59769/ 78729	761	12(9)	11(9)
O30560	cct2	0455	59298	676	12(7)	12(7)
D4GXF1	porA	1305	68470	672	16(8)	11(7)
D4GYQ7	dppA1	0062	67307	652	38(18)	10(7)
D4GSQ3	maeB2	2436	81459	591	11(9)	9(9)
D4GZ78	ppk1	1650	93657	589	12(6)	12(6)
D4GZ02	**gyrA**	1573	94853	577	10(8)	10(8)
D4GZU0	atpI	0311	79504	551	11(8)	11(8)
D4GY19	dsa1	2960	54912	534	13(5)	11(4)
D4GP53	chllD	B0051	73309	518	13(7)	10(5)
D4GU82	Hypothetical	2069	89505	477	7(6)	7(6)
D4GVM4	purL	1048	72964	442	8(5)	8(5)
D4H037	Putative iron-sulphur protein	1692	80206	423	10(6)	9(5)

P25062	csg	2072	85189	383	5(4)	5(4)
D4GXK9	rpap1	1337	64868	361	6(5)	5(4)
D4GUV5	korA	0888	64026	297	5(5)	5(5)
D4GXY5	valS	2945	98944	261	5(2)	5(2)
D4GUB6	pykA	0806	62042	254	7(4)	5(4)
D4GY98	ligN	3000	75811	251	5(4)	5(4)
D4GYD4	gnaD	1488	45794	251	4(4)	4(4)
D4GZ84	DUF460 domain protein	0140	77555	249	5(3)	4(3)
D4GTB5	bccA	2486	65346	248	5(3)	5(3)
D4GZV3	argS	0324	65034	242	5(3)	5(3)
D4GUW1	acs1	0894	74351	201	4(2)	4(2)
D4H019	**uvrA**	0393	108678	192	4(2)	4(2)
D4GSD5	Sensor box histidine kinase	1811	94101	188	5(2)	5(2)

Band c (~50 kDa)

Protein accesion	**Identified proteins**	**HVO number**	**Predicted MW (Da)**	**MASCOT score**	**No. of peptides**	**Peptide sequences**
D4GZY6/ D4GWR0	Elongation factor 1 α1/2	0359/ 2413	45759/ 45628	1987	139(83)	29(20)
D4GYK9	**Hel308a**	0014	90355	1731	34(25)	27(22)
D4GS20	pgk	0480	43636	1340	24(18)	18(15)
D4GVD7	**CetZ1**	2204	42064	1169	29(22)	17(13)
D4GY38	thrC3	2969	44404	1084	19(15)	15(11)
D4GWX0	enolase	2774	42005	1031	38(24)	14(10)
D4GV06	FAD-dependent oxidoreductase (Homolog to geranylgeranyl reductase)	2650	50690	1022	23(12)	17(9)
D4GUU7	Coiled-coil protein	0880	36337	893	31(15)	16(11)
D4GVQ6	GFO family oxidoreductase	2690	40028	886	31(19)	14(11)
Q48327	**FtsZ1**	0717	39805	873	29(18)	14(10)
D4GXM6	GTPase HflX	1346	48212	852	16(10)	15(10)
D4GT62	sucC	2465	41005	842	31(22)	13(11)
D4GW49	**ribonuclease J**	2724	50115	814	25(14)	16(12)
D4GXG2	glyA1	2862	44349	808	19(13)	14(9)
D4GU92	icd	2588	45810	807	21(12)	15(9)
D4GZV0	arf1	0321	46582	789	17(10)	14(9)
D4GVH6	DRTGG domain protein	0999	39826	774	24(14)	12(9)
D4GYV5	sufS	0109	46552	709	14(11)	12(9)
D4GYD4	gnaD	1488	45794	676	20(16)	11(10)
D4GVT3	purB	1084	50817	660	13(9)	12(8)
D4GVU2	trpD2	2226	39646	660	25(16)	13(8)
D4GZX7	**rpoA2**	0350	46140	652	17(10)	11(6)
D4GXX3	tsgA11	1401	39212	648	11(10)	7(7)
D4GS04	ilvA1	0464	43680	642	14(8)	11(7)
D4GYZ9	**top6A**	1570	41829	641	11(8)	10(7)
D4GUJ7	panA	0850	45802	626	15(7)	13(5)
D4GUK7	sufB1	0860	53247	616	13(10)	10(7)

D4GXW6	M20 family amidohydrolase	1395	45410	570	10(7)	9(7)
D4GVK2	acaB2	1025	40732	559	10(7)	10(7)
D4GV77	rbcL	0970	43922	553	13(7)	10(6)
D4GPB3	Mandelate racemase	B0111	44272	547	16(7)	12(6)
D4GWT1	Peptidase M42 family protein	2759	37046	503	10(5)	8(5)
D4GZX4	rpoB2	0347	58897	474	10(6)	9(6)
D4GVT2	gdh	1083	38990	466	10(7)	9(6)
D4GYK3	lysC	0008	41700	459	8(4)	7(4)
D4GUA4	stomatin family protein	0801	46401	453	9(6)	9(6)
D4GW95	metE2	2743	36800	436	7(5)	7(5)
D4GZT6	Hypothetical	0307	49060	436	10(4)	10(4)
D4GW14	YfiH family protein	2708	35406	429	9(7)	6(4)
Q977V2	srp54	0123	50918	426	7(4)	7(4)
D4GTE0	**cdc48d**	1907	53410	418	11(3)	9(3)
O30561	cct1	0133	58925	412	6(5)	6(5)
D4GTE5	kef1	1916	66269	400	6(4)	6(4)
D4GW10	GFO family oxidoreductase	2707	41032	395	8(4)	8(4)
D4GYJ0	dhaK	1546	34847	388	8(3)	7(3)
D4GRS7	Hypothetical	A0619	48181	387	8(3)	8(3)
D4GTX4	rpl15	2542	17828	381	6(4)	6(4)
D4GUK7	sufB1	0860	53247	379	8(4)	7(4)
D4GW79	Hypothetical	1129	43003	364	7(4)	7(4)
D4GYN8	argD	0043	39767	362	6(5)	5(4)
D4GY69	gdhA2	1453	45730	353	8(3)	8(3)
D4GP49	ABC-type transport system periplasmic substrate-binding protein	B0047	44239	351	6(3)	6(3)
D4GQT3	DUF234 domain protein	A0254	52691	344	8(3)	8(3)
D4GW23	rtcB	2712	53434	331	7(3)	7(3)
Q48328	**RadA**	0104	37320	328	8(4)	7(3)
D4GYS6	hemL	0081	47811	323	6(5)	6(5)
D4GZ82	tyrS	0138	36111	323	8(4)	6(3)
O30560	cct2	0455	59298	321	5(4)	5(4)
D4GST7	pepB2	1829	39365	318	6(4)	6(4)
D4GYK5	tnaA	0009	48543	317	7(2)	7(2)
D4GP54	PQQ repeat protein	B0052	44110	309	6(2)	6(2)
D4GT99	HD family hydrolase	2481	42772	309	7(3)	7(3)
D4H0E4	**orc8**	C0057	45953	303	8(4)	6(4)
D4GUN8	Pyridoxal phosphate-dependent aminotransferase	2091	47993	292	5(3)	5(3)
D4GRM1	hutI	A0560	42409	290	4(3)	4(3)
D4GW51	idsA1	2725	38205	289	5(5)	4(4)
D4GUC9	PQQ repeat protein	2606	46919	286	6(2)	6(2)
D4GYQ3	dppF1	0058	50184	267	5(4)	5(4)

Band d (~150 kDa)

Protein accesion	Identified proteins	HVO number	Predicted MW (Da)	MASCOT score	No. of peptides	Peptide sequences
D4GYK9	**Hel308a**	0014	90355	3276	133(91)	48(39)
D4GP52	**cobN**	B0050	141254	2467	86(46)	43(28)
D4GWK9	Carbamoyl-phosphate synthase large chain	2361	117302	2360	95(63)	34(27)
D4GT33	**nrdA**	2452	114376	2038	45(30)	35(24)
D4GZX6	**rpoA1**	0349	108913	1926	48(32)	28(25)
D4GWM8	**Cdc48a**	2380	81989	1883	42(29)	33(24)
D4GRZ3	leuS	0452	109563	1821	52(41)	28(22)
D4GYJ1	ileS	1547	117994	1773	42(29)	26(21)
D4GUK1	**rad50**	0854	100087	1619	29(20)	28(20)
D4H056	sga1	1711	166218	1586	33(17)	32(17)
D4GV93	Muc19 precursor	2160	232813	1469	29(25)	17(16)
O30561	cct1	0133	58925	1460	44(36)	17(15)
D4GXY5	valS	2945	98944	1457	30(23)	19(15)
O30560	cct2	0455	59298	1404	36(27)	21(18)
D4GZY6/ D4GWR0	Elongation factor 1 α1/2	0359/ 2413	45759/ 45628	1309	38(18)	25(15)
D4GSM3	dppDF2	0627	97733	1283	26(16)	21(15)
D4GZF1	alaS1	0206	102464	1261	28(25)	21(19)
D4GPF6	Receiver/ sensor box histidine kinase	B0154	144236	1170	21(14)	20(14)
D4H042	FAD-dependent oxidoreductase	1697	111281	1136	20(15)	18(14)
D4GXM0	**RecJ4**	2889	79159	1130	22(16)	19(16)
D4GZ00	**Top6B**	1571	87438	1083	25(13)	19(11)
D4H019	**uvrA**	0393	108678	1005	18(13)	18(13)
D4GWU7	aldH2	1189	56171	939	22(10)	17(9)
D4GS73	citB2	0541	102210	927	23(9)	19(8)
D4GUK7	sufB1	0860	53247	890	29(19)	16(13)
D4GTI4	CitB1	1955	70406	868	18(12)	13(11)
P25062	csg	2072	85189	815	45(39)	8(7)
D4GXF1	porA	1305	68470	788	18(12)	11(9)
D4GUV5	korA	0888	64026	781	20(15)	14(11)
D4GUG0	ppc	2621	102399	772	16(9)	14(8)
D4GVS7	**cdc48b**	2700	82817	748	15(10)	14(9)
Q1XBW2	DnaK	1590	67250	694	11(9)	10(9)
D4GZ02	**gyrA**	1573	94853	669	13(9)	13(9)
D4GU92	icd	2588	45810	666	12(8)	11(8)
D4GUC3	metG	0809	81936	654	10(7)	10(7)
D4GYH4	aglB	1530	113675	652	13(8)	12(8)
D4GVH7	acdA	1000	74554	644	13(9)	11(8)
D4GUL6	gltB	0869	163908	607	14(6)	13(6)
D4GT75	pccB2	2471	56772	560	9(6)	9(6)
D4GUK8	sufB2	0861	44685	549	15(11)	9(7)
D4GY19	Dsa1	2960	54912	532	11(4)	10(4)
D4GXP5	fumC	2900	50093	529	11(6)	10(6)

D4GT20	**topA**	0681	93686	520	10(5)	10(5)
D4GYQ7	dppA1	0062	67307	498	12(7)	8(5)
D4GVJ5	**RecJ3**	1018	69793	481	11(6)	10(6)
D4GT59	sucD	2464	29864	478	10(8)	7(5)
D4GZX4	**rpoB2**	0347	58897	465	11(6)	10(5)
Q9V2V6/ Q9V2V5	Protease subunit α1/2	1091/ 2923	27613/ 26727	459	10(4)	9(4)
D4GTB5	BccA	2486	65346	452	7(6)	7(6)
D4GVF7	nuoCD	0980	63484	451	9(5)	9(5)
D4GW49	**ribonuclease J**	2724	50115	444	10(7)	9(7)
D4GPB9	Smc-like protein Sph2	B0118	73801	439	8(5)	8(5)
D4GUM1	mRNA 3-end processing factor-like protein	0874	71947	438	9(6)	9(6)
D4GXG2	glyA1	2862	44349	430	7(5)	7(5)
D4GT30	smc	0689	138186	420	8(5)	8(5)
D4GYD4	gnaD	1488	45794	412	10(7)	6(5)
D4GVZ2	Hypothetical	2276	161868	404	9(4)	9(4)
D4GUJ7	panA	0850	45802	399	8(4)	8(4)
D4H0D0	Helicase domain protein	C0042	144537	377	8(6)	8(6)
D4GYJ0	dhaK	1546	34847	362	5(4)	5(4)
Q48328	**RadA**	0104	37320	355	7(5)	6(4)
D4GRM3	hutU	A0562	67982	349	6(5)	6(5)
D4GUG6	pyrG	2624	60829	335	7(4)	7(4)
D4GUB6	pykA	0806	62042	330	7(5)	6(5)
D4GVJ7	HEAT-PBS family protein	1020	45550	320	6(3)	6(3)
D4GW83	purA	1132	48231	317	7(5)	7(5)
Q48332	atpA	0316	64515	316	6(4)	6(4)
D4GZU0	atpI	0311	79504	315	7(3)	7(3)
D4GR46	hyuB2	A0378	64260	307	5(3)	5(3)
D4GVQ6	GFO family oxidoreductase	2690	40028	299	6(4)	6(4)
D4GSA2	katG	1778	80004	295	5(3)	5(3)
D4GS21	gap2	0481	37656	294	5(2)	5(2)
D4GRT2	cadA	A0624	92544	290	7(4)	6(4)
D4GZ33	aspC1	1604	40193	290	5(4)	5(4)
D4GP72	Pyridoxal phosphate-dependent aminotransferase	B0070	46183	288	5(3)	5(3)
D4GUA4	stomatin family protein	0801	46401	276	6(5)	5(4)

Band e (~100 kDa)

Protein accesion	**Identified proteins**	**HVO number**	**Predicted MW (Da)**	**MASCOT score**	**No. of peptides**	**Peptide sequences**
D4GYK9	**Hel308a**	0014	90355	4774	379(246)	65(50)
D4GZY6/ D4GWR0	Elongation factor 1 α1/2	0359/ 2413	45759/ 45628	2865	174(96)	47(29)
D4GXM0	**RecJ4**	2889	79159	2803	96(72)	38(32)

D4GPB9	Smc-like protein Sph2	B0118	73801	2030	60(40)	31(23)
Q1XBW2	**DnaK**	1590	67250	1898	91(61)	27(24)
D4GWM8	**Cdc48a**	2380	81989	1872	45(29)	35(22)
D4GZ00	**Top6B**	1571	87438	1742	66(36)	28(20)
D4GUG0	Phosphoenolpyruvate carboxylase	2621	102399	1730	43(32)	28(21)
D4GUD0	Phosphoenolpyruvate synthase	0812	83133	1444	31(19)	26(17)
D4GUC3	metG	0809	81936	1430	34(24)	24(18)
D4GTI4	CitB1	1955	70406	1414	52(37)	19(16)
D4GR47	hyuA2	A0379	72968	1407	37(26)	19(17)
D4GT20	**topA**	0681	93686	1406	27(20)	24(18)
O30561	cct1	0133	58925	1310	36(28)	16(14)
D4GY19	dsa1	2960	54912	1268	31(17)	20(11)
D4GSA2	katG	1778	80004	1263	42(25)	20(15)
D4GVH7	acdA	1000	74554	1188	32(21)	20(13)
O30560	cct2	0455	59298	951	22(14)	16(12)
Q9HHA2	cct3	0778	55242	912	16(14)	12(11)
D4GZX6	**rpoA1**	0349	108913	889	17(13)	14(11)
D4GZF1	alaS1	0206	102464	855	16(11)	15(10)
D4GUK7	sufB1	0860	53247	852	21(12)	15(10)
D4GTB5	bccA	2486	65346	839	15(11)	15(11)
D4GUV5	korA	0888	64026	835	19(12)	15(10)
D4GTG8/ D4GSE8	**MutLa/b**	1939/ 0551	59769/ 78729	825	13(10)	11(9)
D4GXF1	porA	1305	68470	820	30(19)	11(9)
D4GZ78	ppk1	1650	93657	809	14(10)	13(9)
D4GU82	Hypothetical	2069	89505	790	11(10)	10(10)
D4GYQ7	dppA1	0062	67307	771	54(31)	11(9)
D4GXK6	DUF814 domain protein	2883	78933	734	15(9)	13(9)
D4GYP9	glyS	0054	64952	707	14(14)	10(10)
D4GZU0	atpI	0311	79504	615	12(9)	12(9)
D4GXK1	**lhr2**	1333	104455	608	13(6)	13(6)
D4GXK9	rpap1	1337	64868	590	9(6)	8(5)
D4GZ02	**gyrA**	1573	94853	584	10(8)	10(8)
Q9V2V6/ Q9V2V5	Protease subunit α1/2	1091/ 2923	27613/ 26727	559	13(10)	10(7)
D4GZP8	thrS	1684	73486	502	10(6)	10(6)
D4GVM4	purL	1048	72964	490	8(5)	8(5)
P25062	csg	2072	85189	476	7(6)	4(4)
D4H037	Putative iron-sulphur protein	1692	80206	447	10(7)	9(6)
D4GW86	Hypothetical	1134	80683	425	9(5)	9(5)
D4GZV3	argS	0324	65034	420	11(6)	9(5)
D4GP53	chllD	B0051	73309	405	7(5)	7(5)
D4GYD4	gnaD	1488	45794	401	10(7)	6(4)
D4GYH4	aglB	1530	113675	400	7(5)	7(5)
D4GUW1	acs1	0894	74351	397	7(7)	6(6)
D4GSD5	Sensor box histidine kinase	1811	94101	372	8(4)	8(4)
D4GWV5	dna2	2767	97589	366	8(3)	8(3)
D4GW05	folCP	1088	89156	347	7(3)	6(3)

D4GW14	YfiH family protein	2708	35406	340	4(3)	4(3)
D4GVF7	nuoCD	0980	63484	335	9(2)	8(2)
D4GVJ5	**RecJ3**	1018	69793	333	7(3)	7(3)
Q48332	atpA	0316	64515	321	7(4)	6(3)
D4GV93	Muc19 precursor	2160	232813	315	4(4)	4(4)
D4GYJ0	dhaK	1546	34847	311	5(3)	4(3)
D4GUM1	mRNA 3-end processing factor-like protein	0874	71947	307	6(4)	6(4)
D4GZX4	**rpoB2**	0347	58897	304	5(3)	5(3)
D4H019	**uvrA**	0393	108678	295	5(3)	5(3)
D4GP52	**cobN**	B0050	141254	294	6(3)	6(3)
Q48328	**RadA**	0104	37320	289	5(2)	5(2)
D4GZG5	**MCM**	0220	78855	276	5(3)	5(3)
D4GUB6	pykA	0806	62042	275	6(3)	5(3)
D4GVZ1	N-6 adenine-specific DNA methylase domain protein	2275	97168	271	6(3)	5(3)
D4GU92	Isocitrate dehydrogenase	2588	45810	269	7(2)	6(2)
D4GT09	aspS	0677	48726	258	5(2)	5(2)
D4GXP5	fumC	2900	50093	248	5(3)	5(3)
D4GRM3	hutU	A0562	67982	220	4(4)	3(3)
D4GRL2	Receiver/sensor box histidine kinase	A0550	95922	219	5(3)	4(3)

Band f (~50 kDa)

Protein accesion	**Identified proteins**	**HVO number**	**Predicted MW (Da)**	**MASCOT score**	**No. of peptides**	**Peptide sequences**
D4GZY6/ D4GWR0	Elongation factor 1 α1/2	0359/ 2413	45759/ 45628	2181	204(132)	30(26)
D4GYK9	**Hel308a**	0014	90355	1851	33(28)	30(26)
D4GS20	pgk	0480	43636	1349	31(22)	20(16)
D4GY38	thrC3	2969	44404	1229	27(20)	17(14)
D4GZV0	arf1	0321	46582	1172	27(17)	20(15)
D4GT62	sucC	2465	41005	1169	53(43)	16(15)
D4GWX0	enolase	2774	42005	1156	39(27)	15(12)
D4GW49	**ribonuclease J**	2724	50115	1147	43(25)	21(13)
D4GV06	FAD-dependent oxidoreductase	2650	50690	1117	32(17)	19(11)
D4GVQ6	GFO family oxidoreductase	2690	40028	1085	46(30)	16(12)
D4GVT3	purB	1084	50817	1042	22(13)	17(12)
Q48327	**ftsZ1**	0717	39805	1025	37(25)	16(12)
D4GU92	Isocitrate dehydrogenase	2588	45810	981	28(19)	16(13)
D4GXM6	GTPase HflX	1346	48212	969	17(14)	16(14)
D4GXG2	glyA1	2862	44349	954	25(18)	14(10)
D4GX38	sdhA	2808	67641	949	22(16)	14(11)
D4GY69	gdhA2	1453	45730	889	18(13)	15(11)
D4GP54	PQQ repeat protein	B0052	44110	871	14(8)	14(8)
D4GVH6	DRTGG domain protein	0999	39826	841	30(22)	13(11)

D4GYV5	sufS	0109	46552	816	15(14)	12(11)
D4GYQ3	dppF1	0058	50184	816	17(8)	15(8)
D4GUK8	sufB2	0861	44685	798	23(17)	12(11)
D4GUC9	PQQ repeat protein	2606	46919	786	18(11)	13(9)
D4GYH5	aglM	1531	45967	786	14(10)	13(9)
D4GPB3	Mandelate racemase	B0111	44272	776	28(12)	15(7)
D4GYK5	tnaA	0009	48543	768	19(7)	15(7)
D4GUJ7	panA	0850	45802	740	18(8)	14(7)
D4GYD4	gnaD	1488	45794	725	22(14)	13(8)
D4GZX7	**rpoA2**	0350	46140	663	18(10)	11(6)
D4GW10	GFO family oxidoreductase	2707	41032	657	13(11)	11(9)
D4GW23	rtcB	2712	53434	657	13(10)	12(9)
D4GVU2	trpD2	2226	39646	656	24(11)	12(8)
D4GVK2	acaB2	1025	40732	654	12(10)	11(10)
D4GW14	YfiH family protein	2708	35406	651	19(15)	10(8)
D4GS07	pchB	0467	43289	650	14(7)	13(7)
D4GYZ9	**top6A**	1570	41829	649	11(10)	9(8)
D4GV77	rbcL	0970	43922	637	13(9)	10(7)
D4GXX3	tsgA11	1401	39212	611	13(12)	7(7)
D4GVT2	gdh	1083	38990	608	17(14)	11(10)
D4GST7	pepB2	1829	39365	601	8(7)	8(7)
D4GYN3	Cytochrome P450	0038	52201	592	15(5)	14(5)
D4GYJ0	dhaK	1546	34847	588	15(8)	9(7)
D4GXP5	fumC	2900	50093	585	12(9)	10(8)
D4GT59	sucD	2464	29864	584	11(9)	8(6)
D4GTE0	**cdc48d**	1907	53410	571	13(7)	11(7)
Q9HHA2	cct3	0778	55242	557	12(6)	10(4)
D4GWT1	Peptidase M42 family protein	2759	37046	556	11(9)	8(7)
D4GZX4	**rpoB2**	0347	58897	555	11(6)	11(6)
D4GW51	idsA1	2725	38205	549	11(9)	9(7)
D4GT99	HD family hydrolase	2481	42772	524	13(6)	11(6)
D4GYK3	lysC	0008	41700	521	9(8)	8(7)
O30561	cct1	0133	58925	521	8(5)	8(5)
D4GU28	Hypothetical	2015	37758	496	8(5)	8(5)
D4GTW8	TIGR00300 family protein	2536	44855	494	13(9)	9(5)
D4GS04	ilvA1	0464	43680	491	12(6)	9(4)
D4GUA4	stomatin family protein	0801	46401	478	9(7)	9(7)
D4GRS7	Hypothetical	A0619	48181	473	10(5)	9(5)
D4GSQ0	Peptidase M24 family protein	2435	39985	471	7(7)	7(7)
D4GW79	Hypthetical	1129	43003	467	9(6)	7(5)
D4GXR9	proA	1370	48117	461	9(8)	7(6)
D4GXW6	M20 family amidohydrolase	1395	45410	460	9(6)	7(6)
D4GYS6	hemL	0081	47811	459	11(5)	8(4)
D4GY39	cysK1	1439	34757	457	9(5)	8(5)
D4GTD4	tif2c	1901	43941	450	8(6)	8(6)
Q59468	hmgA	2583	41085	447	8(7)	7(7)

O30560	cct2	0455	59298	446	7(5)	7(5)
D4GXX4	pmm3	1402	47500	445	8(7)	8(7)
D4GW95	metE2	2743	36800	442	8(5)	7(5)
D4H023	Homolog to glycine cleavage system protein T	0397	39014	441	10(5)	9(5)
D4GZT6	Hypothetical	0307	49060	441	9(5)	9(5)
D4GSV2	ywaD2	1849	47130	439	6(6)	6(6)
D4GRF0	cbiA	A0487	46883	438	9(3)	9(3)
D4GVP4	coaBC	1062	39978	417	7(6)	6(5)
D4GXV5	Putative nucleoside deaminase (Cytosine/guanine deaminase)	2929	47098	415	6(5)	6(5)
D4GXW1	Translation-associated GTPase	2932	43517	412	13(5)	9(4)
Q977V2	srp54	0123	50918	411	9(6)	7(5)
D4GXY7	metB2	2946	43142	394	10(6)	7(4)
D4GQT3	DUF234 domain protein	A0254	52691	394	10(2)	9(2)
P43386	glnA	0239	50522	383	6(3)	6(3)
D4H0E4	**orc8**	C0057	45953	381	9(4)	7(4)
D4GVY2	Hypothetical	2266	56711	377	7(4)	7(4)
D4GW22	lysA	1098	44445	376	7(3)	7(3)
D4GRG1	Hypothetical	A0498	42738	376	8(4)	8(4)
D4GTG3	tif2Ba	1934	43201	373	8(4)	8(4)
D4H069	**orc5**	1725	44549	368	8(5)	7(4)
D4GZ33	aspC1	1604	40193	367	6(6)	5(5)
D4GTB8	asd	2487	36612	348	11(5)	7(3)
D4GZ77	mat	1649	43498	345	9(3)	8(3)
D4GP59	chiH2	B0057	36370	344	11(7)	6(5)
D4GQC4	acd6	A0092	42192	336	8(2)	7(2)
D4GWM8	**Cdc48a**	2380	81989	334	8(4)	8(4)
D4GW90	metE1	2742	40298	331	11(4)	7(3)
D4GTB5	BccA	2486	65346	331	7(3)	7(3)
Q48328	**RadA**	0104	37320	330	7(4)	6(4)
D4GWN2	CBS domain protein	2384	42521	322	6(3)	6(3)
Q1XBW2	**DnaK**	1590	67250	320	7(4)	7(4)
D4GZ82	tyrS	0138	36111	306	9(7)	5(4)
D4GYP4	argG	0049	44483	299	6(4)	6(4)
D4GP72	Pyridoxal phosphate-dependent aminotransferase	B0070	46183	296	5(4)	5(4)
D4GXM0	**RecJ4**	2889	79159	293	6(4)	6(4)
D4GWR6	hmgB	2419	50103	291	4(4)	4(4)
D4GU72	agl12	2059	34141	280	5(2)	5(2)
D4GS17	pepB3	0477	33406	280	4(3)	4(3)
D4GYN8	argD	0043	39767	278	6(3)	5(3)
D4GRM1	hutI	A0560	42409	269	5(5)	5(5)
D4GYK0	NamA family oxidoreductase	0004	39289	267	4(4)	4(4)
D4GS58	cca	0521	50492	264	6(3)	6(3)

D4GYQ7	dppA1	0062	67307	262	5(2)	5(2)
D4GT09	aspS	0677	48726	261	6(2)	6(2)
D4GVS1	**PriS**	2697	42674	260	6(3)	5(3)
D4GXR1	folP2	2908	40813	258	4(3)	4(3)
D4GSG0	malK	0565	41743	257	6(3)	6(3)
D4GYI1	**orc15**	1537	42433	256	5(2)	5(2)
D4GQW0	tsgA5	A0283	49100	254	5(2)	5(2)
D4GS21	gap2	0481	37656	251	4(3)	4(3)
D4GWT0	NAD kinase	1179	36763	247	4(3)	4(3)
D4GZT3	etfA1	0304	33519	239	4(4)	3(3)
D4GQN6	**cas7**	A0207	38184	238	4(3)	4(3)

Protein accession: the UniProt database entry code, e.g. D4GWN0**.**

Predicted MW (Da): The molecular weight in Daltons (Da) of the protein sequence listed as identified by MASCOT.

MASCOT score: the score associated with protein identification.

$$Ions\ score = -10log_{10}(P)$$

P = probability that the observed match is a random event

Individual ions scores > 45 indicate identity or extensive homology ($p<0.05$). Protein scores are derived from ions scores as a non-probabilistic basis for ranking protein hits.

Number of peptides: number of peptides associated with protein identification by MASCOT.

Peptide sequences: the number of distinct peptide sequences associated with the protein identified by MASCOT.

Appendix 2: PIPS Reflective Statement Template

Note to examiners:

This statement is included as an appendix to the thesis in order that the thesis accurately captures the PhD training experienced by the candidate as a BBSRC Doctoral Training Partnership student.

The Professional Internship for PhD Students is a compulsory 3-month placement which must be undertaken by DTP students. It is usually centred on a specific project and must not be related to the PhD project. This reflective statement is designed to capture the skills development which has taken place during the student's placement and the impact on their career plans it has had.

PIPS Reflective Statement

For my PIP placement I decided to explore teaching. My placement was at Suffolk One Sixth-Form centre, Scrivener Dr, Ipswich IP8 3SU.

For this placement I worked predominantly as a teaching assistant in both Biology and Chemistry, predominantly with the year 12 classes but also occasionally assisting year 13 classes. Additionally, I helped the science technicians one day a week in preparing practical lessons for the following week. During this time I also gained experience tutoring one-to-one as a number of students often sought my help when they encountered difficulty with the work. I often planned and lead sections of the lessons, and towards the end of my placement I even took a couple of lessons completely on my own. This enabled me to develop my communication skills and helped build my confidence in this area.

In addition to an increase in communication skills, I improved my skills in leadership and teaching throughout my placement. I also developed my problem solving and explanation skills – for example if a student did not understand something it was up to me to explain it in a different way to ensure that they did. I believe that the placement gave me more confidence in different situations and improved my ability to think on my feet.

I have a passion for learning and therefore, do enjoy teaching. In fact, whilst I finish writing my thesis (September 2019) I have been working as a Science Instructor at Stoke High Academy teaching Biology, Chemistry, and Physics at KS3 and KS4 levels. However this is only temporary, long term I plan to pursue an academic career and therefore may include lecturing at a university level.