Using a tryptophan-regulated promoter system to study chromosomal DNA replication in archaeal model organism *Haloferax volcanii*

Agnieszka Skowyra

A Thesis Submitted to the University of St Andrews for the Degree of Doctor of Philosophy

November 2012

Table of Contents

Table of Contents i

List of Figures v

List of Tables viii

Declaration ix

Abbreviations x

Abstract xiii

Acknowledgements xiv

Chapter 1 Introduction 1

1.1 Archaea as a domain of life and model system 1

1.2 Chromosomal DNA replication in Archaea 6

1.2.1 Replication initiation 9

1.2.1.1 Replication origins and Cdc6/Orc proteins 9

1.2.1.2 MCM helicase 9

1.2.1.3 GINS complex 15

1.2.1.4 Single-stranded DNA-binding proteins 19

1.2.2 Replication elongation 27

1.2.2.1 Primase 27

1.2.2.2 DNA Polymerase 28

1.2.2.3 PCNA, the DNA sliding clamp 30

1.2.2.4 Replication Factor C, the clamp loader 32

1.2.2.5 Topoisomerases 34

1.2.3 Okazaki fragment maturation 35

1.2.3.1 Flap Endonuclease 1 35

1.2.3.2 RNase H 35

1.2.3.3 DNA Ligase 36

1.2.4 Halophiles as a model for archaeal genetics 37

1.3 Working with *Haloferax volcanii* 39

1.4 Transcription initiation and promoter structure in Archaea 42

1.5 Objectives 46

Chapter 2 Materials and Methods 47

2.1 Materials **47**
2.1.1 *Haloferax volcanii* strains 47
2.1.2 *Escherichia coli* strains 50
2.1.3 Other strains 51
2.1.4 Plasmids 51
2.1.5 Chemicals and reagents 55
2.1.5.2 *Haloferax volcanii* growth media 55
2.1.5.3 *Escherichia coli* growth media 56
2.1.5.4 *Haloferax volcanii* buffers and solutions 56
2.1.5.5 Other buffers and solutions 56
2.2 Methods **57**
2.2.1 Manipulation and analysis of nucleic acids 57
2.2.1.1 PCR amplification 57
2.2.1.2 Plasmid DNA extraction 58
2.2.1.3 Restriction digest, dephosphorylation of vector DNA, ligation of DNA 58
2.2.1.4 DNA sequencing and oligonucleotide synthesis 59
2.2.2 *Escherichia coli* microbiology 59
2.2.2.1 Growth of *Escherichia coli* and storage 59
2.2.2.2 *Escherichia coli* transformation 59
2.2.2.3 Generation of unmethylated (*dam*$^-$) plasmid DNA 59
2.2.3 *Haloferax volcanii* microbiology 59
2.2.3.1 Growth of *Haloferax volcanii* and storage 59
2.2.3.2 Extraction of genomic DNA from *Haloferax volcanii* 60
2.2.3.3 Extraction of RNA from *Haloferax volcanii* 60
2.2.3.4 *Haloferax volcanii* transformation 60
2.2.3.5 Mating assay 60
2.2.4 *Haloferax volcanii* phenotyping assays 60
2.2.4.1 Spotting technique 60
2.2.4.2 Growth curves 61
2.2.4.3 Viability curves 61
2.2.4.4 DNA damage sensitivity assays 61
2.2.5 Gene expression analysis by RT-PCR 62
2.2.6 Monitoring DNA synthesis 64
2.2.7 *Haloferax volcanii* reverse genetics 65
2.2.7.1 Construction of conditional mutants of replication genes 65
2.2.7.2 Gene deletion in *Haloferax volcanii* 67
2.2.8 Western blot of Flag epitope-tagged RpaC proteins 68
2.2.9 Site-directed mutagenesis of *tna* promoter 69

2.2.10 Expression of *Haloferax volcanii* RpaC 70
2.2.10.1 Construction of plasmids for the *H.volcanii* RpaC expression 70
2.2.10.2 Protein expression 70
2.2.11 Expression of *Sulfolobus solfataricus* PriS CTD 71
2.2.12 Bioinformatics 71

Chapter 3

Conditional inactivation of replication genes 72

3.1 Introduction 72

3.2 Optimizing conditions to use *tna* promoter system 75

3.3 *tna*-plasmid integrant strains 76

3.3.1 Construction of *tna*-plasmid integrant strains 76
3.3.2 Phenotypic analysis of *tna* plasmid-integrant strains 79
3.3.3 Analysis of gene expression by qRT-PCR 82
3.3.4 Phenotypic analysis of *pcna tna* plasmid-integrant strain 85

3.4 *tna* promoter replacement strains 86

3.5 Improving the *tna* promoter system 89

3.5.1 Construction of truncated forms of the *tna* promoter 90
3.5.2 Placing transcriptional terminator L11 upstream of *tna* promoter 91
3.5.3 Mutating the p*tna* TATA box consensus 93

3.6 Discussion 97

Chapter 4 100

Identification of putative single-stranded DNA-binding proteins in *Haloferax volcanii* 100

4.1 Introduction 100

4.2 *H.volcanii* encodes three putative RPA homologous 100

4.3 Deletion analysis of RPA homologous in *H.volcanii* 104

4.3.1 Construction of plasmids for gene deletion 104
4.3.2 Construction of a single *rpa* deletions 105
4.3.3 Phenotypic analysis of *rpaA* and *rpaB* deletion strains 108
4.3.4 Construction of a double Δ*rpaA* Δ*rpaB* deletion 109

4.4 Discussion 111

Chapter 5 Structure-function analysis of RpaC 115

5.1 Conditional down-regulation of *rpaC* gene expression 115

5.1.1 RpaC involvement in DNA replication 116
5.1.1.1 PolB involvement in DNA replication 118
5.2 Interplay between RPA proteins 120
5.3 DNA damage tests on strain overexpressing of *H.volcanii* RpaC 122
5.3.1 UV sensitivity assay 123
5.3.2 MMS sensitivity test 126
5.3.3 Phleomycin sensitivity test 127
5.4 Cross-species complementation 128
5.4.1 Complementation of *H.volcanii* RpaC by the full-length *H.borinquense* and *H.walsbyi* RpaC orthologues 128
5.4.2 Functional analysis of mutated RpaC proteins 130
5.4.3 DNA sensitivity of ΔNTD and ΔCTD RpaC 135
5.5 Expression of *H.volcanii* full-length RpaC and RpaC NTD 136
5.6 Discussion 140

Chapter 6 144
Characterisation of the PriS-GinS operon in Archaea 144
6.1 Introduction 144
6.2 PriS-GINS operon in Archaea 145
6.3 Separation of the PriS-GINS operon in *H.volcanii* 146
6.4 Bioinformatic analysis of PriS CTD 149
6.5 Expression and purification of PriS CTD 153
6.6 Discussion 156

Chapter 7 General discussion 159

References 164
Appendix 182
Table A1 Oligonucleotides used in this study 182
Table A2 SSB-encoding genes in completely sequences haloarchaeal genomes 186
Table A3 SSB-encoding proteins in fourteen sequenced haloarchaeal genomes 188
Figure A1 Multiple sequence alignment of RpaC proteins 191
Published papers
(Skowyra and MacNeill, 2012)
(Swiatek and MacNeill, 2010)

List of Figures

Figure 1.1 Maximum likelihood phylogenetic tree of the archaeal domain 2
Figure 1.2 Core components of archaeal replisome 8
Figure 1.3 Domain organization of the monomeric MCM helicases 12
Figure 1.4 Two possible DNA unwinding modes by MCM helicase 14
Figure 1.5 Schematic representation of the domain organization of eukaryotic and archaeal GINS proteins 16
Figure 1.6 The crystal structure of *T.kodakaraensis* GINS 18
Figure 1.7 Recent model of assembly of archaeal replisome 19
Figure 1.8 Structure of OB fold domain and OB fold/nucleic acid complex 20
Figure 1.9 Structures of eukaryotic RPA protein bound to DNA 22
Figure 1.10 Diversity of domain architecture of single-stranded DNA-binding proteins in three domains of life 25
Figure 1.11 Structure of archaeal PCNA proteins 32
Figure 1.12 The clamp loading process 33
Figure 1.13 DNA topology and its relevance in replication 34
Figure 1.14 Different conformations of archaeal ATP-dependent ligases 37
Figure 1.15 Properties of *H.volcanii* 40
Figure 1.16 Comparison of the transcription initiation complex in Eukarya and Archaea 43
Figure1.17 The structure of archaeal promoter region 44
Figure 1.18 Overall structure of RNA polymerases from bacteria, archaea and eukaryotes 45
Figure 2.1 Construction of plasmids for gene deletion 67
Figure 2.2 Pop-in/pop-out gene deletion strategy 68
Figure 2.3 Site-directed mutagenesis using PCR overlap extension 69
Figure 3.1 Growth of *H.volcanii* DS70 on HvMin medium supplemented with different concentrations of tryptophan 75
Figure 3.2 Construction of pNPM01-series plasmid for regulated expression 77
Figure 3.3 Construction of *tna* plasmid-integrant strains 78
Figure 3.4 Growth of the *tna* plasmid-integrant strains in the presence and absence of tryptophan 80
Figure 3.5 Growth curves of selected *tna* plasmid-integrant strains 81
Figure 3.6 Analysis of transcript level of selected *tna* plasmid-integrant

strains 84
Figure 3.7 Growth of the *tna pcnA* plasmid-integrant strains in the presence of different tryptophan concentrations 85
Figure 3.8 Construction of pNPM-Bcl series plasmids for regulated expression 86
Figure 3.9 Construction of *tna* promoter replacement strains 87
Figure 3.10 Growth of the *tna* promoter replacement strains in the presence and absence of tryptophan 89
Figure 3.11 Construction of *tna*2-5 truncated promoters 91
Figure 3.12 Growth of the strains with truncated *tna* promoters in the presence and absence of tryptophan 92
Figure 3.13 Location of the TATA box element in *tna* promoter 94
Figure 3.14 Mutations generated in *tna* promoter TATA box element 94
Figure 3.15 Growth of the cells expressing PolB, PolD1 and Mcm from mutated *tna* promoters in the presence and absence of tryptophan 95
Figure 3.16 Growth of the cells expressing PolD2 and RpaC from mutated *tna* promoters in the presence and absence of tryptophan 96
Figure 3.17 Comparison analysis of transcript level of selected genes expressed from the wt *tna*- and *tna*M promoter 97
Figure 4.1 RPA proteins in *H.volcanii* and *M.acetivorans* 102
Figure 4.2 Construction of plasmids for *rpa* gene deletions 105
Figure 4.3 Schematic diagram of pop-in/pop-out gene deletion strategy exemplified for *rpaA* 106
Figure 4.4 Screening of *rpa* gene deletions 108
Figure 4.5 Response to UV treatment of *rpaA* and *rpaB* deletion strains 109
Figure 4.6 Screening of Δ*rpaA* Δ*rpaB* double deletion 110
Figure 4.7 Growth of the Δ*rpaA* strain carrying integrated pNPM-tna-RpaB 111
Figure 5.1 Growth of the p*tna-rpaC* conditional lethal mutants 115
Figure 5.2 Monitoring of DNA synthesis in p*tna-rpaC* and p*tna*M3-*rpaC* strains 117
Figure 5.3 Monitoring of DNA synthesis in p*tna-polB* and p*tna*M3-*polB* strains 119
Figure 5.4 Rescue of p*tna-rpaC* strain by elevated expression of RpaB 121
Figure 5.5 Strains overexpressing RpaC used in DNA sensitivity assays 123
Figure 5.6 UV response of p*tna-rpaC* and p*fdx-rpaC* strains 124
Figure 5.7 4-NQO response of p*tna-rpaC* and p*fdx-rpaC* strains 126
Figure 5.8 MMS response of p*tna-rpaC* and p*fdx-rpaC* strains 127
Figure 5.9 Phleomycin response of p*tna-rpaC* and p*fdx-rpaC* strains 128

Figure 5.10 Rescue of ptna-rpaC strain by *H.borinquense* and *H.walsbyi* RpaC orthologues 130
Figure 5.11 Complementation of *H.volcanii* RpaC by mutated *H.borinquense* orthologues 133
Figure 5.12 Sensitivity of cells expressing RpaC-ΔNTD and RpaC-ΔCTD to DNA damage 136
Figure 5.13 Sequence and predicted structure of RpaC N-terminal domain (NTD) 137
Figure 5.14 A small-scale expression of CBD, RpaC NTD-CBD and FL RpaC-CBD proteins 138
Figure 5.15 Large-scale purification of CBD, RpaC NTD-CBD and FL RpaC-CBD proteins 140
Figure 6.1 Growth of the p*tna-priS-ginS* conditional lethal mutants 146
Figure 6.2 Rescue of p*tna-priS-ginS* strain by *in trans* expression *H.volcanii* PriS-GINS and GINS proteins 147
Figure 6.3 Rescue of p*tna-priS-ginS* strain by *H.borinquense* PriS and GINS orthologues 149
Figure 6.4 Multiple sequence alignment of archaeal primase CTD, archaeal GINS B-domain and eukaryotic GINS B-domain 151
Figure 6.5 Comparison of the three-dimensional structures of *S.solfataricus* PriS-CTD with the B-domain of human GINS subunits 152
Figure 6.6 Model for acquisition of the CTD by PriS 153
Figure 6.7 Expression of *S.solfataricus* PriS CTD in *E.coli* 155

List of Tables

Table 1.1. Model organisms for genetics in the domain Archaea 5

Table 2.1 *H.volcanii* strains used in this study 47

Table 2.2 *E.coli* strains used in this study 50

Table 2.3 Plasmids used in this study 51

Table 2.4 Supplements added to different type of media 55

Table 2.5 *H.volcanii* media supplements 56

Table 2.6 Reagents for PCR amplification using Long PCR Enzyme Mix 57

Table 2.7 Reagents for PCR amplification using MyTaq Red Mix 57

Table 2.8 PCR reaction conditions 58

Table 2.9 Final concentrations for DNA damaging agents used is sensitivity assays 62

Table 2.10 SYBR Green Supermix reagents for RT-PCR 63

Table 2.11 One-Step RT-PCR with SYBR Green reagents for reverse transcriptase RT-PCR 63

Table 2.12 Reverse transcriptase RT-PCR reactions 64

Table 2.13 Details about construction of pNPM-tna-derived plasmids 66

Table 3.1 Selected components of *H.volcanii* chromosomal DNA replication machinery 73

Table 3.2 Experimental design for gene expression analysis 82

Table 3.3 Mutant tna promoter constructs 90

Table 4.1 Number of colonies obtained in gene knockout experiments 107

Table 5.1 Mutated derivatives of *H.borinquense* RpaC 131

Table 6.1 Condition tested for expression of PriS CTD 154

Declaration

I, Agnieszka Skowyra, hereby certify that this thesis, which is approximately 60,000 words in length, has been written by me, that it is the record of work carried out by me and that it has not been submitted in any previous application for a higher degree.

I was admitted as a research student in April 2009 and as a candidate for the degree of PhD in April 2010; the higher study for which this is a record was carried out in the University of St Andrews between 2009 and 2012.

Date 12.11.2012 Signature of candidate

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

Date 12.11.2012 Signature of supervisor

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

The following is an agreed request by candidate and supervisor regarding the electronic publication of this thesis:

Access to all of printed copy but embargo of all of electronic publication of thesis for a period of two years on the following ground(s): publication would be commercially damaging to the researcher, or to the supervisor, or the University; publication would preclude future publication;

Date 12.11.2012 Signature of candidate

Date 12.11.2012 Signature of supervisor

Abbreviations

3’	3 prime DNA end
^{3}H	Tritium
4NQO	4-Nitroquinoline 1-oxide
5’	5 prime DNA end
5FOA	5-fluoroorotic acid
6-4 PPs	6-4 Photoproducts
ADP	Adenosine diphosphate
APS	Ammonium persulphate
ATP	Adenosine 5’-triphosphate
BER	Base excision repair
BLAST	Basic local alignment search tool
bp	Base pair
BRE	B-factor recognition element
CBD	Cellulose binding domain
CFU	Colony factor unit
COG	Cluster of orthologous groups
CTD	C-terminal domain
DBD	DNA binding domain
DEPC	Diethylpyrocarbonate
DMSO	Dimethyl sulfoxide
DNA	Deoxyribose nucleic acid
DSB	Double strand break
dsDNA	Double-stranded DNA
E.coli	*Escherichia coli*
EDTA	Ethylenediaminetetraacetic acid
Fen1	Flap endonuclease 1
FRET	Fluorescence resonance energy transfer
GINS	Go, Ichi, Ni, San
GST	Glutathione-S-transferase

Hbo	*Halogeometricum borinquense*
Hfx	*Haloferax volcanii*
HR	Homologous recombination
Hv-Ca	*Haloferax volcanii* casamino acids media
HV-Min	*Haloferax volcanii* minimal media
Hv-YPC	*Haloferax volcanii* yeast extract, peptone and casamino acids media
Hwa	*Haloquadratum walsbyi*
IPTG	Isopropyl β-D-1-thiogalactopyranoside
kDa	Kilo Dalton
LigA	ATP-dependent ligase
LigN	NAD^+-dependent ligase
Mac	*Methanosarcina acetivorans*
MALDI-MS	Matrix-assisted laser desorption/ionization mass spectrometry
Mb	Mega base
MCM	Minichromosome maintenance
MMS	Methyl methanesulfonate
NER	Nucleotide excision repair
nt	Nucleotide
OB	Oligosaccharide-oligonucleotide
OD	Optical density
ORC	Origin recognition complex
ORF	Open reading frame
PBS	Phosphate buffered saline
PCNA	Proliferating cell nuclear antigen
PCR	Polymerase chain reaction
PDB	Protein data bank
PEG	Polyethylene glycol
P*fdx*	Ferredoxin promoter
Pfu	*Pyrococcus furiosus*
PIP box	PCNA interacting protein box
PMSF	Phenylmethylsulfonyl fluoride
Pol	Polymerase

PriL	Primase large subunit
PriS	Primase small subunit
p*tna*	*Tna* promoter
RT PCR	Reverse transcriptase polymerase chain reaction
RFC	Replication factor C
RNA	Ribonucleic acid
RNAP	RNA polymerase
RPA	Replication protein A
RPE	Rpa-associated phosphoesterase
Rpm	Revolutions per minute
SDS PAGE	Sodium dodecyl sulphate-polyacrylamide gel electrophoresis
SSB	Single-stranded DNA binding protein
TBP	TATA-box binding protein
TF	Transcriptional factor
Tm	Melting temperature
tRNA	Transfer RNA
Trp	Tryptophan
Ura	Uracil
UV	Ultraviolet light
v/v	Volume per volume
w/v	Weight per volume
wt	Wild-type

Abstract

Chromosomal DNA replication is an essential process for all forms of cellular life. Archaea, the third domain of life, possess a DNA replication apparatus that shares significant protein sequence similarity with eukaryotic replication factors, making the Archaea a good model system for understanding the biology of chromosome replication in Eukaryotes.

The results described in this thesis contribute to the genetic and physiological characterisation of DNA replication in the model organism *Haloferax volcanii*, a halophilic euryarchaeon. The thesis documents the generation of conditional lethal mutants of replication genes in *H.volcanii* using the tryptophan- regulated *tna* promoter and its TATA-box mutant, *tna*M3. This system was used to study the cellular function of the triple OB fold containing single-stranded DNA binding protein RpaC and for characterisation of PriS-GINS operon.

Deletion analysis of three putative SSB proteins in *H.volcanii* indicated that RpaA1 and RpaB1 are individually non-essential for cell viability but share an essential function, whereas RpaC protein is essential. Loss of RpaC function can however be rescued by elevated expression of RpaB, indicative of functional overlap between the two classes of haloarchaeal SSB. Down-regulation of RpaC caused growth retardation and a significant reduction of DNA synthesis *in vivo* suggesting, that RpaC is required for DNA replication. In addition, RpaC overexpression increased resistance to various types of DNA damage implying its role in DNA repair. This function is probably mediated by the N-terminus as deletion of this region makes cells sensitive to DNA damaging agents.

Analyses of PriS-GINS operon indicated that GINS could be down-regulated without severe consequences for the cells suggesting that it is non-essential protein in *H.volcanii*. In addition, bioinformatics studies identified sequence similarity between the C-terminal domain of the catalytic subunit of archaeal primase (PriS CTD) and B-domain of GINS51 and B-domain of GINS23 that offer insights into mechanisms for the evolution of these proteins.

Acknowledgements

This research project would not have been possible without the support of the kind people around me, only some of whom it is possible to give particular mention here.

Above all, I would like to express my deep gratitude to my supervisor, Dr Stuart MacNeill, for being my mentor and friend, for his never-ending patience, enthusiasm and guidance. You made me believe in my own abilities, which was the first and crucial step to becoming a scientist. I could not have imagined having a better supervisor for my PhD study.

During my project I have been lucky to work with a fantastic group of people who were not only my labmates but also best friends. I thank all members (past and present) of the MacNeill lab, especially Jasmin Schnick, Dr Xavier Giroux, Elizabeth King, Kazim Ogmen, Antonia Evripioti, Fiona Gray, Anna Bota Cherino, Kaori Kashi and Kacper Sendra for all scientific discussions, encouragements, making me laugh when I really needed it and our long talks about life. Special thanks to Elizabeth King, for helping me construct the *tna*M3 promoter. I also thank all members of the Malcolm White, Peter Coote and Jo Parish lab for their support in all its forms and friendly atmosphere in the lab.

Many thanks go to the members of the mass spectrometry team Dr Catherine Botting, Dr Sally Shirran and Dr Matt Fuszard who helped me analysed protein samples.

I would like to also acknowledge the administrative and technical staff of the Biomedical Sciences Research Complex, especially Ian, Charlie, Allan and Rey. Thank you for your dedicated work, smiling faces and helping me every time I needed you, even though sometimes I am not the easiest person to work with.

Finally, my warmest thanks go to my family, my parents in law and my husband for their unconditional love, constant encouragement and listening patiently to “cells didn’t grow”. You are the reasons for everything I have and ever will do.

Chapter 1
Introduction

1.1 Archaea as a domain of life and model system

Before the 1970s, only two domains of life- Bacteria and Eukarya- were recognised, and Archaea were wrongly classified as bacteria. The distinct status of Archaea was revealed when the American microbiologist Carl Woese employed the technology of nucleic acid sequencing and used small-subunit rRNA (SSU rRNA) as a molecular chromometer; rRNA, as an essential component of translational apparatus of all self-replicating organisms, shows strong sequence conservation. Woese suggested that prokaryotes are much more diverse than previously thought and that the existing division of living organisms should be reconsidered (Woese and Fox, 1977). In the redrawn phylogenetic tree Archaea are designated as the third domain of life and their closer relationship with Eukaryotes than Bacteria is indicated (Woese and Fox, 1977, Woese et al., 1990). The status of the Archaea as a domain is generally accepted but the phylogenetic relationship between species within the domain remains open to debate. Until recently, the archaeal domain was divided into two phyla, the Euryarchaeota (from the Greek '*euryos*', meaning diversity) and the Crenarchaeota (from the Greek '*crenos*', meaning spring or origin). The Euryarchaeota is the largest group, including halophiles and methanogens, whereas the Crenarchaeota include only hyperthermophiles (Woese et al., 1990). The unclear phylogenetic position of *Cenarchaeum symbiosum*, a marine organism living as an endosymbiont of sponge *Axinella mexicana* (Preston et al., 1996) showed that using only small subunit rRNA sequences is not enough to resolve the deepest nodes of the archaeal phylogeny. That problem is caused by the small size of SSU rRNA, which limits the number of nucleotide positions that can be used in analysis. A combined analysis of SSU rRNA, large subunit rRNA (LSU) and ribosomal proteins sequences gives more reliable results in phylogenetic analysis. Recently, the two new main phyla, Thaumarchaeota and Aigarchaeota, were proposed (Brochier-Armanet et al., 2011) (Figure 1.1) Thaumarchaeota (from the Greek '*thaumas'*, meaning wonder) are believed to be an ancient lineage of Archaea.

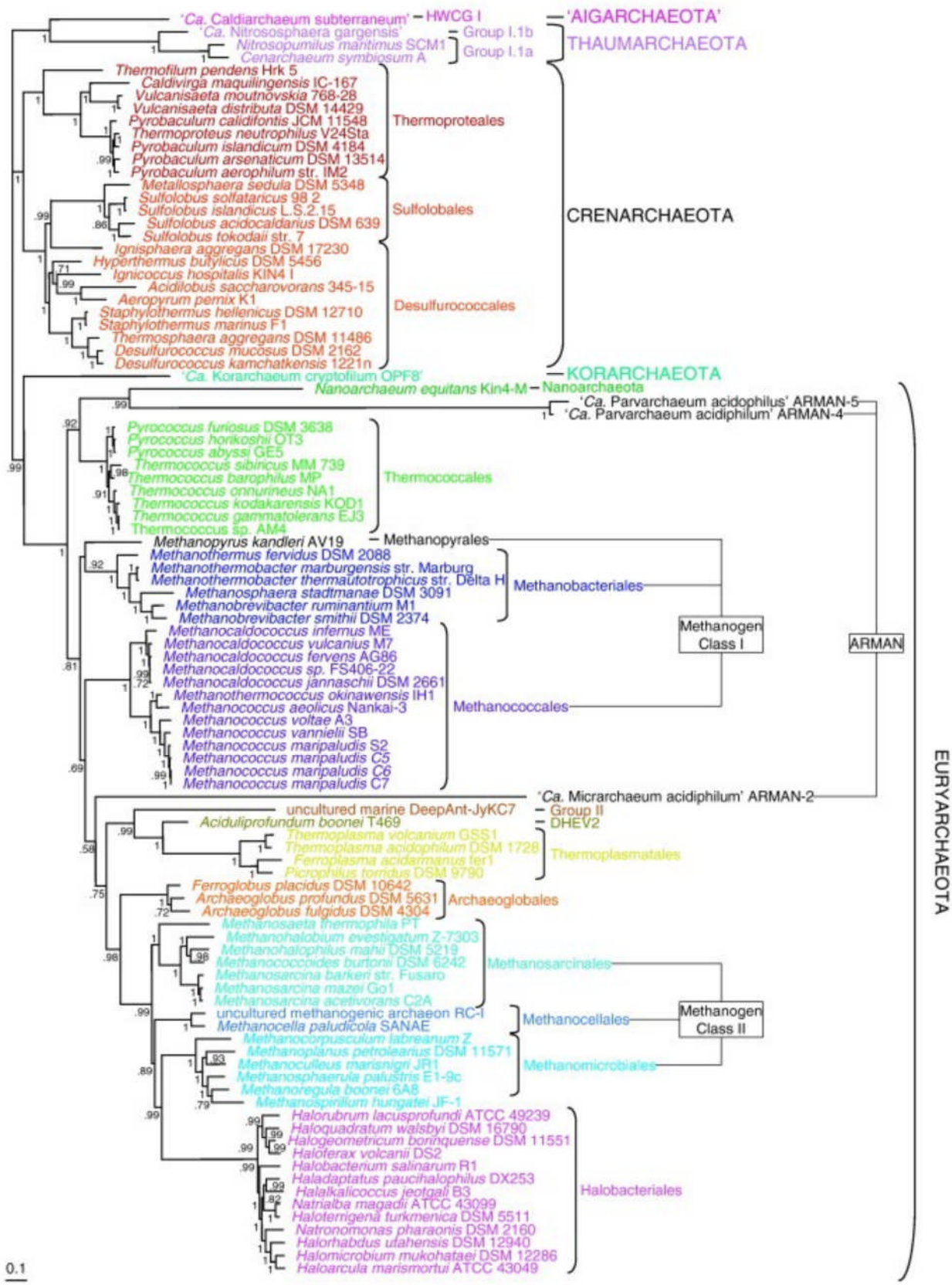


Figure 1.1 Maximum likelihood phylogenetic tree of the archaeal domain based on a concatenation of 57 ribosomal proteins, adapted from (Brochier-Armanet et al., 2011).

The most characteristic feature of Archaea is their ability to survive in extreme environments not available for most of Bacteria and Eukaryote. They are well suited to conditions that might have existed on the early Earth (Archaean), which was the reason that they were given their name. Extremophilic Archaea are members of four main physiological groups: the halophiles, thermophiles, alkaliphiles, and acidophiles. The first group, comprising species like *Haloferax volcanii* or *Halobacterium salinarum*, requires molar salt concentrations so can be found in salt lakes like the Dead Sea. Thermophiles, including *Sulfolobus solfataricus,* grow best at temperatures above 45^{o}C in places such as hot springs. The Hyperthermophile *Methanopyrus kandleri* grows at 122^{o}C, which is the highest recorded temperature at which any organism will grow. Alkaliphiles requires a high pH for life, usually higher than 10. In contrast to alkaliphiles, the last mentioned group, the acidophiles, need a pH lower than 3 for growth. An example of species belonging to that group is *Picrophilus torridus*. It grows at pH 0, which is equivalent to thriving in 1.2 molar sulphuric acid.

Archaea are a chimera of bacterial, eukaryotic and unique features. They resemble bacteria in general cell morphology and physiology, including pathways involved in energy production, nitrogen fixation and polysaccharide synthesis (Schafer et al., 1999). In contrast, the principal components of the information processing system, including DNA replication and translation are much more similar to eukaryotic systems (Allers and Mevarech, 2005). Archaea have also some unique features, not present in either of other two domains, like the ability to survive in extreme environments, ether-linked membrane lipids and unique metabolic pathways (Ferry, 2010, Sakuraba et al., 2004, van de Vossenberg et al., 1998). In fact, around 40-50% of archaeal genes have no apparent homologous in Bacteria or Eukarya (Allers and Mevarech, 2005).

Although Archaea are well known from being extermophiles, cultivation-independent studies showed archaeal abundance in all known habitats. They are widespread, may contribute up to 20% of the total biomass on Earth and are a vital player in biochemical cycles essential for the proper function of ecosystems (Schleper et al., 2005).

Some archaeal proteins had been studied even before the third domain of life was proposed. For example, bacteriorhodopsin, the major photosyntheic protein, was discovered in *Halobacterium salinarum* in 1971 (Oesterhelt and Stoeckenius, 1971). Enzymes from extermophilic species are tested for their industrial relevance and

there are many examples of success in that field. Thermostable DNA polymerase (Pfu DNA polymerase) from *Pyrococcus furiosus* is commonly used for molecular cloning because of its proofreading properties. Amylases, galactosidases and pullulanases from other *Pyrococcus* species are used in food processing (Egorova and Antranikian, 2005). Thermo- and halophilic proteins are also very suitable for structural studies. The structure of the ribosome was solved using the complex from *Haloarcula marismortui* (Ban et al., 2000).

The ecological significance of Archaea are also studied by many research groups. It was recently descoverd that ammonia-oxidising archaea, which are among the most abundant organisms on this planet, have a great role in the global cycling of nitrogen (Schleper and Nicol, 2010).

Research on archaeal cell biology and genetics was for a long time far behind that of Bacteria or Eukarya. One of the reasons was a lack of the basic genetic techniques and Extremophiles, in general, were believed to be difficult to culture. In fact, many archaeal species can be cultivated relatively easy and there has been tremendous progress in the development of genetic tools for the major archael phyla. Currently, four groups of Archaea are particularly well suited for genetic studies: methanogens, halophiles, *Thermococcales* (thermophilic euryarchaea) and crenarchaea (*Sulfolobales*) but all four have both advantages and disadvantages. Table 1.1 summarises key species in all four groups and examples of research that these species can be used for.

Table 1.1. Model organisms for genetics in the domain Archaea

Group of Archaea	Key species		Scientific interest/features
Methanogens	*Methanococcus*	*Methanococcus maripaludis* *Methanocaldococcus janaschii*	▪ The methanogenic pathway (De Vrieze et al., 2012) ▪ DNA Replication and transcription (Walters and Chong, 2009) ▪ Structural biology (Thomas and Cavicchioli, 2000) Methanogens are strict anaerobes that requires special measures in the lab
	Methanosarcina	*Methanosarcina acetivorans* *Methanosarcina barkeri*	
Halophiles	*Haloferax volcanii* *Halobacterium salinarum*		▪ Adaptation to saline environments (Mevarech et al., 2000, Morgunova et al., 2009) ▪ DNA replication (Capes et al., 2011, MacNeill, 2009, Norais et al., 2007a) ▪ DNA repair (Crowley et al., 2006, Lestini et al., 2010) ▪ Structural biology and biotechnology (Ban et al., 2000, Oren, 2010) ▪ Gene exchange and LGT (Papke et al., 2007, Papke et al., 2004)
Thermococcales	*Thermococcus*	*Thermococcus kodakaraensis*	▪ Protein thermostability (Daniel et al., 2010) ▪ DNA replication and repair (Hopkins and Paull, 2008, Yoshimochi et al., 2008b) ▪ Cellular response to stress (Laksanalamai and Robb, 2004, Neves et al., 2005) Optimal growth at high temperature coupled with anaerobic growth requires special measures in the lab
	Pyrococcus	*Pyrococcus furiosus* *Pyrococcus abysi*	
Sulfolobales	*Sulfolobus solfataricus* *Sulfolobus islandicus* *Sulfolobus acidocaldarius*		▪ DNA replication and repair (Paytubi et al., 2012, Gristwood et al., 2012) ▪ Structural biology (Hirata et al., 2008, Liu et al., 2008a) ▪ CRISPR system(Zhang et al., 2012, Lawrence and White, 2011) ▪ Systems biology (Albers et al., 2009) Genetic methodology developed only recently

1.2 Chromosomal DNA replication in Archaea

DNA replication is an essential process for all forms of cellular life. It requires the complex interplay of many enzymes and other protein factors. Only the temporal and spatial coordination of all participants can ensure highly efficient and accurate chromosome replication, which is crucial to maintain the integrity of the genetic information. The general course of DNA replication is similar in all three domains of life. DNA replication starts at specific sequences called origins of replication, at which origin-binding proteins bind and locally unwind the DNA duplex. Additional proteins are loaded and a helicase assembles around DNA to form the pre-initiation complex. The helicase unwinds DNA, resulting in formation of the replication fork that extends away from the origin. Single-stranded DNA (ssDNA), exposed as a result of helicase activity, is coated and protected by ssDNA binding protein. The movement of the replication fork and bidirectional DNA synthesis involves the action of primase and DNA polymerase. The latter is tethered to the DNA template by the ring-shaped sliding clamp. Due to the antiparallel nature of DNA and the fact that DNA polymerases can add free nucleotides only to the 3' end of the newly forming strand, one strand of the chromosome (called the leading strand) is synthesised continuously whereas the second strand (the lagging strand), is copied discontinuously as a series of Okazaki fragments that are later joined.

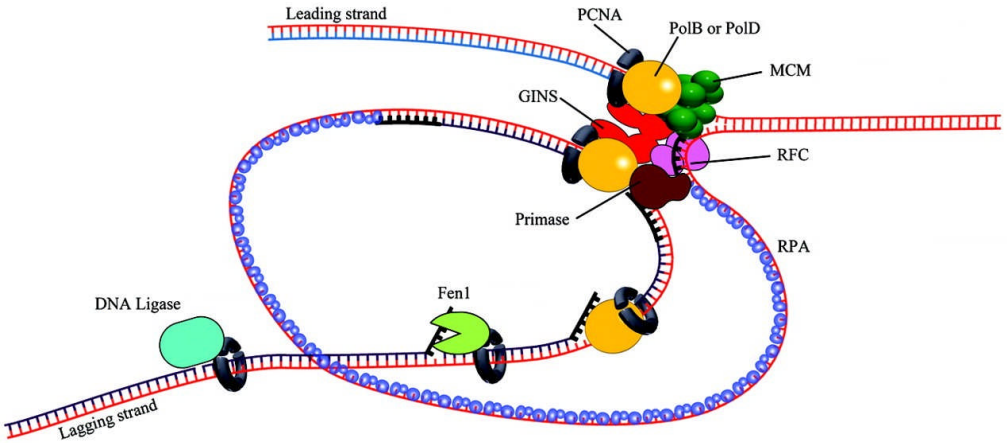
Although the basic principles of DNA replication are very similar in all domains of life, comparative genome and structural analysis indicated that eukaryotic and archaeal replication machineries are much more similar to each other than to the bacterial machinery. Several putative archaeal replication proteins display significant sequence similarity with eukaryotic proteins that have no apparent counterpart in Bacteria. Furthermore, in those cases where related replication proteins are present in all three domains, the archaeal and eukaryotic homologous are more similar (Grabowski and Kelman, 2003, MacNeill, 2001, Barry and Bell, 2006, MacNeill, 2009).

Active research on archaeal DNA replication started in about 1990, far behind similar studies in Bacteria (studied before 1960) or Eukarya (studied before 1980). Why is it worthwhile to study DNA replication in Archaea? Similarities between archaeal and eukaryotic replication proteins make Archaea a good model system for understanding this process in eukaryotic cells. Based on predicted number of encoded proteins, the archaeal DNA replication apparatus is simplified version of

that in eukaryotes. Often, heteromeric eukaryotic factors are homomeric, or at least simpler in structure, in archaea. Also, many archaeal proteins show enhanced stability, which is a useful feature for structural studies. Close similarities between archaeal and eukaryotic replication machinery together with the fact that Archaea possess bacterial-like circular genome, encourage questions of the history of cellular life. Studying DNA replication in Archaea offers an insight into this process from an evolutionary perspective. Finally, as many Archaeal species are extremophiles, working on DNA replication is especially interesting in order to understand the mechanisms by which cells live and maintain their genomic integrity in extreme environmental conditions.

In Archaea, approximately 20-25 proteins have been identified that are likely to be involved in chromosome replication, 10-15 of which can be considered as core components of the replication apparatus (see Figure 1.2 for a schematic diagram of archaeal replication core and list of proteins involved in replication process) (Barry and Bell, 2006) The following section will summarise current knowledge of the structure and function of molecular components of archaeal replication apparatus, considering similarities and differences with their eukaryotic homologous.

A



B

Protein	Structure/Function
PolB	Monomeric family B DNA polymerase, related to catalytic subunit of eukaryotic replicative polymerase
PolD	Dimeric family D DNA polymerase, unique to Euryarchaea. Unknown function
MCM	Catalytic core of the replicative helicase. Heterohexameric in eukaryotes, homohexametic in most archaea
GINS	In eukaryotes heterotertameric part of the CMG complex. In Archaea dimer of dimers in some archaea, homotetramer in others. Exact function in archaeal cells unknown
Primase	Dimeric primase consist of catalytic small subunit and non-catalytic large subunit
RPA	Single stranded-DNA binding protein, various configurations in archaea
PCNA	Sliding clamp that tether DNA polymerase into DNA. Heterotrimer in eukaryotes, homo- or heterotrimer in archaea
RFC	PCNA loader. Heteropentamer in eukaryotes, various configurations in archaea, but probably all pentameric
Fen1	Monomeric endonuclease involved in Okazaki fragment maturation
DNA ligase	DNA ligase involved in Okazaki fragment joining

Figure 1.2 Core components of archaeal replisome

A Cartoon showing the overall structure of replisome indicating position of proteins involved in DNA replication, adapted from (Li et al., 2010)
B Main protein factors involved in DNA replication

1.2.1 Replication initiation

1.2.1.1 Replication origins and Cdc6/Orc proteins

To accommodate the large size of the genome, eukaryotic cells initiate DNA replication at hundreds, if not thousands, of origins. Origins are rich in adenine and thymine bases, including one or more AT-stretches. Long inverted repeat sequences are located at the both ends of the origins, with several shorter repeats in between (Sun and Kong, 2010). Origins are subsequently bound by many proteins, including ORC (origin-recognition complex), Cdc6, Cdt1 and MCM (minichromosome maintenance) helicase. Although the first origin of replication identified in archaea was the single origin from *Pyrococcus abyssi* (Myllykallio et al., 2000), other archaeal species contain multiple origins: *S.solfataricus* and *S.acidocaldarius* have three origins *oriC1*, *oriC2* and *oriC3* (Lundgren et al., 2004), *Hbt.salinarum* two (Zhang and Zhang, 2003) and *H.volcanii* five (Norais et al., 2007b). In *H.volcanii*, the origins are distributed on the different replicons; two are present on the main chromosome. Interestingly, origins show a functional hierarchy or differential usage - one of the origins from smaller chromosome appears dominant. This differential usage of the origins might reflect replication control under certain growth condition (Norais et al., 2007a). The most archaeal genomes encode at least one homologue of Cdc6/Orc1; *H.volcanii* for example has 14 homologous, from which at least three (*orc1*, *orc5* and *orc10*) can be deleted (Norais et al., 2007a). Like their eukaryotic counterparts, archaeal Cdc6/Orc1 homologous are members of the AAA+ family of ATPases (ATPases Associated with diverse cellular Activities). The three-dimentional structure of the *Pyrobaculum aerophilum* Cdc6 shows that protein is composed of three distinct domains. Two domains form an AAA^{+}-type nucleotide-binding fold. The two domains form a cleft, which is a place of Mg-ATP binding and hydrolysis. The third domain comprises of three α-helices and three β-sheets that folds into a winged helix, a structure found in a many proteins that bind DNA (Liu et al., 2000).

1.2.1.2 MCM helicase

In eukaryotic and archaeal cells the “core” of replicative helicase is MCM (minichromosome maintenance) complex. The name reflects the first identification of MCM proteins in budding yeast, in a screen of genes whose mutation abolished the ability of the cells to maintain a plasmid containing a centromere and a

replication origin (minichromosome) (Maine et al., 1984). Eukaryotic MCM is a heterohexamer, composed of six polypeptides, MCM2-7, all of which are essential for cell viability (Forsburg, 2004, Bell and Dutta, 2002). That heterohexameric complex further forms a complex with Cdc45 and GINS, called the CMG complex, which is believed to function as replicative helicase *in vivo* (see below). The archaeal MCM helicase exists as a homohexamer of a single polypeptide that is homologous to each of the MCM2-7 subunits in eukaryotes. Within Archaea, the best studied MCM proteins comes from the species *M.thermautotrophicus* (mthMCM) and *S.solfataricus* (ssoMCM). Those proteins serve as a simplified model for understanding the structure and function of replicative helicases. Both proteins oligomerize to form single or double hexamers in solution and both exhibit single- and double-stranded DNA binding, ATPase activity and 3'-5' helicase activity on forked DNA substrates *in vitro* (Slaymaker and Chen, 2012).

As mentioned above, most Archaeal species encode a single homologue of MCM. Atypically, the *Thermococcus kodakaraensis* genome encodes three MCM homologous, designated MCM1-3 (Pan et al., 2011). All three proteins share the sequence similarity with other archaeal and eukaryotic counterparts, but MCM1 and MCM3 display unique extensions at their N-termini. Biochemical studies using recombinant proteins indicated that MCM2 and MCM3 assemble into homohexamers and exhibit DNA binding, helicase and ATPase activities *in vitro*. MCM1 and MCM2, however, can be deleted in *T.kodakaraensis* and deletion strains show no obvious defect in viability or growth rate. Therefore, like other archaeal organisms, *T.kodakaraensis* possesses only one essential MCM homologue, MCM3, and that protein is probably the replicative helicase in this organism. Additional studies, like testing the sensitivities of Δ*mcm1* and Δ*mcm2* mutant strains to different types of DNA damage, are required to determine physiological the function of non-essential MCMs (Pan et al., 2011).

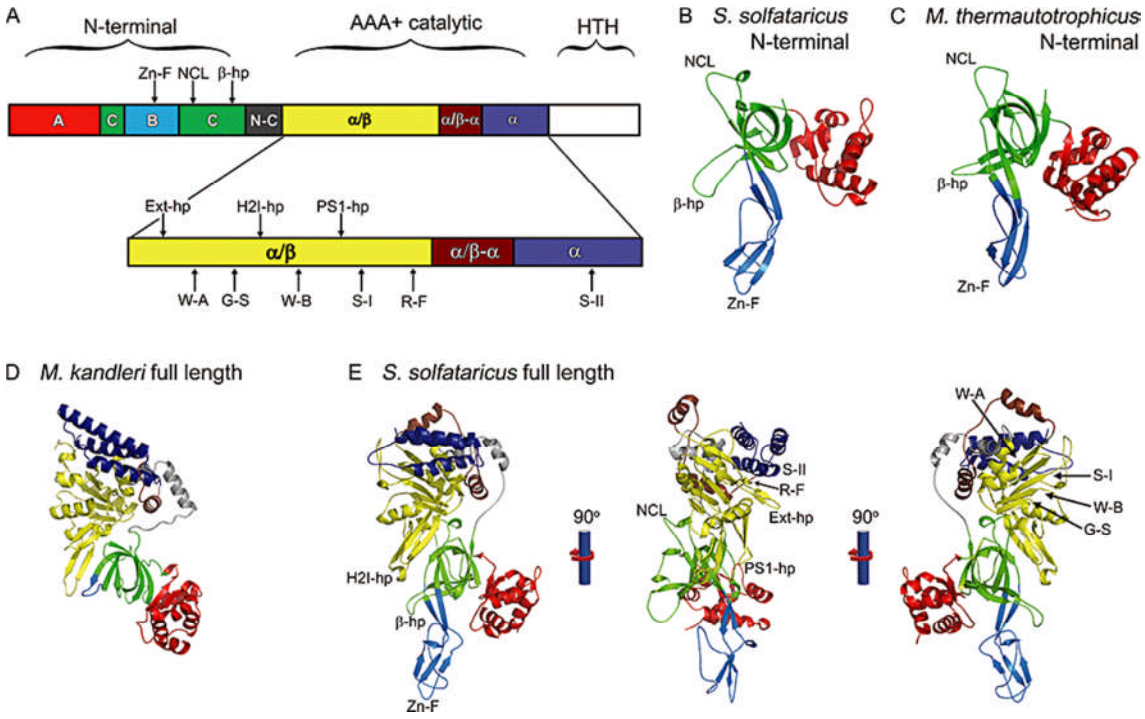
Thermococcus kodakaraensis is not the only archaeal organism with more than one *mcm* gene. Analyzing the genomic context of those genes showed that often they are located within mobile elements, most likely originated from viruses or plasmids. Some element-associated MCMs are structurally distinct from their cellular counterpart, with one case of novel domain organization (Krupovic et al., 2010). The number of MCM homologous is especially high in *Methanococcales* and varies from two in *M.voltae* A3 to eight in *M.maripaludis* C6 (Walters and Chong, 2010). Co-expression of recombinant MCMs from *M.maripaludis* C2 allowed co-purification of

all four proteins encoded by this organism, indicating unique among archaea ability to form heteromeric MCM complexes (Walters and Chong, 2010).

To date, the following crystal structures of archaeal MCM homologous are available: the N-terminal part of *M.thermautotrophicus* MCM (Fletcher et al., 2003) and *S.solfataricus* MCM (Liu et al., 2008b), a near full-length ssoMCM (Brewster et al., 2008) and a full-length inactive MCM from *M.kandleri* (Bae et al., 2009). The overall structure of ssoMCM monomer reveals an elongated form, with five subdomains organized into large N- and C-terminal domains (Figure 1.3). The N-terminal domain consists of three small subdomains (A, B and C) and does not have an helicase activity but it is important for DNA binding and enzyme processivity (Barry et al., 2007). The C-terminal domain is the catalytic part of the protein that binds, hydrolyses ATP and releases ADP and orthophosphate, exploiting the energy from this process to effect DNA melting and translocation along DNA. That domain consists of an AAA+ ATPase core and a small winged helix bundle. The catalytic core contains the conserved motifs found in other AAA+ proteins, including Walker A and Walker B motif. ATP binding and hydrolysis takes place at an interface between two monomers: one monomer provides the tri-phosphate binding loop (P-loop or Walker A and Walker B motif) for ATP binding and the other monomer contributes residues in trans to interact with the ATP (Brewster et al., 2008). The AAA+ core contains also a “glutamate switch” which is believed to regulate ATPase activity upon ligand binding by controlling the orientation of the conserved glutamate residue of the Walker B motif, switching it between active and inactive conformations. The N- and C-terminal domains communicate with each other due the conserved loop called the allosteric control loop (ACL) (Zhang and Wigley, 2008).

Figure 1.3 Domain organization of the monomeric MCM helicases, from (Sakakibara et al., 2009).

A Schematic representation of motifs present in the archaeal MCM proteins. The N-terminal part is divided into three domains, A (red), B (light blue) and C (green). The N-terminal domain is linked to the AAA+ catalytic part by an N–C linker (grey). The AAA+ part is divided into two domains, α/β (yellow) and α (blue), which are connected by a α/β–α linker (brown). The C-terminal HTH region is shown in white. Arrows indicate the main structural motifs. Zn-F, zinc finger; NCL, N-terminal communication loop; β-hp, β-hairpin; Ext-hp, β-hairpin on the exterior of the helicase; W-A, Walker-A; G-S, glutamate switch; H2I-hp, helix-2 insertion β-hairpin; W-B, Walker-B; PS1-hp, pre-sensor 1 β-hairpin; S-I, sensor-1; R-F, arginine finger; S-II, sensor-2.
B and **C:** Structures of the monomeric N-terminal part of *S.solfataricus* (B) and *M. thermautotrophicus* (C) MCM proteins. Colours and domains are as in A.
D and **E:** Structures of the full-length *M.kandleri* (D) and *S.solfataricus* (E) MCM proteins. Colours and domains are as in A.



The structure of the N-terminal domain of the *M.thermautotrophicus* MCM (crystallized as a double-hexamer) was used to generate a model of ssoMCM hexamer (Brewster et al., 2008). The general shape proposed with this model is consistent with earlier electron-microscopy studies (Pape et al., 2003). The hexamer has a wide central channel, narrowing toward the N-terminus and a six side channels extending radially through the side "wall" of the C-terminal helicase domain near the N-terminal side of the ATP binding pocket (Brewster et al., 2008). The central channel is wide enough to accommodate both single- and double-stranded DNA whereas the side channels have dimensions allowing ssDNA to be threaded through which is the base of the proposal that the side channel are the exits for unwound DNA as in the SV40 Large tumour antigen (LTag) hexameric helicase structure (Li et al., 2003, Gai et al., 2004).

Eukaryotic and archaeal MCM complexes, similar to many other AAA+ proteins, assembly into a variety of oligomeric arrangements. The most commonly observed form is a hexamer that might associate in a double hexamer architecture (reviewed by (Slaymaker and Chen, 2012). Double hexamerization of mthMCM in the presence and absent of DNA was shown by X-Ray crystallography and cryo-EM. In these studies mthMCM's N-terminal domains come together in a "head to head" configuration (Fletcher et al., 2003(Gomez-Llorente et al., 2005). Interestingly, the oligomerization state of mthMCM appears to be salt and protein concentration dependent (Gomez-Llorente et al., 2005, Costa et al., 2006). The similar situation is found for LTag where *in vitro* conditions can promote double hexamerization or inhibit it. Electron microscopy data of LTay-dsDNA unwinding complex shows the presence of a "rabbit ear" structures that are indication of bidirectional fork unwinding localized to a double hexamer complex (Wessel et al., 1992).

Like the oligomerization state of MCM helicase *in vivo*, the exact mechanism of the DNA binding, translocation and unwinding, have been the focus of much research effort. Over the last few years, several unwinding models that were proposed for both a single and double hexamer: the steric exlusion model, the rotary pump model, the strand exclusion model and the ploughshare model (Brewster and Chen, 2010, Takahashi et al., 2005). The structural and biochemical data of ssoMCM allowed to limit the list of models to two: steric exclusion and side channel extrusion (Figure 1.4) (Brewster et al., 2008).

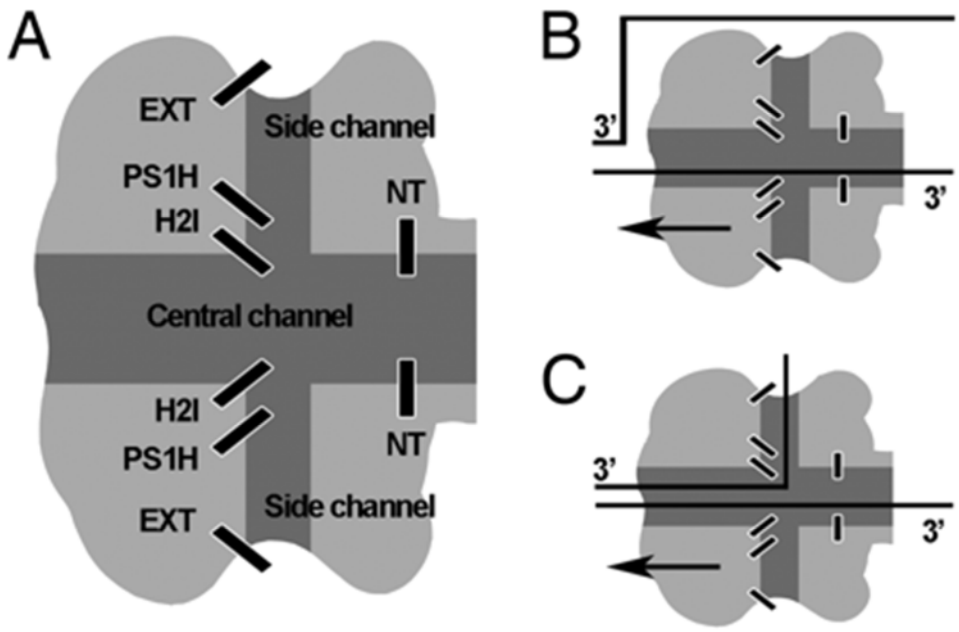


Figure 1.4 Two possible DNA unwinding modes by MCM helicase, adapted from (Brewster et al., 2008).

A Schematic representation of a MCM hexamer helicase. The four β-hairpins (NT, H2I, PS1, and EXT hairpins) are represented by short solid bars; the central channel and the side channels are in darker shades.
B Steric exclusion model for a single hexameric MCM helicase.
C Side-channel extrusion model, showing ssDNA extruding from the side channel. DNA is shown as black lines. Arrows indicate direction of helicase movement.

The steric exclusion model (Figure 1.4 B): DNA duplex is separated prior to the entry to the MCM central channel; one strand proceeds through the central channel, whereas the other strand is displaced away from the channel. The second model (Figure 1.4 C) proposed that double stranded DNA enters the central channel, it is unwound, and one strand is extruded out through a side channel. In both models, the extracted or excluded 5'-strand has no further contact with MCM and does not play any role in the unwinding mechanism (Brewster et al., 2008). It is in opposition with previous evidence from single molecule FRET studies showing a highly dynamic interaction between the 5'-tail and the exterior surface of ssoMCM (Rothenberg et al., 2007). Recently, the steric exclusion and wrapping model of DNA unwinding, termed the SEW model, was proposed (Graham et al., 2011). It is a modified steric exclusion model in which displaced 5'-strand wraps the exterior MCM surface along specific paths, resembling a spool of thread. It is hypothesized that wrapping of the 5'-tail to MCM promotes DNA unwinding and prevent re-annealing behind the helicase (Graham et al., 2011).

In eukaryotic cells, the MCM complex on its own is not an active helicase. It requires association of other protein factors: the tetrameric GINS complex and the CDC45 protein. The three components form together a complex referred as the CMG

complex (CDC45, MCM, GINS), which is believed to be the active replicative helicase *in vivo*. Both GINS and CDC45 seem to increase MCM activity, translocation speed or base pair separation activity (Pacek et al., 2006, Forsburg, 2004, Aparicio et al., 2009, Moyer et al., 2006). Archaeal genomes encode GINS homologous, which also form a tetramer, however until recently it was believed that Archaea does not possess CDC45 homologous. In 2012 a phylogenetic study was published showing that CDC45 might be in fact the eukaryotic orthologue of the bacterial RecJ family nucleases (Makarova et al., 2012). Using the arCOG database, which contains clusters of orthologous genes from the sequenced archaeal genomes, revealed the wide distribution of RecJ homologous (arCOG00427) in archaea. It was previously shown that *S.solfataricus* GINS associates with RecJdbd, which is homologue of DNA-binding domain of bacterial RecJ (Marinsek et al., 2006) and *T.kodakaraensis* GINS associates with GAN which is a homologue of the full-length bacterial RecJ (Li et al., 2010). Considering that Archaea typically possess homologous of the essential components of the eukaryotic replication apparatus, it was proposed that Archaea form a RecJ-MCM-GINS complex that is functional equivalent of eukaryotic CDC45-MCM-GINS complex (Makarova et al., 2012). This hypothesis raises a lot of questions. The obvious one is why the CDC45/RecJ homologous, if they are real homologous/orthologues, are so divergent in comparison to other proteins involved in DNA replication.

1.2.1.3 GINS complex

As described above, in eukaryotic cells, the helicase activity in DNA replication process is performed by the CMG complex, consisting of CDC45, MCM and GINS proteins. GINS itself is a heterotetramer composed of four distinct but related subunits: Sld5, Psf1, Psf2 and Psf3. The name of GINS comes from the Japanese names for the numerals in the subunits Go, Ichi, Ni and San meaning five, one, two and three. Although the exact role of GINS is unknown, this complex is essential for the initiation and elongation stages of chromosome replication (reviewed in (MacNeill, 2010, Labib and Gambus, 2007). Each of four GINS subunit comprises two distinct protein domains: an A domain made up largely of α-helices and a smaller B domain composed mainly of β-strands. Interestingly, the order of the two domains is circularly permuted in the Sld5 and Psf1 subunits in comparison to Psf2 and Psf3 subunits. In Sld5 and Psf1, the A-domain is N-terminal and the B-domain is C-terminal. In opposition, in Psf2 and Psf3 subunits, the B-domain is N-terminal and

the A-domain is C-terminal (Figure 1.5) (MacNeill, 2010, Labib and Gambus, 2007). The two domains are separated by the interdomain loop, whose length varies between species. In 2007 three research groups published the crystal structure of human complex (Boskovic et al., 2007, Chang et al., 2007b, Kamada et al., 2007).

All archaeal genomes sequenced to date encode a single protein with similarity to the eukaryotic Sld5 and Psf1 subunit, including the A and B domain composition. That protein is named GINS51. In addition, some archaeal species also encode a second GINS protein, named GINS23, due its similarities to eukaryotic Psf2 and Psf3 subunits. Interestingly, GINS23 is present in organisms belonging to *Thaumarchaeota*, the ancient lineage of Archaea, which might suggest that the last common archaeo-eukaryotic ancestor encoded both proteins.

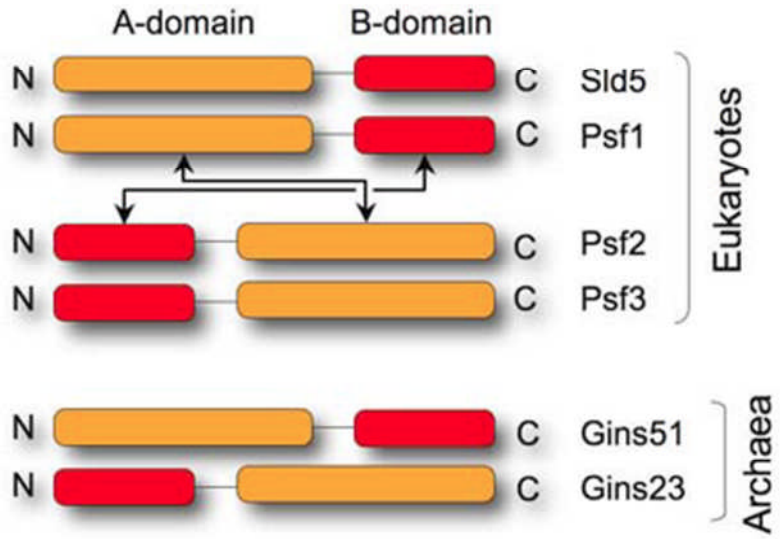


Figure 1.5 Schematic representation of the domain organization of eukaryotic and archaeal GINS proteins, adapted from (MacNeill, 2010).

Biochemical and structural studies indicate that archaeal GINS complex is simplified version of the eukaryotic counterpart. GINS homologous from *S.solfataricus* and *P.furiosus* have been shown to interact with MCM (Yoshimochi et al., 2008, (Marinsek et al., 2006). A chromatin immunoprecipitation assay revealed that *P.furiosus* GINS is detected preferentially at the *oriC* region during exponential but not stationary growth phase. In addition, the GINS complex stimulates the ATPase and helicase activity of MCM *in vitro* (Yoshimochi et al., 2008b). The basic question about archaeal GINS was what is the structure of the complex? Do species with both GINS51 and GINS23 proteins form a tetrameric complex that resembles the eukaryotic one and what is architecture of complex in organisms apparently lacking GINS23? In 2011, the crystal structure of the GINS complex from *T.kodakaraensis*

was determined (Oyama et al., 2011). The *T.kodakaraensis* GINS is a tetramer comprising GINS51 and GINS23 dimers (Figure 1.6). That tetramer is similar to human GINS tetramer in overall architecture and size. The two major differences were seen between those structures: the position of the B domain of the GINS51 and difference on the contact surface between GINS51 and GINS23 (upper-lower). The most significant difference is position of the B domain of the GINS51 in comparison to the corresponding domain from human Sld5 and Psf1. In the human GINS, the B domain of Sld5 is involved in the contact with Psf2 and stable tetramer formation. In contrast, B domain of Psf1 is more mobile and is not essential for complex stability. (Boskovic et al., 2007, Chang et al., 2007b, Kamada et al., 2007). In this regard the B domain of GINS51 is more similar to the human Psf2 B domain. To determine whether the GINS51 B domain played a structural role in tetramer formation, *T.kodakaraensis* mutant GINS51 lacking B domain was constructed and oligomeric state of the GINS complex was tested by gel filtration. The truncation mutant formed a stable tetramer indicating that GINS51 B domain is not required to maintain the complex native architecture (Oyama et al., 2011). The fact that the B domain appears mobile allows speculation that it might be responsible for interaction with other components of replication machinery (Bell, 2011).

The crystal structure of *T.kodakaraensis* GINS tetramer was used for modelling of the putative homotetrameric GINS complex from *Thermoplasma acidophilum.* This thermophilic organism encodes a GINS51 protein only. Gel filtration and electron microscopy analysis indicated that *T.acidophilum* GINS forms an α_4 homotetramer. In the homotetramer model, on the upper layer the B domains are placed in the same orientation as in *T.kodakaraensis* GINS51 and on the bottom layer, domains are positioned according to BA-type. In that model, the homotetrameric GINS complex is asymmetric due to the B domains repositioning. That differential positioning is allowed by the inter-domain loop, which is elongated in comparison to other homologous and, considering its amino acids content, is prone to form an intrinsically disordered structure. The loop might also prevent non-functional polymerization on the tetramer surface (Oyama et al., 2011).

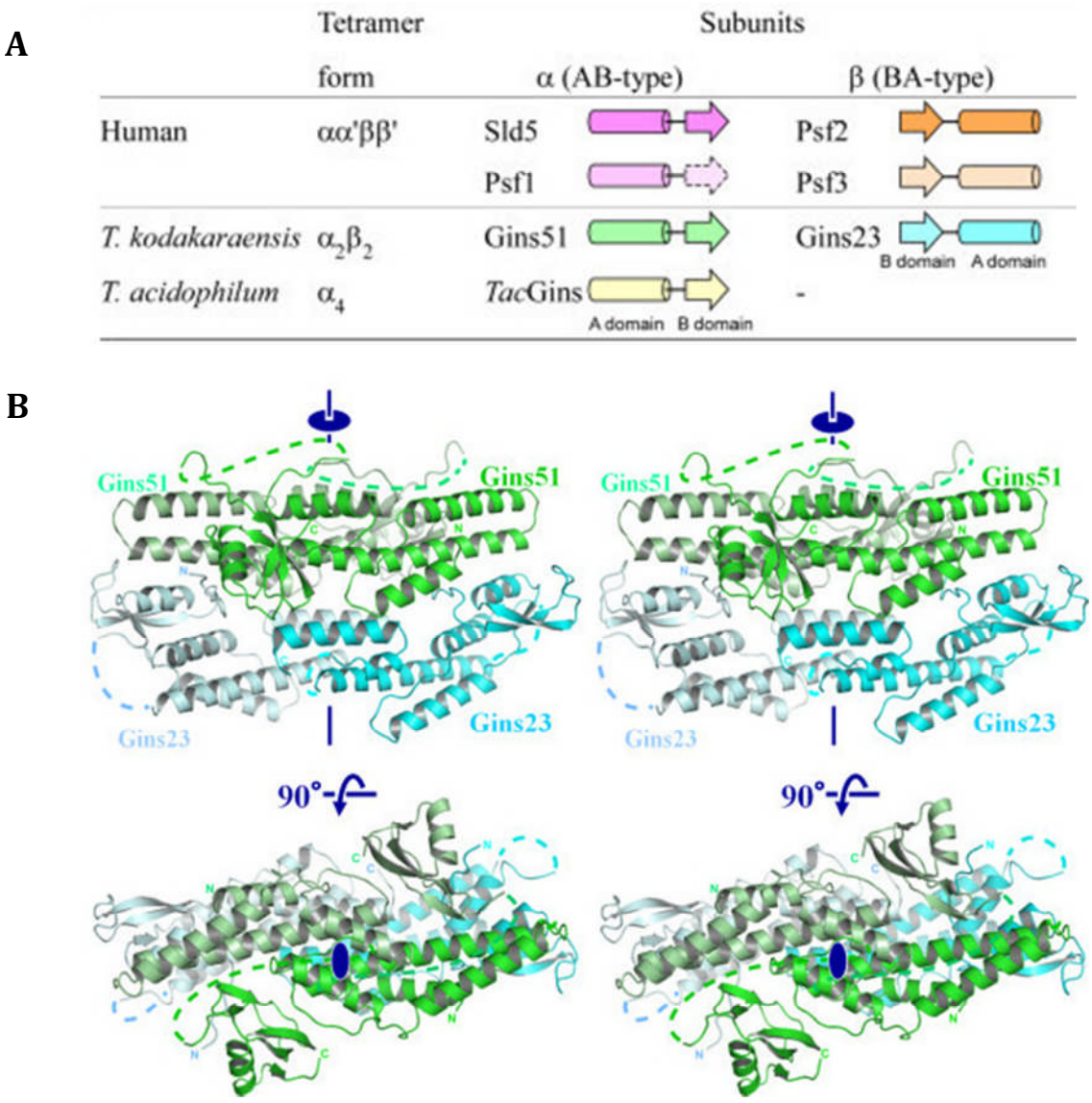


Figure 1.6 The crystal structure of *T.kodakaraensis* GINS, adapted from (Oyama et al., 2011).

A Domain organization of GINS from human, *Thermococcus kodakaraensis*, and *Thermoplasma acidophilum*. Cylinders represent A-domains mainly composed of α-helices and arrows show B-domains rich in β-strands. The B domain of human Psf1 is missing in the crystal structures.
B Ribbon representation of the structure. The GinS51 subunits are colored green, and the GinS23 subunits are cyan. Missing parts are shown with dotted lines. The crystallographic two-fold axis is indicated.

The structure of archaeal GINS complex is a significant step towards understanding the architecture of the replication fork assembly. Based on the recent findings in the eukaryotic and archaeal research field, a model for the initial assembly of the archaeal replisome was proposed (Figure 1.7) (Bell, 2011). This is a highly speculative model that assumes that MCM is loaded at replication origins as a head to head double hexamer and RecJ/GAN are indeed parts of the archaeal CMG

complex. RecJ/GAN would interact with GINS51 B domain and direct single-stranded DNA generated by the MCM helicase activity to the catalytic site of primase, ensuring coupling of DNA unwinding and priming activity.

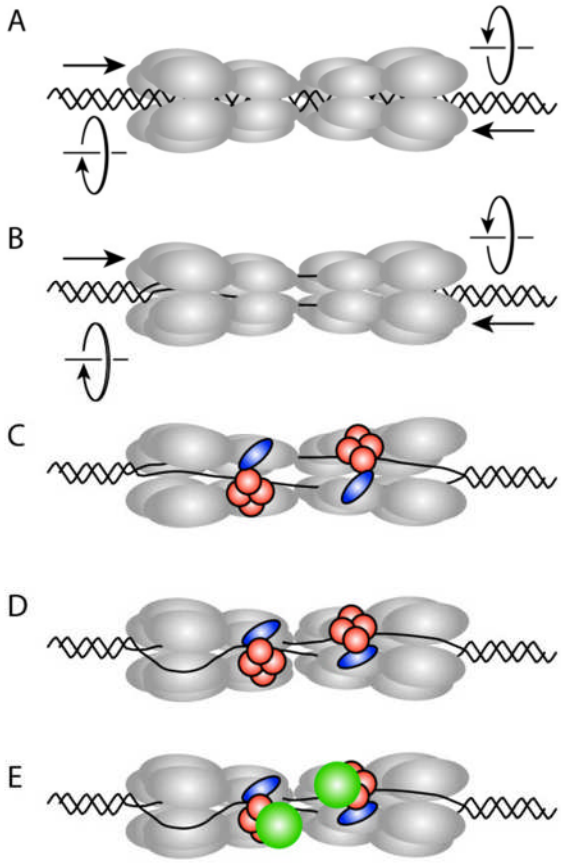


Figure 1.7 Recent model of assembly of archaeal replisome, from (Bell, 2011)

A A double hexameric MCM (grey) is loaded on dsDNA at replication origin
B DNA is pumped into the central cavity of the assembly by the two MCM hexamers held together
C The GINS complex (orange) together with RecJ/GAN (blue) stabilizes an open form of the MCM and promotes extrusion of one DNA strand (the lagging strand)
D MCM is released, one DNA strand (the leading strand) is passing through the center of the helicase and the displaced strand is trapped between the outside of MCM and GINS
E GINS recruits DNA primase (green)

1.2.1.4 Single-stranded DNA-binding proteins

Single-stranded DNA-binding proteins (SSBs) are widespread proteins playing indispensable roles in many aspects of DNA metabolism including replication, repair and recombination, making them central factors in the maintenance of genomic integrity (Flynn and Zou, 2010, Pestryakov and Lavrik, 2008, Broderick et al., 2010). As this thesis presents genetic analysis of RPA function in the archaeal organism *H.volcanii*, single-stranded DNA-binding proteins will be described in greater detail, beyond their function in DNA replication.

Single-stranded DNA-binding proteins accomplish the essential process of stabilizing exposed ssDNA regions, protecting them from chemical and nucleolytic attack and preventing formation of secondary structures that might affect on-going cellular processes. In addition to binding ssDNA, SSBs interact with many protein factors, playing roles in orchestrating the protein-DNA and protein-protein interaction.

The characteristic feature of all SSBs is the presence of one or more OB fold (oligosaccharide-oligonucleotide binding fold) domains (Bochkarev and Bochkareva, 2004). The OB fold is a compact structural motif that ranges from 75 to 150 residues in length and, despite low degree of sequence similarities shows structural conservation (Murzin, 1993). The OB fold consists of a five-stranded β-sheet that is coiled to form a closed β-barrel, often capped by an α-helix located between the third and fourth strands. The variation in length among OB fold domains from different proteins is caused by the presence of additional loops located between well-conserved β-strands (Flynn and Zou, 2010, Murzin, 1993, Pestryakov and Lavrik, 2008) (Figure 1.8).

The OB fold motif is found in various protein backgrounds; according to the SCOP (Structural Classification of Proteins) database (Murzin et al., 1995), the OB fold motif is present in eight distinct superfamilies. In addition to SSBs, where ssDNA is bound with no sequence specificity, OB folds are present in proteins, such as the bacterial transcriptional terminator Rho, that recognize specific single-stranded regions or in proteins that interact with mainly non-helical structured nucleic acids. An example of a protein belonging to the latter family is the ribosomal *Thermus thermophiles* ribosomal initiation factor 1 (Flynn and Zou, 2010).

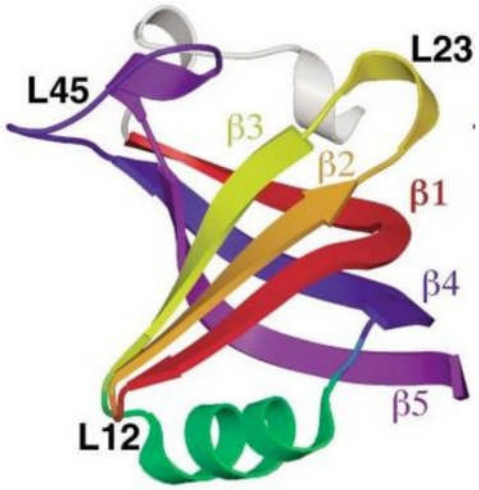


Figure 1.8 Structure of OB fold domain and OB fold/nucleic acid complex, adapted from (Theobald et al., 2003)
Structure of canonical OB fold domain. From the N terminus to the C terminus, strand β1 is shown in red, β2 in orange, β3 in yellow, the helix between β3 and β4 in green, β4 in blue, and β5 in violet. An α-helix, which is found in half of the OB-folds in these complexes, is shown in white at the top of the figure, just N-terminal to strand β1. Variable loops between strands are indicated in black text.

A number of structures of SSB-DNA complexes have been solved, providing insights into the molecular mechanism of the OB fold-ssDNA interaction. Two kinds of interactions are particularly important to mediate the protein-nucleotide interaction: [1] stacking interactions between nucleotide residues and conserved aromatic amino acids side chains in the protein and [2] packing interactions with

hydrophobic side chains or the aliphatic parts of more polar groups like lysine and arginine (Theobald et al., 2003).

An additional structural element found in eukaryotic and euryarchaeal SSBs is the zinc finger. Zinc ions are widely used as a cofactors and the zinc finger motif has a functional role in nucleic acid binding, protein-protein interactions and the binding of small ligands (Matthews and Sunde, 2002). The zinc finger present in eukaryotic proteins belongs to the $X_3CX_{2-4}CX_{12-15}CXC$ family (where C is a cysteine residue and X is any amino acid residue), and in human protein (RPA70, see below) is located in the C-terminal domain of the large protein subunit within the variable loop of the OB fold (Iftode et al., 1999). Zinc finger motifs were also found in several euryarchaeal single-stranded DNA-binding proteins but the amino acid sequence is different and includes one histidine residue (general consensus is $CX_2CH_8CX_2H$) (Lin et al., 2005). Although the zinc finger is not essential for human RPA protein binding to ssDNA, biochemical study of *Methanosarcina acetovorans* mutant proteins indicated that the zinc finger might regulate ssDNA binding as a redox agent (Lin et al., 2005).

SSBs display a wide variety of domain and subunit organizations across evolution (Pestryakov and Lavrik, 2008). In Bacteria, the great majority of SSB family members comprise a single OB fold located at the N-terminus and a flexible C-terminal tail that mediates protein-protein interaction (Pestryakov and Lavrik, 2008). Four individual SSB subunits assemble to form a homotetrameric complex that may bind to ssDNA in a variety of modes depending on solution conditions, for example at low salt concentration *E.coli* SSB binds to approximately 35 nucleotides using only two out of four subunits (SSB_{35} mode), whereas at high salt concentration SSB binds to ssDNA using all four subunits and occluding approximately 65 nucleotides (SSB_{65} mode) (Roy et al., 2007). In some bacterial lineages, such as in the *Deinococcus* and *Thermus* genera, SSB proteins adopt a different architecture. *Deinococcus radiodurans* SSB (DrSSB) is a homodimer, with both monomers comprising two OB folds linked by a conserved linker sequence. The X-ray crystal structure of the DrSSB-ssDNA complex showed strong similarities in DNA binding mechanisms employed by the two classes of bacterial SSB (George et al., 2012).

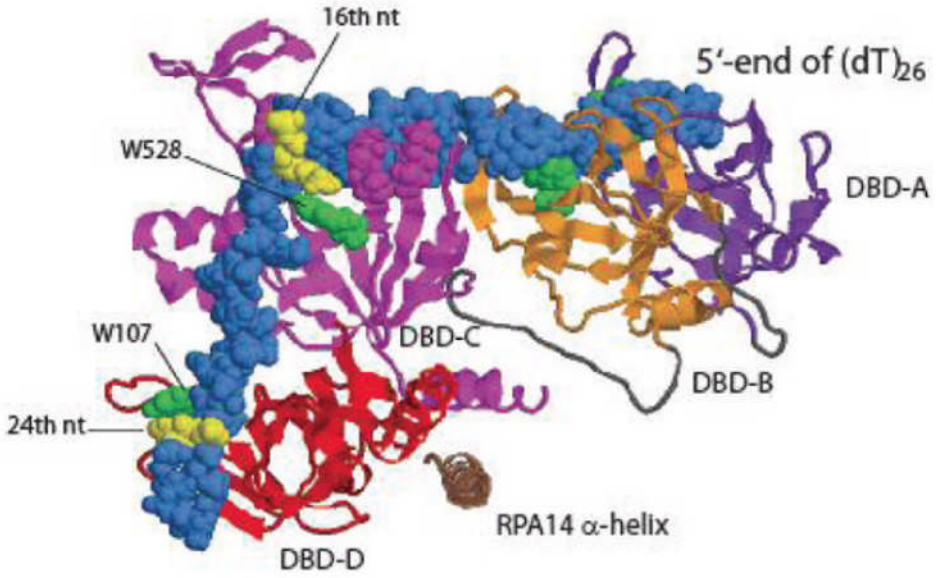
In eukaryotes, the major cellular SSB is replication protein A (RPA), a heterotrimeric complex comprising of three distinct subunits of approximately 70, 32 and 14 kDa (RPA70, RPA32 and RPA14, respectively). The RPA trimer has a six OB fold domains referred to as DBD-A to -F (where DBD means DNA binding domain and

the order A-F indicates binding efficiency to ssDNA). RPA70 contains four OB folds (DBD-A to –C and DBD-F) and RPA32 and RPA14 one each (DBD-D and DBD-E, respectively). In fact, only four OB folds (DBD-A to -D) participate in ssDNA binding; the remaining two OB folds mediate protein-protein interactions. In addition to OB folds, RPA70 also possesses the already mentioned zinc finger motif and RPA32 possesses an extended N-terminal domain which is a target for regulatory phosphorylation, and a winged helix-turn-helix domain at the C terminus (Broderick et al., 2010).

RPA binds to ssDNA in sequential manner: first, DBD-A and –B bind to 8-10 nucleotides of ssDNA; this causes a conformational changes in the complex allowing subsequent interaction DBD-C and DBD-D with the ssDNA substrate. The full binding mode occludes 28-30 nucleotides (Figure 1.9) (Fanning et al., 2006).

Figure 1.9 Structures of eukaryotic RPA protein bound to DNA, adapted from (Fan and Pavletich, 2012)

A model for the hetrotrimeric eukaryotic RPA bound to 26 nucleotides of single-stranded DNA $(dT)_{26}$, coloured blue. The DNA-binding domains (DBD) A, B, C, and D are coloured purple, orange, magenta and red. The DBDA-C are located in the RPA70 subunit and the DBDD in RPA32. Four tryptophan residues, coloured green, are thought to interact with ssDNA, specifically dT16 and dT24, which are highlighted in yellow.



Although the canonical RPA is believed to be the major cellular SSB in eukaryotic cells, a number of additional SSBs were discovered during last few years in various

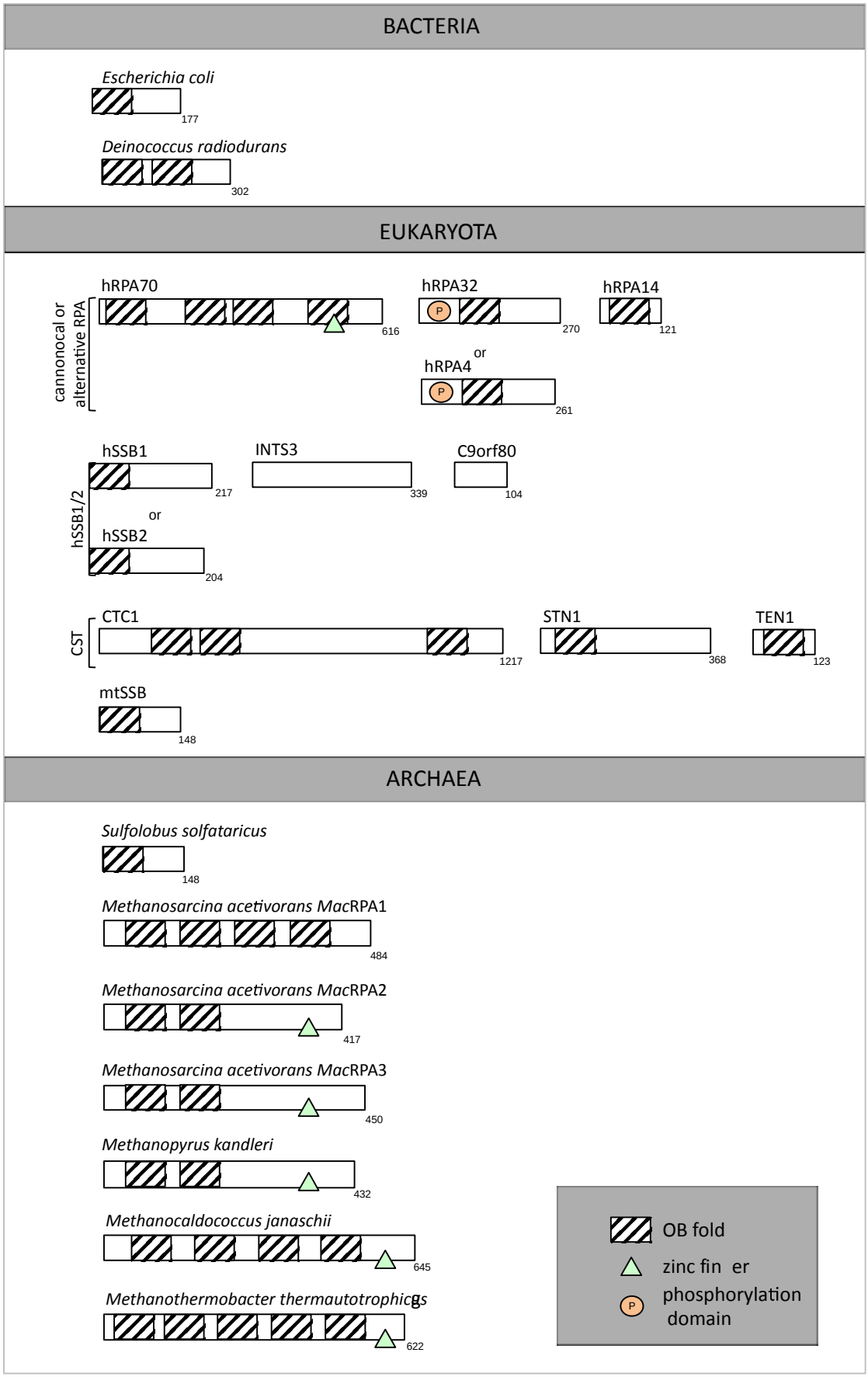
eukaryotic lineages (Flynn and Zou, 2010). A 34-kDa homologue of RPA32 termed RPA4 was identified in humans (Keshav et al., 1995) and in a seed plants (Sakaguchi et al., 2009). RPA4, like RPA32, carries one OB fold and winged helix-turn-helix domain and is capable of forming a stable complex with RPA70 and RPA14. The alternative RPA complex (RPA70-RPA4-RPA14) might be involved in DNA repair as it interacts with Rad52 and Rad51, stimulates Rad51 strand exchange, supports the dual incision/excision reaction of the nucleotide excision repair pathway (Kemp et al., 2010) and does not support chromosomal DNA replication (Haring et al., 2010). Another RPA-like complex, named the CST complex, is involved in telomere maintenance. This complex is composed of Cdc13, Stn1 and Ten1 subunits and like RPA contains multiple OB folds (reviewed by (Price et al., 2010). Two additional mammalian SSBs identified recently are named hSSB1 and hSSB2 (Richard et al., 2008). Based on amino acid sequence alignment and domain composition these proteins are more closely related to bacterial SSB than to RPA. Both hSSB1 and hSSB2 are composed of a single OB fold at the N-terminus and a C-terminal tail predicted to play a role in protein-protein interaction. Three independent research groups showed that hSSB1 and hSSB2 form two distinct heterotrimeric complexes with the integrator complex subunit 3 (INTS3) and the previously uncharacterized protein, C9orf80 (Huang et al., 2009, Li et al., 2009, Zhang et al., 2009). HSSB1/2 are the middle subunits in the corresponding complexes and provide the only OB folds in the complexes, as neither INTS3 or C9orf80 appear to contain this structural motif. hSSB1 was proved to participate in the early stage of repair of DNA double-strand breaks (DSBs) by the homologous recombination pathway by direct interaction with Mre11-Rad50-NBS1 complex, recruiting it to the DSBs and stimulating its nuclease activity (Richard et al., 2011). Finally, eukaryotes also encode a mitochondrial SSB (mtSSB) involved in mitochondrial DNA replication (Tiranti et al., 1993).

In the third domain of life, the archaea, SSBs display an unusual variety of architectures. In general SSB proteins are divided into two groups: those that resemble bacterial SSB (found in Crenarchaea) and those that resemble eukaryotic RPA (found in Euryarchaea). The best-characterized SSB from the Crenarchaeal phylum comes from the thermophilic organism *Sulfolobus solfataricus* (SsoSSB) and contains a single OB fold followed by a C-terminal tail that is not essential for DNA bonding but which is involved in protein-protein interactions. Although the domain organisation of SsoSSB resembles its bacterial counterpart, the structural data

indicates that the OB fold domain present in the crenarchaeal protein is more similar to those from eukaryotic RPA (Kerr et al., 2003).

Single-stranded-DNA binding proteins found in Euryarchaea display diverse subunit organization and distribution but again, the OB folds in euryarchaeal proteins are more related to eukaryotic RPAs than bacterial SSBs at both amino acid sequence and structure level. RPAs from species representative of several major lineages have been identified and characterised. The hyperthermophilic archaeon *Pyrococcus furiosus* encodes three RPA proteins, RPA41, RPA32 and RPA14, which assemble to form a heterotrimeric complex. The complex, in contrast to each subunit on its own, was shown to bind ssDNA (Komori and Ishino, 2001). The methanogenic species *Methanosarcina acetivorans* encodes three RPA proteins named MacRPA1-MacRPA3; each MacRPA can act as a distinct SSB and oligomerize to form a homomultimeric complex (Lin et al., 2008, Robbins et al., 2004b). Particular MacRPA proteins vary in terms of domain organisation. MacRPA1 is composed of four tandem OB folds while MacRPA2 and MacRPA3 contain two OB folds and a $CX_2CX_8CH_2H$ zinc finger motif (Lin et al., 2005). The combination of multiple OB folds and a putative zinc finger is also found in other methanogenic and non-methanogenic RPAs. As examples, *Methanopyrus kandleri* RPA contains two OB folds and a zinc finger (Robbins et al., 2005), *Methanocaldococcus jannaschii* RPA contains four OB folds and a zinc finger (Kelly et al., 1998), *Methanothermobacter thermautotrophicus* RPA contains five OB folds and a zinc finger, *Ferroplasma acidarmanus* and *Thermoplasma volcanium* RPAs contains two OB folds and a zinc finger) (Lin et al., 2008). It is believed that the complexity in euryarchaeal RPAs distribution that we see today is due to several gene duplication, fission and/or intermolecular recombination events (Lin and Ha, 2006, Lin et al., 2008, Robbins et al., 2005).

Figure 1.10 Diversity of domain architecture of single-stranded DNA-binding proteins in three domains of life. Cartoon showing OB folds (hatched boxes) and the zinc finger motifs (green triangle). Proteins drawn in one line form a complex. Protein lengths are indicated. The domains are not drawn to scale.



As single-stranded DNA-binding proteins are vital players in many cellular processes, all living organisms are believed to encode SSB homologous. Intriguingly, in the *Thermoproteales*, a clade of hyperthermophilic Crenarchaea, the canonical SSB is substituted by a distinct ssDNA binding protein named ThermoDBP. These proteins, exemplified by recently identified protein Ttx1576 from *Thermoproteus tenax*, do not contain classical OB fold motif and bind to ssDNA via an extended cleft with a aromatic core interacting with the nucleotide bases and a positively charged region suitable for interactions with the phosphate backbone of DNA (Paytubi et al., 2012). Lack of the canonical OB fold-containing SSB in the *Thermoproteales* is an interesting example of non-homologous gene displacement in genome evolution.

Single-stranded DNA-binding proteins play essential roles in DNA replication. In fact, RPA, as the name suggests, was originally identified as a protein required for SV40 virus to replicate its DNA (Fairman and Stillman, 1988). In general, SSB binds to ssDNA regions generated by the replicative helicase, stabilizes their structure and mediates the assembly of other proteins. Recently studies identified at least fourteen replication proteins interacting with human RPA, shedding light on the complex role that RPA plays in the regulation and progression of DNA synthesis (Nakaya et al., 2010). At the initiation step of DNA replication, when the pre-replication complex is formed, human RPA interacts with DNA polymerase α/primase complex, increasing DNA polymerase processivity and accuracy (Maga et al., 2001). Similar results were obtained for archaeal RPA proteins: three RPA homologous from *Methanosarcina acetivorans* were shown to increase processivity of polymerase B1 (Robbins et al., 2004a). In eukaryotes, RPA is also involved in the switch from DNA polymerase α to polymerase δ at the transition from the initiation to the elongation step of replication. This switch is mediated by replication factor C, which competes with Pol α-RPA interaction (Maga and Hubscher, 1996). Finally, RPA participates also in the removal of RNA primers from Okazaki fragments. RPA was shown to coordinate the sequential action of two endonucleases, Dna2 and Fen1 in yeast (Bae et al., 2001).

Many experimental outcomes provide evidence of the essential role of SSBs in DNA damage checkpoints and all major types of DNA repair including nucleotide excision repair, base excision repair, mismatch repair and double-strand break repair. The properties of mammalian RPAs in this area remain under extensive study as defects in RPA-associated cellular pathways lead to genomic instability and, as a consequence, many diseases including cancer (reviewed by (Broderick et al., 2010, Fanning et al., 2006, Pestryakov and Lavrik, 2008, Zou et al., 2006). RPAs are one

of the first proteins that interact with damaged DNA and they are involved in recruiting a number of repair proteins at the damage site (for example XPA, XPG, uracil-DNA glycosylase, Rad51, MRN complex)(Ramilo et al., 2002, DeMott et al., 1998, Reardon and Sancar, 2005, Eggler et al., 2002). DNA repair pathways in Archaea are still not fully understood; the step of DNA damage recognition is especially puzzling as most archaea lack any obvious homologous of MutS or the XPA and XPC proteins that initiate the mismatch- and nucleotide excision repair, respectively (Kelman and White, 2005). SSBs are strong candidates to play a role in DNA damage recognition in Archaea (Cubeddu and White, 2005).

1.2.2 Replication elongation

Once the origin DNA is unwound and the replication fork is formed, DNA primase synthesises a short RNA oligonucleotide that serves as a primer for the initiation of replication on the leading strand and for the initiation of each Okazaki fragment on the lagging strand. However, before DNA polymerase can start its activity in DNA synthesis, it has to be tethered to DNA by the accessory factor PCNA. PCNA is a toroidal trimer that encircles the DNA forming a sliding clamp.

1.2.2.1 Primase

Primase is a DNA-dependent RNA polymerase that synthesizes short, 8-12 nucleotide primers, which are used by DNA polymerase to start elongating the leading strand and to form the Okazaki fragments. In bacteria, DNA primase is a single polypeptide, DnaG. In eukaryotes, primase is a dimer consisting of a small, catalytic subunit and large, non-catalytic subunit. These subunits associate with the catalytic and B-subunits of DNA polymerase α forming a four-subunit complex, the Polα-primase complex (reviewed in (Kuchta and Stengel, 2009)). Archaea possess bacterial- and eukaryotic-like primase genes in their genomes, but DnaG homologous appear to be non-essential for cell viability and most likely are not involved in DNA replication (*dnaG* can be deleted in *H.volcanii* and deletion strain does not show any obvious phenotype under laboratory conditions) (Le Breton et al., 2007). Furthermore, in *S.solfataricus*, the DnaG protein associates with the exosome-like complex, suggesting a role in RNA metabolism in this species (Evguenieva-Hackenberg et al., 2003). In contrast, the eukaryotic-like DNA primase was investigated in several archaeal species and was shown to interact with replication protein A in *P.furiosus* supporting its involvement in replication (De Falco et al., 2004, Bocquier et al., 2001, Desogus et al., 1999, Lao-Sirieix and Bell, 2004, Liu et

al., 2001, Matsui et al., 2003). In archaea, the small and large primase subunits are designated PriS and PriL, respectively. The crystal structures of the PriS proteins from *P.furiosus* (Augustin et al., 2001), *P.horikoshii* (Ito et al., 2003) and *S.solfataricus* (Lao-Sirieix and Bell, 2004) are available, offering structural insights into archaeal primase function. PriS from *Pyrococcus* species is composed of three distinct domains: an α/β domain that harbours the catalytic prim domain, a small α-helical domain of unknown function and a zinc-binding domain probably responsible for the interaction with DNA. Archaeal primase small subunits contain also a triple-aspartate motif that resembles part of a PolX family of DNA polymerases. In eukaryotic cells PolX enzymes are involved in DNA replication, repair and recombination. Within Archaea only *M.thermautotrophicus* has been shown to possess a PolX homolog, which might suggest that primase is taking part in DNA repair (Le Breton et al., 2007). In comparison to PriS from *Pyrococcus* species, *S.solfataricus* PriS has an α-helical domain reduced to a single irregular helix and the zinc-binding motif is located at the end of an extended β hairpin structure that is absent from the *Pyrococcus* proteins. In addition, *S.solfataricus* PriS, like majority of PriS proteins from other archaeal species, possess a mixed α/β domain at the C-terminus (termed the PriS-CTD). The functional and evolutionary role of that domain will be discussed in Chapter 6. The primase large subunit consists of a α-helical domain with a small α/β domain that mediates PriL- PriS interaction. Biochemical studies on *P.abyssi* enzyme showed that the small subunit on its own has no RNA synthesis activity but can synthesize long, up to 3 kb, DNA strands. Addition of the large subunit increases the rate of DNA synthesis but at the same time decreases the length of synthesized fragments and also confers the ability to RNA synthesis (Le Breton et al., 2007).

1.2.2.2 DNA Polymerase

DNA replication is achieved by a DNA-dependent DNA polymerase that uses ssDNA as a template to synthesize the complementary strand. Based on amino acid sequences DNA polymerase enzymes are classified into at least six main families: A, B, C, D, X and Y (Ito and Braithwaite, 1991). In eukaryotic cells three B polymerases participate in DNA replication: Polα, Polδ and Polε. Briefly, the four subunits Polα/primase is initially loaded onto the origin of DNA replication and makes short RNA primers that are extended by the polymerase to about 30 bases.

Then elongation is taken over by Pol ε (on the leading strand) and δ (on the lagging strand) (reviewed by (Stillman, 2008).

The archaeal world is divergent in term of encoded genes for DNA polymerases. *Crenarchaeota* possess single or multiple B-type polymerases (PolB), whereas members of *Euryarchaeota* encode a monomeric PolB and a dimeric D-type polymerase (PolD) (Gueguen et al., 2001).

PolB is a monomer, with one exception for *M.thermautotrophicus* enzyme, which is split into two polypeptides (Kelman et al., 1999). Crystal structures have been revealed for PolB proteins from several archaeal species (Hopfner et al., 1999, Rodriguez et al., 2000, Zhao et al., 1999). Those enzymes have similar amino acid sequences, domain organisation and overall structures, and all possess 3'-5' exonuclease proofreading activity, however there are some minor biochemical differences between them. Those differences might be reflection of adaptation to diverse environmental conditions under which the different species live (Perler et al., 1996). The activity of PolB proteins from several *Crenarchaeota* and *Euryarchaeota* is stimulated by PCNA and RFC, which support the theory that B-type enzymes are likely to function as the replicative polymerases in archaea.

An interesting feature that has been demonstrated for archaeal family B polymerases is the ability to sense uracil in DNA template and to then stall replication four bases ahead of that pro-mutagenic residue. This ability is facilitated by a small, conserved pocked located in the N-terminal domain of the polymerase that was shown to binds uracil with high affinity (Connolly et al., 2003, Fogg et al., 2002). The 3'-5' exonuclease proofreading activity of PolB hinders the copying of template strand beyond deaminated bases (Russell et al., 2009).

Polymerase D is a euryarchaeal-specific DNA polymerase that was identified for the first time in *P.furiosus* (Ishino et al., 1998). PolD is a two-subunit enzyme comprised of small, non-catalytic and large, catalytic subunits, PolD1 and PolD2, respectively. Uniquely, in *Pyrococcus*, genes encoding both subunits are in an operon with additional genes involved in DNA metabolism and with *oriC* located in close proximity (Gueguen et al., 2001). PolD1 shows homology to the non-catalytic subunits of the eukaryotic polymerases, which belong to the large, calcineurin-like phosphoesterase superfamily based on the presence of five conserved motifs. This superfamily consists of enzymes with a common di-metal active site, but diverse functions (Aravind and Koonin, 1998). It was proposed that the phosphoesterase

domain is responsible for the 3'-5' exonuclease activity of PolD (Gueguen et al., 2001).

The large subunit, PolD2, does not have sequence similarity to any known protein and without association with PolD1 possess only limited polymerase activity. The active site for DNA polymerization was identified as Asp1122 and Asp1124, both located in the most conserved region of PolD2 (Shen et al., 2001).

The most extended biochemical studies on archaeal DNA polymerases were performed for *P.abyssi* enzymes (Castrec et al., 2010, Castrec et al., 2009, Henneke et al., 2005). Although it was shown that both PolB and PolD are DNA polymerizing enzymes they have different DNA binding properties. PolD is a primer-directed DNA polymerase independent of the primer composition. In contrast, PolB discriminates between RNA or DNA and can only extend a DNA-primed template. PolD, but not PolB requires interaction with PCNA for efficient DNA synthesis (Henneke et al., 2005).

Biochemical and structural features of both PolB and PolD makes them potential candidates to be involved in chromosomal DNA replication. It was even proposed that PolB synthesizes the leading strand and PolD is responsible for the lagging strand replication (Henneke et al., 2005). However, that hypothesis cannot be accepted without performing more experimental studies since there is no direct evidence that either PolB or PolD are involved in DNA replication *in vivo*.

1.2.2.3 PCNA, the DNA sliding clamp

PCNA (Proliferating Cell Nuclear Antigen) is a ring-shaped complex that encircles double-stranded DNA and tethers DNA polymerases to DNA. This increases its polymerase processivity (from tens to thousands of nucleotides, as has been shown in *in vitro* studies), and acts as a platform onto which diverse factors involved in replication are assembled. The sliding clamp, conserved in structure and function, is present in all three domains of life indicating its indispensability in DNA synthesis (reviewed in (Moldovan et al., 2007)). PCNA forms a ring-shaped complex (homodimeric in bacteria, homotrimeric in eukaryotes and homo- or heterotrimeric in archaea) with pseudohexameric symmetry. Each monomer displays the two globular domains, linked by long loop, named the interdomain connecting loop. Monomers arranged head-to-tail form the ring with an inner positively charged surface (altered in halophilic archaea, see below) made of α-helices and an outer surface composed of β-sheets (Moldovan et al., 2007). The positive charge of the

pore is proposed to reduce repulsion with the negatively charged phosphate backbone of DNA (Winter et al., 2009). As mentioned above, PCNA interacts with many protein factors, including Flap endonuclease 1 (Fen1), Replication factor C (RFC) and DNA ligase (McNally et al., 2010, Meslet-Cladiere et al., 2007, Maga and Hubscher, 2003, Vivona and Kelman, 2003). The interaction between PCNA and its partner is commonly facilitated through a conserved motif named PIP motif (PCNA-interacting motif) (Warbrick, 1998). The core element of the PIP box is a peptide with the sequence QxxΨ (Ψ being the hydrophobic residues L, M, or I) (Xu et al., 2001). The PIP motif is usually located at the N- or C-terminal ends of proteins, although it also might be positioned internally. In addition, some proteins that contain a PIP motif may use different regions to interact with PCNA (Moldovan et al., 2007). Because PCNA is a trimer and each monomer has a interdomain connecting loop, up to three different proteins can potentially be loaded onto a single PCNA complex simultaneously, making PCNA a "molecular tool-belt" (Strzalka and Ziemienowicz, 2011).

Archaeal PCNA complexes fall into two groups. Euryarchaea usually encode a single PCNA that forms a homotrimer. In contrast species belonging to Crenarchaea, encode multiple PCNA proteins, for example *S.solfataricus* has three proteins, which form a heterotrimer (Dionne et al., 2003).

The crystal structures of several archaeal PCNAs have been solved, including homotrimeric complexes from *A.fulgidus* (Chapados et al., 2004), *P.furiosus* (Matsumiya et al., 2001) and *H.volcanii* (Morgunova et al., 2009, Winter et al., 2009) and heterotrimer from *S.solfataricus* (Williams et al., 2006). Although the overall structure of all those complexes is conserved, there are some differences in amino acid composition and three-dimensional structures that are believed to be adaptations to different environmental conditions (Figure 1.11). *P.furiosus* PCNA shows an increase in charged residues, shortening of loop length and the gain of ion pairs (Figure 1.11a, left) (Matsumiya et al., 2001). *H.volcanii* PCNA is almost completely devoid of the usual positive surface charge in the inner pore. Typically, the inner pore, a part of the protein through which DNA passes, is positively charged due to the presence of lysine and arginine residues on the twelve α-helices that line the pore. *Hfx*PCNA possesses only two those basic residues (per monomer) in comparison to 9-12 from a typical sliding clamp. Lys143 and Lys205 extend into the channel to contribute to the proposed water-mediated interactions with the phosphate backbone of DNA passing through the pore. PCNA proteins are usually highly

acidic proteins, however that effect is more apparent in *Haloferax*, as can be clearly seen by mapping the electrostatic potential of *Haloferax* and *Pyrococcus* PCNA surfaces (Figure 1.11b) (Winter et al., 2009). An increase in acidic residues on the protein surface is a common adaptation to high salt conditions (Madern et al., 2000, Mevarech et al., 2000). Finally, *S.solfataricus* PCNA is composed of three monomers, SsoPCNA1, SsoPCNA2 and SsoPCNA3. SsoPCNA1 appears to have a distinct structure and plays the key role in the complex assembly. The trimer is formed by a strong interaction between subunits 1 and 2, followed by a much weaker interaction with subunit 3 which might explain the self-loading phenomenon (see below) (Hlinkova et al., 2008).

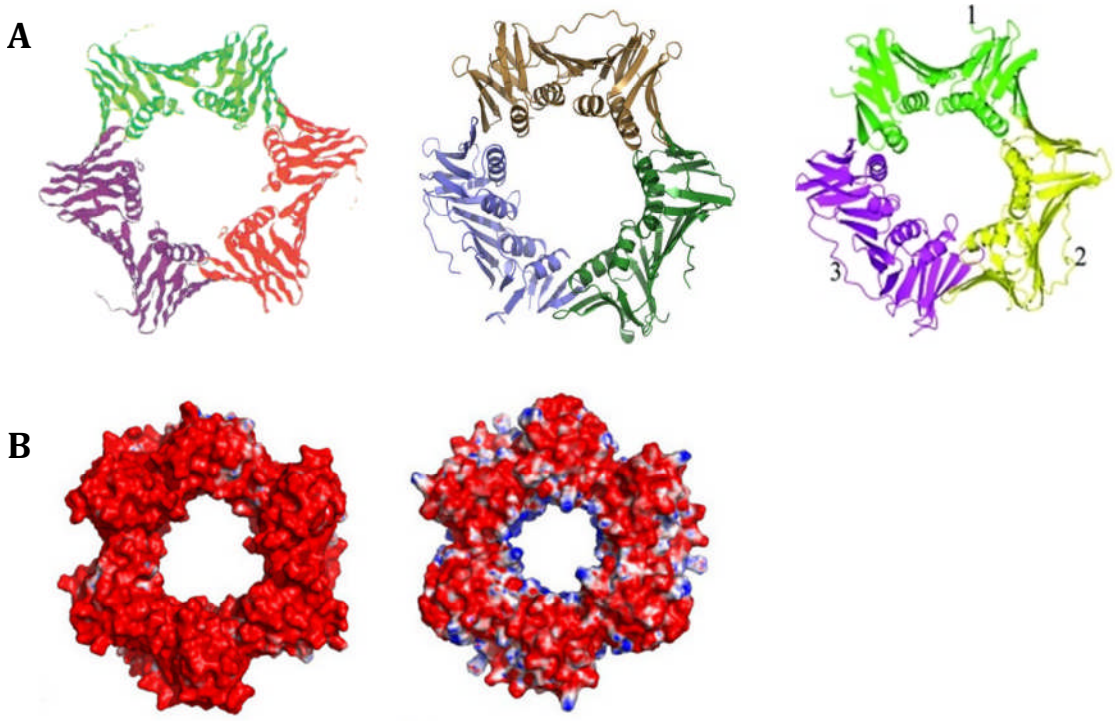


Figure 1.11 Structure of archaeal PCNA proteins

A Comparison of the PCNA structures from the three archaeal species: *P.furiosus* (left), *H.volcanii* (middle) and *S.solfataricus* (right), adapted from (Matsumiya et al., 2001, Williams et al., 2006, Winter et al., 2009)

B Surface charge distribution of PCNA proteins from *P.furiosus* (left) and *H.volcanii* (right). The accessible surface area is colored according to the electrostatic potential, adapted from (Winter et al., 2009)

1.2.2.4 Replication Factor C, the clamp loader

In bacteria and eukaryotes, PCNA is not capable of spontaneous assembly onto DNA but it requires a conserved, chaperone-like loader called RFC (Replication Factor C) (Majka and Burgers, 2004). In eukaryotes, RFC is a complex consisting of five

subunits, one large and four small. All five subunits are essential for protein function and they share significant similarity in seven regions based on the amino acid sequence. The large subunit also contains an additional domain on the N-terminus (Cullmann et al., 1995). RFC loads PCNA onto DNA in an orientation-dependent manner. It recognises the template-primer DNA sequence, binds PCNA and positions it toward the 3' end of the elongating DNA. This ensures that DNA polymerase, tethered to DNA by PCNA, is also orientated toward to growing end (Moldovan et al., 2007). RFC functioning as a clamp loader requires ATP binding. The RFC subunits belong to the AAA+ family of ATPases that bind ATP at the junction of two subunits. ATP binding, but not hydrolysis, is needed to open PCNA ring to a structure that can bind DNA (Seybert and Wigley, 2004). As shown by fluorescence resonance energy transfer in yeast, PCNA is initially open about 34 Å in the plane of the ring. Since a duplex DNA has a diameter of 20 Å, the gap in PCNA ring is wide enough to encircle the DNA. When the loading reaction progresses, the PCNA is held open with a gap of about 5 Å. The final step involves an in-plane closing of the PCNA ring around DNA (Zhuang et al., 2006).

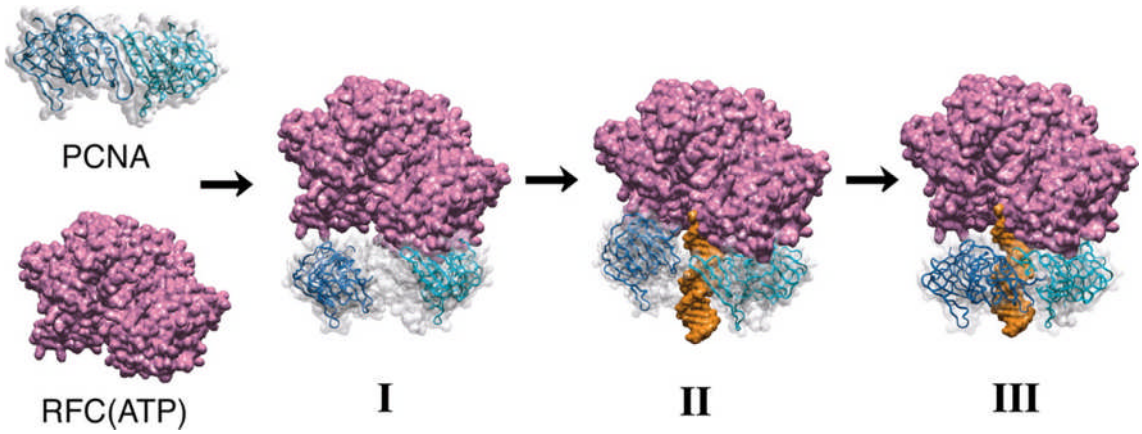


Figure 1.12 The clamp loading process, from (Zhuang et al., 2006)

I In-plane opening PCNA complex with RFC
II Out-of-plane closing of PCNA onto duplex DNA
III Final in-plane closing of PCNA onto DNA

Each of the complete archaeal genomes encodes at least a pair of RFC proteins, one large and one small. RFC complex from *S.solfataricus* is a pentamer comprised of one large and four small subunits.

In bacteria and eukaryotes, the presence of RFC is essential for PCNA loading onto DNA. Surprisingly, the PCNA proteins from both Crenarchaea and Euryarchaea can

self-assemble onto DNA. This self-loading phenomenon can be explained by weaker interaction between the subunit in a trimer, which might cause spontaneous opening of PCNA (Hingorani and O'Donnell, 2000, Williams et al., 2006). Although archaeal PCNA does not require RFC activity, it was shown that assembly of PCNA is much more efficient the presence of a clamp loader (Cann et al., 2001, Kelman and Hurwitz, 2000, Kelman et al., 1999, Pisani et al., 2000).

1.2.2.5 Topoisomerases

DNA topoisomerases are enzymes that regulate the degree of genomic DNA supercoiling. All topoisomerases catalyse changes in the linkage of DNA strands or helices by a conserved mechanism of transient DNA strand cleavage and relegation. Those enzymes are divided into two families: type I cleaves and reseals one strand of duplex DNA in the absence of ATP and type II cleaves and re-ligates both DNA strands in the presence of ATP (Koster et al., 2010).

DNA replication, similarly to translation or recombination, impacts DNA topology. As the DNA duplex is unwound by MCM helicase, a replisome will accumulate positive supercoils in front of the replication fork. At the same time negative supercoils are produced behind it. Topoisomerases have the essential role to relieve the super-helical strains (Koster et al., 2010, Postow et al., 2001, Rampakakis et al., 2010).

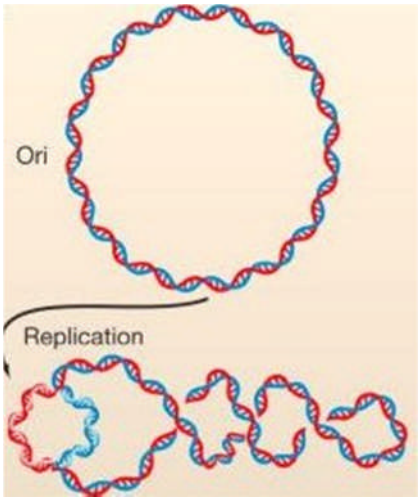


Figure 1.13 DNA topology and its relevance in replication, adapted from (Koster et al., 2010)

When a circular DNA is replicated, DNA at the origin is unwound by replicative helicase. By conservation of linking number, this generates positive supercoils ahead of the forks

The crystal structures of three archaeal topoisomerases have been determined: type I enzymes from *A.fulgidus* (Rodriguez and Stock, 2002) and *S.tokodaii* (Matoba et al., 2002) and type II enzyme from *M.jannaschii* (Nichols et al., 1999). The topoisomerases from *A.fulgidus* and *S.tokodaii* are reverse gyrases, enzymes found

only in thermophilic organisms. As many archaea live under extreme environmental conditions that are challenging for genome integrity, topoisomerases might have additional functions in stabilizing DNA.

1.2.3 Okazaki fragment maturation

Whereas the leading DNA strand is synthesised in a continous manner, the lagging strand is synthesised in short fragments known as Okazaki fragments. Those fragments are joined together to create a mature duplex DNA. The major proteins involved in this process are flap endonuclease1 (Fen-1), RNaseH, and DNA ligase.

1.2.3.1 Flap Endonuclease 1

Fen-1 (Flap Endonuclease 1) is a structure-specific nuclease, conserved in eukaryotes and archaea. Its enzymatic activity is removing 5'-overhanging flaps and the RNA/DNA primer during maturation of the Okazaki fragment. All archaeal species with genome sequence available have clear Fen-1 homologous. Biochemical studies on proteins from *M.jannaschii* (Hosfield et al., 1998a, Rao et al., 1998) and *P.horikoshii* (Matsui et al., 1999) provide information about Fen-1's enzymatic properties *in vitro,* which seem to be similar to its eukaryotic counterparts. Also, like eukaryotic Fen-1, enzymes form *S.solfataricus* and *P.furiosus* interact with PCNA via the PIP motif (Dionne et al., 2003, Hosfield et al., 1998b). Crystal structures for archaeal Fen-1 proteins are also available (Hwang et al., 1998, Matsui et al., 1999, Mase et al., 2011). *P.furiosus* Fen-1 consists of a single globular domain with a deep, positively charged groove and a protruding C-terminal region that extends beyond the body of the enzyme. The catalytic site of Fen-1 is located within the groove (Matsui et al., 1999).

1.2.3.2 RNase H

RNase H is an enzyme that specifically degrades the RNA strand of an RNA-DNA hybrid making it essential protein in the Okazaki fragment maturation process. Eukaryotic cells encode two classes of RNase H, type 1 (RNase H1) and 2 (RNase H2) (Ohtani et al., 1999). Although they have low similarity at the amino acid sequence level, both types use the same catalytic mechanism to cleave at the 5'-end of RNA phosphodiester bonds (Chapados et al., 2001). RNase H2 enzyme has the unique ability to hydrolyse single ribonucleotides embedded in a DNA duplex and probably is a replicative RNase. Archaea were postulated to encode homologous of RNase H2 only, however, RNase H type 1 was identified in *Halobacterium* sp. NRC-

1 has been shown to cleave an RNA-DNA junction (Ohtani et al., 2004). Studies with RNase H2 enzymes from *P.kodakaraensis*, *A.fulgidus* and *M.jannaschii* revealed structural and functional similarities to their bacterial and eukaryotic counterparts (Haruki et al., 1998, Chai et al., 2001, Lai et al., 2000). RNase H2 from *A.fulgidus* is a single protein with the PIP motif and catalytic centre (in opposition to eukaryotic heterotrimeric RNase H2 complex). An archaeal enzyme was shown to interact with PCNA, but the physiological role of that interaction is unknown (Meslet-Cladiere et al., 2007). Recently, the crystal structure of RNaseH2-PCNA complex from *A.fulgidus* was solved (Bubeck et al., 2011). PCNA stimulates RNase H2 activity on ribonucleotides embedded within DNA and on Okazaki-like substrates (Bubeck et al., 2011).

1.2.3.3 DNA Ligase

DNA ligase is an enzyme that catalyses formation of a phosphodiester bond between adjacent 5' phosphate and 3' hydroxyl termini in DNA molecules. Depending on the cofactor required for activity, DNA ligases have been divided into two groups: ATP- and NAD^+- dependent (Martin and MacNeill, 2002, Wilkinson et al., 2001). Apart from the difference in cofactor requirement, the reactions catalysed by the two classes of those two enzymes are identical. First, an AMP moiety derived from the cofactor is covalently attached to a conserved lysine residue within the K*x*DG motif. Then, AMP is transferred from the adenylated enzyme intermediate to the free 5′-phosphoryl group at a nicked site of duplex DNA. In the final step, the AMP group is released from the adenylated DNA intermediate as the phosphodiester bond is formed (Lehman, 1974). ATP-dependent ligases are widespread in evolution, whereas NAD^+- dependent enzymes are found mostly in bacteria. All Archaea possess ATP-dependent ligase (LigA NAD^+- dependent ligase); several of these have been characterised biochemically (Nakatani et al., 2002, Lai et al., 2002, Rolland et al., 2004). In addition, it was found that *H.volcanii* genome encodes bacterial-like NAD^+- dependent ligase (LigN) (Zhao et al., 2006). Phylogenetic analyses show that LigN was acquired by lateral gene transfer, most likely from a δ-proteobacterium. LigA and LigN, individually, are non-essential for *H.volcanii* cell viability, but double deletion is not possible indicating that the two enzymes share an essential function (Zhao et al., 2006). The crystal structures of archaeal ATP-dependent ligases from *P.furiosus* (Nishida et al., 2006), *S.solfataricus* (Pascal et al., 2006) and, recently, *A.fulgidus* (Kim et al., 2009) have been solved showing the presence of

three distinct domains: a DNA-binding domain, an adenylation binding domain and an OB-fold domain. Although these domains share structural similarities between species, the enzymes displayed different conformations in crystal structures (Figure 1.14). *S.solfataricus* DNA ligase adopted an extended highly opened conformation, whereas *P.furiosus* and *A.fulgidus* enzymes adopted a closed conformation (the structure in former species is even more closed). Those structural variations confirm conformational flexibility that is believed to be essential for enzyme function. The action of multi-domain DNA ligases predicts large conformational changes associated with domain rearrangements (Lee et al., 2000).

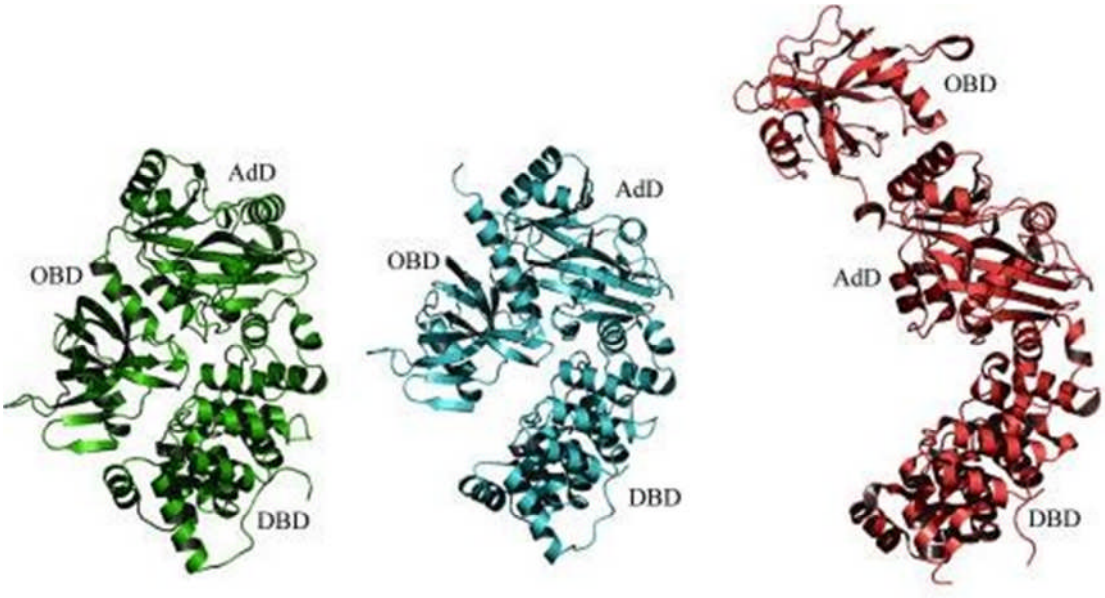


Figure 1.14 Different conformations of archaeal ATP-dependent ligases, adapted from (Kim et al., 2009)

Ribbon diagrams of *A.fulgidus* (left), *P.furiosus* (centre) and *S.solfataricus* (Albers et al.) DNA ligases. DBD, OBD and AdA refer to a DNA-binding domain, an adenylation binding domain and an OB-fold domain, respectively. *A.fulgidus* ligase displays the most closed conformation, with water-mediated hydrogen-bond interactions between the two terminal domains DBD and OBD.

1.2.4 Halophiles as a model for archaeal genetics

Halophiles are heterotrophic, aerobic archaea belonging to *Euryarchaeota* that inhabit the most saline environments on Earth like solar salterns and natural salt lakes. Phylogenetically Haloarchaea might have evolved from methanogens by acquiring bacterial genes for aerobic respiration by the lateral gene transfer (Boucher et al., 2003).

Halophilic archaea maintain an osmotic balance with their medium by accumulating inorganic ions, mainly potassium, in the cytoplasm. Potassium is used because it attracts less water than sodium. This strategy is called the "salt-in" strategy. In contrast, halophilic bacteria (with exception of *Salinibacter ruber*) favor the „salt-out" strategy based on excluding salt from cells and using organic compounds like glycerol and glycine betaine to maintain an osmotic balance (Oren, 2008). This strategy is energetically costly and less suitable at high salt concentration, which explains why archaea predominate in hyper-saline niches (Oren, 1999). The "salt-in" strategy used by Archaea requires several structural adaptations allowing proteins to function in molar salt concentrations when mesohalic proteins are salted out. The haloarchaeal proteins have reduced overall hydrophobicity by replacing large hydrophobic residues on the protein surface with small hydrophilic residues. Another adaptation is an increase in acidic residues. A high density of negative charges coordinates a network of hydrated cations, which maintain the protein in solution (Madern et al., 2000, Mevarech et al., 2000). Finally, some proteins posses an additional damain not present in mesohalic counterparts (Marg et al., 2005).

Haloarchaeal genomes usually consist of one main chromosome and a number of mega-plasmids (Soppa et al., 2008). Number of genome copies varies from 15 to 30 per cell in exponential phase (ploidy is down-regulated in stationary phase) (Breuert et al., 2006). A high G+C content (around 65%) is another signature of haloarchaea. The one known exception is *Hqr.walsbyi* (45% G+C) (Bolhuis et al., 2006). The high G+C content might be linked to the acidic proteome of haloarchaea or might be a way to avoid insertion sequence (IS) elements that target A+T-rich sequences (Hartman et al., 2010, Brugger et al., 2002).

Haloarchaea are particularly suitable for genetic studies. They are easy to cultivate in the laboratory as they grow aerobically at moderate temperature, have a reasonably fast generation time, are resistant to contamination by non-halophilic microorganisms and most importanty, tools for their genetic manipulation have been developed. For years, halophiles were the only archaeal organism that could be transformed with foreign DNA, based on the polyethylene glycol- mediated method described in 1987 (Charlebois et al., 1987). To date, sixteen haloarchaeal genomes have been sequenced: *Halobacterium* sp. NRC-1, *Halobacterium* sp. R-1, *Haloarcula marismortui* ATCC 43049, *Netronomonas pharaonis* DSM 2160, *Haloquadratum walsbyi* DSM 16790, *Haloferax volcanii* DS2, *Halorubrum lacusprofundi* ATCC 49239, *Halomicrobium mukohataei* DSM 12286, *Halorhabdus*

utahensis DSM 12940, *Halogeometricum borinquense* DSM 11551, *Haloterrigena turkmenica* DSM 5511, *Natrialba magadii* ATCC 43099, *Halalkalicoccus jeotgali* B3, *Haloarcula hispanica* ATCC 33960, *Halopiger xanaduensis* SH6 and *Haloferax mediterranei*. To improve data access and facilitate functional genomic studies on haloarchaea, a dedicated database and website, named HaloWeb, was developed. It incorporates all finished genomes, including gene, protein and RNA sequences and annotation data (Dassarma et al., 2010).

Within Haloarchaea, two species have been developed into model organisms for experimental studies: *Hbt.salinarum* and *H.volcanii*. *Halobacterium* is an extreme halophile with optimal salt concentration 2.5-4M whereas *Haloferax* is a moderate halophile (optimal salt concentration 1.7-2.5). *Halobacterium* was isolated first and become the traditional choice for cell and systems biology (DasSarma et al., 2006). The genome of *Halobacterium salinarum* NRC-1 was published in 2000 (Ng et al., 2000) and in 2008 a second annotation was published for *Halobacterium salinarum* R1 (Pfeiffer et al., 2008). They are essential the same species with a structural differences in mega-plasmids. *Hbt.salinarum* is less suitable for genetic studies as it grows slowly, its genome is unstable due to frequent IS-mediated rearrangements and the list of selectable markers is limited. Nowadays, the best haloarchaeal organism for molecular genetics is *H.volcanii*. It will be characterized in grater details in the next section as all the laboratory work described in this thesis was done on *Haloferax.*

1.3 Working with *Haloferax volcanii*

Haloferax volcanii was isolated from bottom sediment of the Dead Sea in 1975 (Mullakhanbhai and Larsen, 1975). It was initially known as *Halobacterium volcanii,* where "*volcanii*" refers to Benjamin Elazari Volcanii who first demonstrated the existence of microbiological life in hyper-saline environment. *H.volcanii* is a genetically stable prototroph that can be cultivated aerobically in complex and minimal media using a single carbon source. The growth optimal conditions are temperature 45^{o}C and salt (sodium chloride) concentration about 2M.

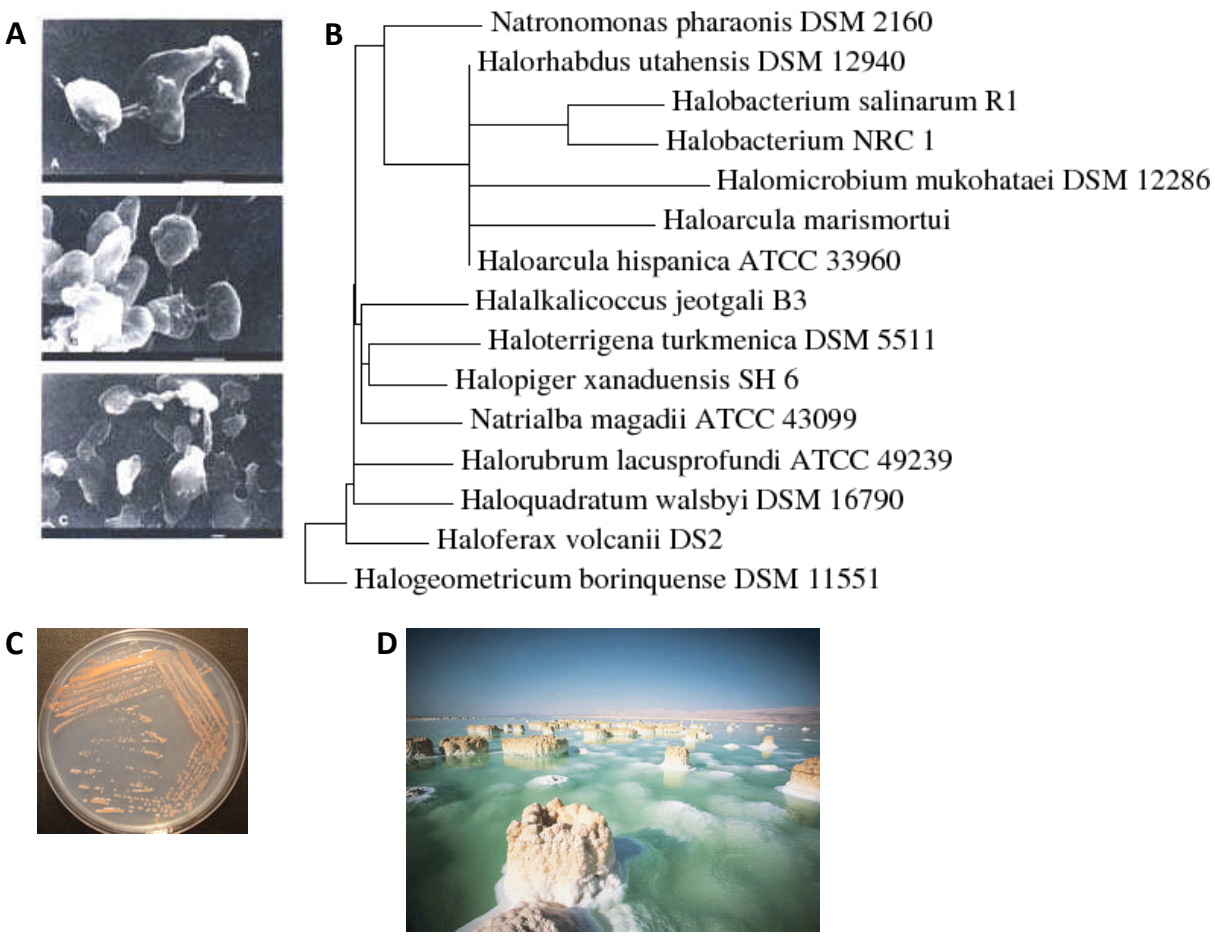


Figure 1.15 Properties of *H.volcanii*

A Microscopy image of *H.volcanii* cells. Image adapted from Prof. Moshe Mevarech, Tel Aviv University.
B Phylogenetic tree of related species based on multiple genome alignment.
C Growth of *H.volcanii* on solid medium.
D Columns of salt rise from extremely saline waters of the Dead Sea, natural habitat of *H.volcanii*, adapted from http://www.photoweeklyonline.com

Haloferax has a long history of genome research going back to 1991, when a physical map of overlapping genomic clones was published (Charlebois et al., 1991). The full genome sequence of *H.vocanii* DS2 was published in 2010 (Hartman et al., 2010). The genome is composed of five circular genetic elements: a 2.848 Mb main chromosome, three smaller chromosomes pHV4, pHV3 and pHV1 and one plasmid pHV2. The total genome size is 4 Mb. The distinction between which elements are chromosomes and which are plasmids is still under debate. Here, following (Hartman et al., 2010), the four largest elements are termed “chromosomes” as they all appear to replicate using chromosome-like replication mechanism (Norais et al., 2007a). The sizes and other feature of particular genetic elements are listed below.

	Main chromosome	pHV4	pHV3	pHV1	pHV2
Size (bp)	2 847 757	635 786	437 906	85 092	6 359
GC content (%)	66.64	61.67	65.56	55.50	56.06
Gene content	3021	635	380	88	6
NCBI Ref.Seq. Accession	NC_013967	NC_013966	NC_013964	NC_013967	NC_013965

The average genomic G+C content is 65%. In total, 4063 predicted proteins were identified in the genome. Of those, 5.5% have no sequence similarities to any proteins in any publicly available complete genome sequences (Hartman et al., 2010). *Haloferax* has two origins of replication on its main chromosome, *oriC*1 and *oriC*2. Chromosome copy number varies from 18 in exponential phase to 15 in stationary phase (Soppa, 2011).

What makes *H.volcanii* so suitable for genetics is development of all tools required for that kind of study. PEG-mediated transformation protocol allows efficient transformation with foreign DNA. This method can only be used for a species without a rigid cell wall, for which it is easy to generate spheroplasts. The cell surface of Haloarchaea is made of paracrystalline glycoprotein layer, called the S layer. Its stability requires magnesium ions. Treatment with EDTA removes Mg^{2+} from the medium, what causes spheroplast formation in which DNA can be introduced. Incubation of cells in rich medium in presence of high concentration of Mg^{2+} allows regeneration the cells (Charlebois et al., 1987). The pop-in/pop-out method for gene deletion (Bitan-Banin et al., 2003), selectable markers (Allers et al., 2004), reporter genes (Holmes and Dyall-Smith, 2000), an inducible promoter (Large et al., 2007) and system for protein overexpression (Allers et al., 2010) have been developed in the last decade. For selection purposes auxotrophic selectable markers are commonly used, specifically genes for amino acid or nucleotide biosynthesis: *pyrE2* for uracil, *trpA* for tryptophan and *hdrB* for thymidine biosynthesis (Bitan-Banin et al., 2003). Combining auxotrophic markers allows construction of multiply deleted/mutated strains.

1.4 Transcription initiation and promoter structure in Archaea

To survive in a competitive environment, organisms need to tightly control production of proteins depending of actual requirements for cell activity and avoid extraneous, energetically costly production. Regulation of transcription initiation is one of the most important steps governing gene expression.

In all three domains of life, the transcriptional machinery consists of transcription factors and RNA polymerase (RNAP) that are recruited to promoter DNA and form pre-initiation complex. The first step in initiation process is binding of the TATA-box binding protein (TBP) to the AT-rich TATA-box promoter sequence located approximately 25 bp upstream of transcription start site. That causes DNA bend and distortion of DNA backbone that allows other transcriptional factors to binding and finally, position single-stranded template DNA near the RNAP active site (Hampsey, 1998). In bacteria, σ factor binds to RNAP to form holoenzyme, which recognises promoter sequence and makes a closed RNAP-promoter complex (Dove et al., 2003, Hampsey, 1998). In contrast to bacterial holoenzyme, eukaryotic RNAP is not capable of sequence specific promoter recognition and requires the presence of additional factors to properly position near the transcription start site. Those factors, in addition to TBP, are: TFIIB, TFIID, TFIIE, TFIIF and TFIIH (Hampsey, 1998).

In Archaea, transcription apparatus, including RNA polymerase and general transcription factors, is similar to the eukaryotic RNAP II apparatus. (Jun et al., 2011). The Archaeal pre-initiation system includes TBP, TFB and TFE, which are orthologous of TBP, TFIIB, and TFIIEα, respectively (Bell and Jackson, 1998). (Figure 1.16)

Archaeal TBPs have approximately 30-40% sequence identity to the C-terminal domain of eukaryotic TBPs and close structural similarity (Soppa and Link, 1997, Kosa et al., 1997). In both domains of life, the C-terminal domain is composed of an imperfect direct repeat. Eukaryotic proteins posses a variable N-terminal domain not found in archaeal molecules. Instead, archaea have unique, 6-10 amino acids long, highly acidic, C-terminal tail (Thomsen et al., 2001).

Archaeal TFB is an orthologue of eukaryotic TFIIB. The C-terminal core region of TFB, composed of two domains, each with five or six α-helices, binds to the B-factor Recognition Element (BRE) (see below), whereas the N-terminal region, which forms a zinc ribbon motif, recruits RNAP and determines the approximate position

of the enzyme catalytic centre with respect to the TATA-box (Littlefield et al., 1999, Zhu et al., 1996). The interaction between TFB and BRE appears to be necessary to determine direction of transcription (Bell et al., 1999). *In vitro* studies in *M.thermautotrophicus* shows also that once 9-12 bases of the nascent transcript are formed, TFB is released from the transcription initiation complex and recycled for the next round of transcription (Xie and Reeve, 2004). TFB release is also required to open the channel for RNA and allow its exit from RNAP (Kostrewa et al., 2009).

The third factor found in Archaea is TFE, an orthologue of the N-terminal domain of the α-subunit of eukaryotic TFIIE (Bell and Jackson, 2001). It has been suggested that TFE stabilizes the RNAP open complex formation by enhancing melting DNA (Naji et al., 2007); transcription from strong promoters, however, is not dependent on TFE (Bell et al., 2001).

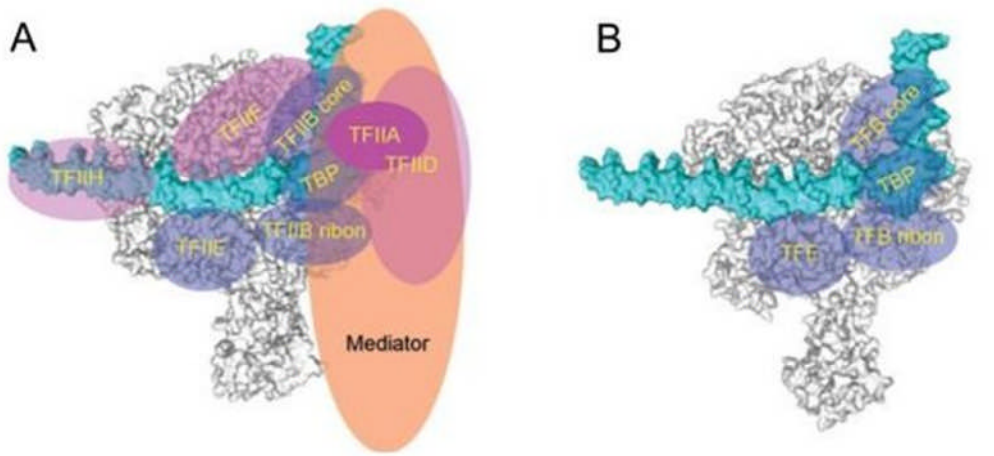


Figure 1.16 Comparison of the transcription initiation complex in Eukarya (A) and Archaea (B)

RNAP (grey) and promoter DNA (cyan) are represented as surface and transcription factors are represented as ellipses. Transcription factors found in Eukarya and Archaea are colored transparent blue, and unique for Eukarya- transparent purple. Adapted from (Jun et al., 2011).

When TBP/TFB/RNAP pre-initiation complex assembles, open complex formation and transcription initiation on linear DNA template can finally take place. It is important to mention that several archaeal genomes encode multiple genes for TBP and/or TFB; *Hbt.salinarum* sp. NRC-1 for example posses six putative TBPs and seven putative TFBs (Ng et al., 2000). The presence of multiple proteins might have a role in gene expression regulation. Intriguingly, there is a tendency within archaea that halophiles, hyperthermophiles and thermoacidophiles have multiple TFBs, whereas methanogens have multiple TBPs. It is not clear if this tendency has any functional significance but considering the roles of TBP and TFB in DNA binding

and DNA opening respectively, it might indicate that for halophiles and hyperthermophiles the DNA opening is the crucial target for gene regulation whereas for methanogens, more important might be regulate DNA binding (Jun et al., 2011).

The structure of archaeal promoters is other feature showing similarity with eukaryotic RNAPII system. Archaeal promoters consist of a TATA-box, positioned approximately 25-30 bp upstream of transcription start site and a purine-rich region called the B-factor Recognition Element (BRE) (Figure 1.17) The TATA-box and BRE are bound by TBP and TFB, respectively. It was experimentally proven that the sequences of those elements directly determine efficient of transcription initiation (Palmer and Daniels, 1995, Gregor and Pfeifer, 2005, Reiter et al., 1990). The TATA-box sequence has different consensus in different phyla of Archaea: TTTTTAAA in Crenarchaea (predominantly in *Sulfolobus* genus), TTTATATA in methanogens and NTTTWWWN in halophiles (W means A or T; N means any nucleotide) (Reeve, 2003).

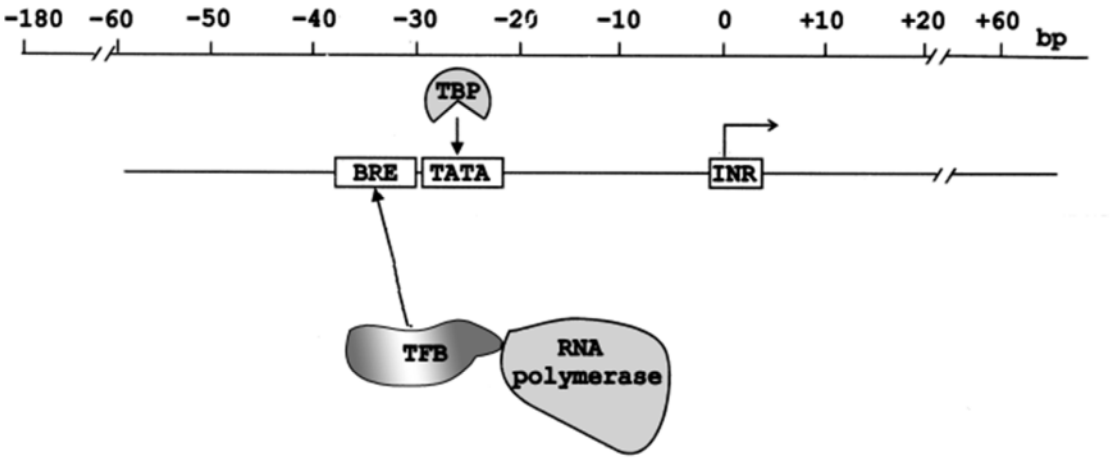


Figure 1.17 The structure of archaeal promoter region, adapted from (Reeve, 2003)

The core enzyme in transcription process is RNA polymerase (RNAP). In 2008, the crystal structure of RNAP from all three domains of life become available (Hirata et al., 2008, Zhang et al., 1999, Cramer et al., 2001). The overall shape of bacterial, archaeal and eukaryotic enzymes resembles a crab claw (Figure 1.18). The largest 2-3 subunits form the bulk of each claw-arm, generating a cleft for double-stranded DNA binding (Hirata et al., 2008, Zhang et al., 1999, Cramer et al., 2001). The structure around the cleft is highly conserved among all RNAPs indicating that transcription machinery have come from the same origin. However, the architecture

of the surface of the enzyme shows significant differences between Bacteria and Eukarya-Archea. The most apparent difference is that archaeal RNAP and all three types of eukaryotic RNAPs (RNAPI, RNAPII and RNAPIII) posses a protruding stalk-like structure that is not present in bacterial RNA polymerase (Armache et al., 2005). Archaeal RNAPs are homologous of all three eukaryotic polymerases, but they do not have subunits homologous to subunits that are present only in RNAP I and RNAPIII. Tremendous progress in understanding of the function of archaeal RNA polymerase was made by re-assembly twelve subunits *Methanococcus janashii* enzyme from individual recombinant proteins. It allowed confirmation of subunit interactions, identified subunits essential for catalysis and residues involved in magnesium ion binding (Werner and Weinzierl, 2002). When comparing the structure of archaeal and eukaryotic RNAP, one of the major differences between them is the absence of a C-terminal domain in the largest subunit of the archaeal enzyme. This domain, containing a heptapeptide-repeat, is a target of regulatory phosphorylation and the place of binding of many transcription factors interacting with RNAPII (Stiller and Hall, 2002, Woychik and Hampsey, 2002).

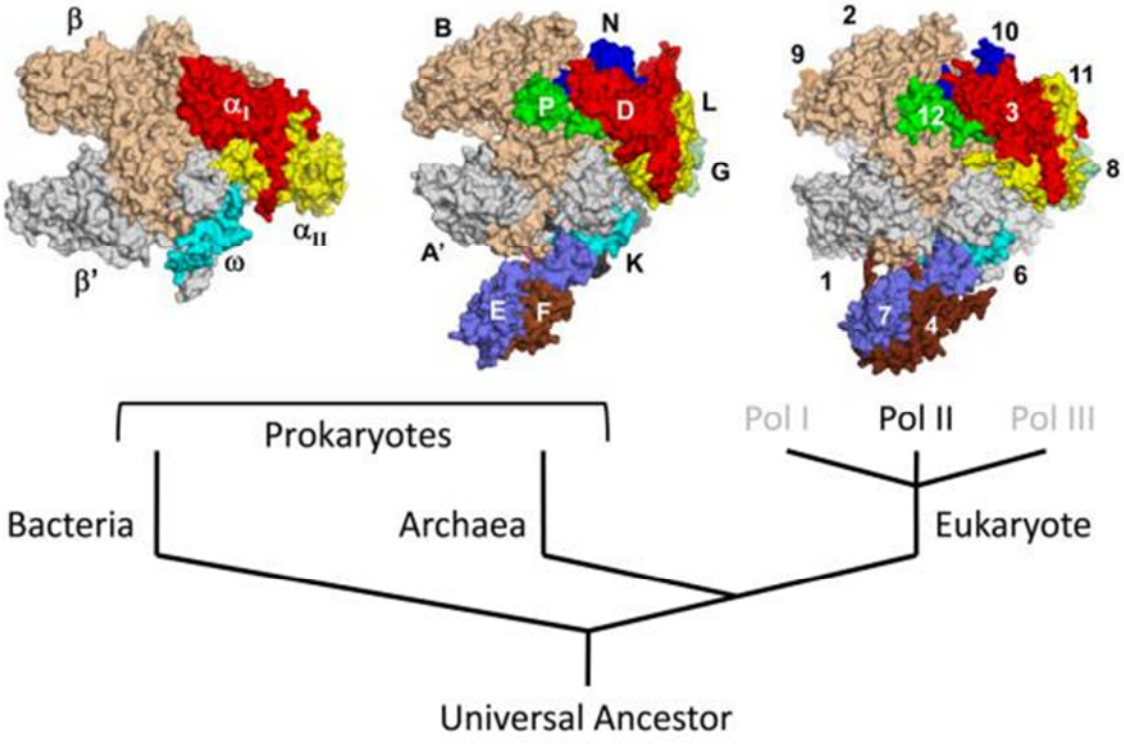


Figure 1.18 Overall structure of RNA polymerases from bacteria (left, *Thermus aquaticus* core enzyme), archaea (middle, *Sulfolobus solfataricus* RNAP) and eukaryotes (right, *Saccharomyces cerevisiae* RNAPII). Orthologous subunits are depicted with the same color. Adapted from (Jun et al., 2011)

1.5 Objectives

This thesis aims to contribute to a genetic and physiological characterisation of archaeal DNA replication using halophilic euryarchaeon *Haloferax volcanii* as a model organism. Because of similarities between the Eukaryal and Archaeal DNA replication machinery, this research can be used to increase our understanding of DNA replication in eukaryotic cells.

The tryptophan- regulated *tna* promoter was used to generate conditional- lethal (promoter shut- off) mutants of replication genes, and these strains were used to investigate the *in vivo* functions of the corresponding gene products.

Chapter 2

Materials and Methods

2.1 Materials

2.1.1 *Haloferax volcanii* strains

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

Strain	Genotype	Notes
DS70	Wild-type strain cured of plasmid pHV2	(Wendoloski et al., 2001)
H26	Δ*pyrE2*	(Allers et al., 2004)
H53	Δ*pyrE2* Δ*trpA*	(Allers et al., 2004)
H98	Δ*pyrE2* Δ*hdrB*	(Allers et al., 2004)
H99	Δ*pyrE2* Δ*hdrB* Δ*trpA*	(Allers et al., 2004)
pNPM-tna- plasmid integrant strains (pop-in)		
SMH719	*polB*:[pNPM-tna-PolB] $pyrE2^+$ $hdrB^+$	H98, integration of PL9 at *polB* locus
SMH720	*polD1*:[pNPM-tna-PolD1] $pyrE2^+$ $hdrB^+$	H98, integration of PL10 at *polD1* locus
SMH721	*polD2*:[pNPM-tna-PolD2] $pyrE2^+$ $hdrB^+$	H98, integration of PL11 at *polD1* locus
SMH722	*pcnA*:[pNPM-tna-PcnA] $pyrE2^+$ $hdrB^+$	H98, integration of PL12 at *polD2* locus
SMH723	*mcm*:[pNPM-tna-Mcm] $pyrE2^+$ $hdrB^+$	H98, integration of PL13 at *mcm* locus
SMH724	*rfcA*:[pNPM-tna-RfcA] $pyrE2^+$ $hdrB^+$	H98, integration of PL14 at *rfcA* locus
SMH725	*rfcB*:[pNPM-tna-RfcB] $pyrE2^+$ $hdrB^+$	H98, integration of PL15 at *rfcB* locus
SMH726	*rfcC*:[pNPM-tna-RfcC] $pyrE2^+$ $hdrB^+$	H98, integration of PL16 at *rfcC* locus
SMH727	*ligA*:[pNPM-tna-LigA] $pyrE2^+$ $hdrB^+$	H98, integration of PL17 at *ligA* locus
SMH728	*rpaA*:[pNPM-tna-RpaA] $pyrE2^+$ $hdrB^+$	H98, integration of PL18 at *rpaA* locus
SMH729	*rpaB*:[pNPM-tna-RpaB] $pyrE2^+$ $hdrB^+$	H98, integration of PL19 at *rpaB* locus
SMH730	*rpaC*:[pNPM-tna-RpaC] $pyrE2^+$ $hdrB^+$	H98, integration of PL20 at *rpaC* locus
SMH731	*priS*:[pNPM-tna-PriS] $pyrE2^+$ $hdrB^+$	H98, integration of PL21 at *priS* locus
SMH732	*priL*:[pNPM-tna-PriL] $pyrE2^+$ $hdrB^+$	H98, integration of PL22 at *priL* locus
SMH200	*polB*:[pNPM02-PolB] $pyrE2^+$ $hdrB^+$	H98, integration of PL33 at *polB* locus

SMH201	*polB*:[pNPM03-PolB] *pyrE2*$^{+}$ *hdrB*$^{+}$	H98, integration of PL36 at *polB* locus
SMH202	*polB*:[pNPM04-PolB] *pyrE2*$^{+}$ *hdrB*$^{+}$	H98, integration of PL39 at *polB* locus
SMH203	*polB*:[pNPM05-PolB] *pyrE2*$^{+}$ *hdrB*$^{+}$	H98, integration of PL42 at *polB* locus
SMH204	*polB*:[pNPM-L11-PolB] *pyrE2*$^{+}$ *hdrB*$^{+}$	H98, integration of PL42 at *polB* locus
SMH205	*polB*:[pNPM-M1tna-PolB] *pyrE2*$^{+}$ *hdrB*$^{+}$	H98, integration of PL45 at *polB* locus
SMH206	*polB*:[pNPM-M2tna-PolB] *pyrE2*$^{+}$ *hdrB*$^{+}$	H98, integration of PL46 at *polB* locus
SMH207	*polB*:[pNPM-M3tna-PolB] *pyrE2*$^{+}$ *hdrB*$^{+}$	H98, integration of PL47 at *polB* locus
SMH208	*polD1*:[pNPM02-PolD1] *pyrE2*$^{+}$ *hdrB*$^{+}$	H98, integration of PL34 at *polD1* locus
SMH209	*polD1*:[pNPM03-PolD1] *pyrE2*$^{+}$ *hdrB*$^{+}$	H98, integration of PL37 at *polD1* locus
SMH210	*polD1*:[pNPM04-PolD1] *pyrE2*$^{+}$ *hdrB*$^{+}$	H98, integration of PL40 at *polB* locus
SMH211	*polD1*:[pNPM05-PolD1] *pyrE2*$^{+}$ *hdrB*$^{+}$	H98, integration of PL43 at *polB* locus
SMH212	*polD1*:[pNPM-L11-PolD1] *pyrE2*$^{+}$ *hdrB*$^{+}$	H98, integration of PL43 at *polB* locus
SMH213	*polD1*:[pNPM-M1tna-PolD1] *pyrE2*$^{+}$ *hdrB*$^{+}$	H98, integration of PL48 at *polB* locus
SMH214	*polD1*:[pNPM-M2tna-PolD1] *pyrE2*$^{+}$ *hdrB*$^{+}$	H98, integration of PL49 at *polB* locus
SMH215	*polD1*:[pNPM-M3tna-PolD1] *pyrE2*$^{+}$ *hdrB*$^{+}$	H98, integration of PL50 at *polB* locus
SMH216	*mcm*:[pNPM02-Mcm] *pyrE2*$^{+}$ *hdrB*$^{+}$	H98, integration of PL35 at *mcm* locus
SMH217	*mcm*:[pNPM03-Mcm] *pyrE2*$^{+}$ *hdrB*$^{+}$	H98, integration of PL38 at *mcm* locus
SMH218	*mcm*:[pNPM04-Mcm] *pyrE2*$^{+}$ *hdrB*$^{+}$	H98, integration of PL41 at *mcm* locus
SMH219	*mcm*:[pNPM05-Mcm] *pyrE2*$^{+}$ *hdrB*$^{+}$	H98, integration of PL44 at *mcm* locus
SMH220	*mcm*:[pNPM-L11-Mcm] *pyrE2*$^{+}$ *hdrB*$^{+}$	H98, integration of PL44 at *mcm* locus
SMH221	*mcm*:[pNPM-M1tna-Mcm] *pyrE2*$^{+}$ *hdrB*$^{+}$	H98, integration of PL51 at *mcm* locus
SMH222	*mcm*:[pNPM-M2tna-Mcm] *pyrE2*$^{+}$ *hdrB*$^{+}$	H98, integration of PL52 at *mcm* locus
SMH223	*mcm*:[pNPM-M3tna-Mcm] *pyrE2*$^{+}$ *hdrB*$^{+}$	H98, integration of PL53 at *mcm* locus
SMH224	*polD2*:[pNPM-M3tna-PolD2] *pyrE2*$^{+}$ *hdrB*$^{+}$	H98, integration of PL56 at *polD2* locus
SMH225	*rpaC*:[pNPM-M3tna-rpaC] *pyrE2*$^{+}$ *hdrB*$^{+}$	H98, integration of PL57 at *rpaC* locus
pNPM-tna- promoter replacement strains (pop-out)		
SMH737	*polB*::*tna-polB* Δ*pyrE2* Δ*hdrB*	Pop-out of PL23, leaving *polB* under *tnaA*-promoter
SMH738	*rpaC*::*tna-rpaC* Δ*pyrE2* Δ*hdrB*	Pop-out of PL24, leaving *rpaC* under *tnaA*-promoter
SMH739	*priS*::*tna-priS* Δ*pyrE2* Δ*hdrB*	Pop-out of PL25, leaving *priS* under *tnaA*-promoter
Strains for constitutive high- level expression of Rpa proteins		

SMH763	*rpaA*:[pNPM-fdx-RpaA] *pyrE2*$^{+}$ *hdrB*$^{+}$	H98, integration of PL26 at *rpaA* locus
SMH764	*rpaB*:[pNPM-fdx-RpaB] *pyrE2*$^{+}$ *hdrB*$^{+}$	H98, integration of PL27 at *rpaB* locus
SMH765	*rpaC*:[pNPM-fdx-RpaC] *pyrE2*$^{+}$ *hdrB*$^{+}$	H98, integration of PL28 at *rpaC* locus
SMH789	*rpaA*:[pNPM-fdx-RpaA] *rpaC*::*tna- rpaC pyrE2*$^{+}$ *hdrB*$^{+}$	Tna-rpaC promoter replacement strain, integration of PL26 at *rpaA* locus
SMH790	*rpaB*:[pNPM-fdx-RpaB] *rpaC*:: P*tna- rpaC pyrE2*$^{+}$ *hdrB*$^{+}$	Tna-rpaC promoter replacement strain, integration of PL27 at *rpaB* locus
rpa gene deletion strains		
SMH787	*rpaA*::*trpA*$^{+}$ Δ*pyrE2* Δ*hdrB*	H99, PL59 pop-out, leaving *rpaA*::*trpA*$^{+}$
SMH788	*rpaB*::*hdrB*$^{+}$ Δ*pyrE2* Δ*trpA*	H99, PL61 pop-out, leaving *rpaB*::*hdrB*$^{+}$
SMH791	*rpaB*::[pNPM-tna-RpaB] *rpaA*::*trpA*$^{+}$ *pyrE2*$^{+}$ *hdrB*$^{+}$	*rpaA*::*trpA*$^{+}$, integration of PL21 at *rpaB* locus
Strains carrying non-integrated plasmids		
SMH751	*rpaC*::*tna-rpaC* [pTA230] Δ*hdrB*	*rpaC*::*tna-rpaC* carrying PL3
SMH223	*rpaC*::*tna-rpaC* [pTA230-Hfx RpaC] Δ*hdrB*	*rpaC*::*tna-rpaC* carrying PL64
SMH752	*rpaC*::*tna-rpaC* [pTA230-Hgm RpaC] Δ*hdrB*	*rpaC*::*tna-rpaC* carrying PL65
SMH753	*rpaC*::*tna-rpaC* [pTA230-Hqr RpaC] Δ*hdrB*	*rpaC*::*tna-rpaC* carrying PL66
SMH754	*rpaC*::*tna-rpaC* [pTA230- HgmRpaC-ΔNTD] Δ*hdrB*	*rpaC*::*tna-rpaC* carrying PL67
SMH755	*rpaC*::*tna-rpaC* [pTA230- HgmRpaC-ΔOB^{A}] Δ*hdrB*	*rpaC*::*tna-rpa*, carrying PL68
SMH756	*rpaC*::*tna-rpaC* [pTA230 HgmRpaC-ΔOB^{B}] Δ*hdrB*	*rpaC*::*tna-rpa*, carrying PL69
SMH757	*rpaC*::*tna-rpaC* [pTA230- HgmRpaC-ΔOB^{C}] Δ*hdrB*	*rpaC*::*tna-rpaC* carrying PL70
SMH758	*rpaC*::*tna-rpaC* [pTA230- HgmRpaC-ΔOB^{AB}] Δ*hdrB*	*rpaC*::*tna-rpaC* carrying PL71
SMH759	*rpaC*::*tna-rpaC* [pTA230- HgmRpaC-ΔOB^{BC}] Δ*hdrB*	*rpaC*::*tna-rpaC* carrying PL72
SMH760	*rpaC*::*tna-rpaC* [pTA230--HgmRpaC-ΔOB^{AC}] Δ*hdrB*	*rpaC*::*tna-rpaC* carrying PL73
SMH761	*rpaC*::*tna-rpaC* [pTA230- HgmRpaC-ΔOB^{ABC}] Δ*hdrB*	*rpaC*::*tna-rpaC* carrying PL74
SMH762	*rpaC*::*tna-rpaC* [pTA230- HgmRpaC-ΔCTD] Δ*hdrB*	*rpaC*::*tna-rpaC* carrying PL75
SMH226	*rpaC*::*tna-rpaC* [pTA230-Flag-HgmRpaC] Δ*hdrB*	*rpaC*:: *tna- rpaC* carrying PL76
SMH227	*rpaC*::*tna-rpaC* [pTA230-Flag-HgmRpaC-ΔNTD] Δ*hdrB*	*rpaC*::*tna-rpaC* carrying PL77
SMH228	*rpaC*::*tna-rpaC* [pTA230-Flag-HgmRpaC-ΔOB^{A}] Δ*hdrB*	*rpaC*::*tna-rpaC* carrying PL78
SMH229	*rpaC*::*tna-rpaC* [pTA230-Flag-HgmRpaC-ΔOB^{B}] Δ*hdrB*	*rpaC*::*tna-rpaC* carrying PL79
SMH230	*rpaC*::*tna-rpaC* [pTA230-Flag-HgmRpaC-ΔOB^{C}] Δ*hdrB*	*rpaC*::*tna-rpaC* carrying PL80
SMH231	*rpaC*::*tna-rpaC* [pTA230-Flag-HgmRpaC-ΔOB^{AB}] Δ*hdrB*	*rpaC*::*tna-rpaC* carrying PL81

SMH232	*rpaC*:: P*tna- rpaC* [pTA230-Flag-HgmRpaC-ΔOB^{BC}] Δ*hdrB*	*rpaC*::*tna-rpaC* carrying PL82
SMH233	*rpaC*:: P*tna- rpaC* [pTA230-Flag-HgmRpaC-ΔOB^{AC}] Δ*hdrB*	*rpaC*::*tna-rpaC* carrying PL83
SMH234	*rpaC*::*tna-rpaC* [pTA230-Flag-HgmRpaC-ΔOB^{ABC}] Δ*hdrB*	*rpaC*::*tna-rpaC* carrying PL84
SMH235	*rpaC*::*tna-rpaC* [pTA230-Flag-HgmRpaC-ΔCTD] Δ*hdrB*	*rpaC*::*tna-rpaC* carrying PL85
SMH798	H26 [pTA233-CCBD] *hdrB*+	H26 carrying PL86
SMH799	H26 [pTA233-HfxRpaC-CCBD] *hdrB*+	H26 carrying PL87
SMH800	H26 [pTA233-HfxRpaC-NTD-CCBD] *hdrB*+	H26 carrying PL88
SMH234	H26 [pTA233-NCBD] *hdrB*+	H26 carrying PL89
SMH801	H26 [pTA233-NCBD-HfxPriS CTD] *hdrB*+	H26 carrying PL90
SMH767	*priS*::*tna-priS* [pTA230] *pyrE2*+ Δ*hdrB*	*priS*::*tna-priS* carrying PL3
SMH768	*priS*::*tna-priS* [pTA230-HfxPriS-GinS] *pyrE2*+ Δ*hdrB*	*priS*::*tna-priS* carrying PL91
SMH769	*priS*::*tna-priS* [pTA230-HfxGinS] *pyrE2*+ Δ*hdrB*	*priS*::*tna-priS* carrying PL92
SMH770	*priS*::*tna-priS* [pTA230-HgmPriSGinS] *pyrE2*+ Δ*hdrB*	*priS*::*tna-priS* carrying PL93
SMH771	*priS*::*tna-priS* [pTA230-HgmPriS] *pyrE2*+ Δ*hdrB*	*priS*::*tna-priS* carrying PL94
SMH772	*priS*::*tna-priS* [pTA230-HfxGinS] *pyrE2*+ Δ*hdrB*	*priS*::*tna-priS* carrying PL95

2.1.2 *Escherichia coli* strains

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

Strain	Genotype	Notes
DH5α	*F*$^-$ *endA*1 *glnV*44 *thi*-1 *rec*A1 *relA*1 *gyrA*96 *deoR nupG* Φ80d*lacZ*ΔM15 Δ(*lacZYA-argF*)U169, *hsdR*17(r_K^- m_K^+), λ^-	Standard cloning strain (Invitrogen)
SCS110	*rpsL* (*Str*^{r}) *thr leu endA thi*-1 *lacY galK galT ara tonA tsx dam dcm supE*44 Δ(*lac-proAB*) [F´ *traD36 proAB lacIqZΔM*15]	*dam*$^-$ *dcm*$^-$ mutant for preparing unmethylated DNA for *H. volcanii* transformations (Stratagene)
Rosetta2 (DE3) [pLysS]	Δ(ara-leu)7697 ΔlacX74 ΔphoA PvuII phoR araD139 ahpC galE galK rpsL (DE3) F'[lac^+ $lacI^q$ pro] gor522::Tn10 trxB pLysSRARE (Cam^R, Str^R, Tet^R)	Strain designed to enhance the expression of proteins that contain codons rarely used in *E. coli* (Novagen)

2.1.3 Other strains

Halogeometricum borinquense (Hbo) DSM11551 and *Haloquadratum walsbyi* (Hwa) DSM16790 were obtained from the DSMZ (German Collection of Microorganisms and Cell Cultures) and grown according to the DSMZ's directions.

2.1.4 Plasmids

Table 2.3 Plasmids used in this study

No.	Plasmid	Relevant properties
Cloning vectors		
PL1	pTA131	pBluescript II with *BamH*I-*Xba*I fragment containing *pyrE2* under ferredoxin promoter (Allers et al., 2004)
PL2	pTA187	Integrative vector with *hdrB* marker (Allers et al., 2004)
PL3	pTA230	pTA131 with *Nco*I-*Hind*III fragment containing pHV2 replication origin (Allers et al., 2004)
PL4	pTA233	pTA230 containing with PCR fragment containing *hdrB* under ferredoxin promoter (Allers et al., 2004)
PL5	pTA298	(Allers et al., 2004)
PL6	pTA409	Shuttle vector containing *pyrE2* and *hdrB* markers under ferredoxin promoter (Holzle et al., 2008)
PL7	pTA927	Overexpression vector with *pyrE2* marker and pHV2 origin, derived from pTA230 by insertion of 131-bp t.L11e terminator at *Kpn*I site and 224-bp pt*naA* promoter at *Apa*I and *Cla*I sites (Allers et al., 2010)
PL8	pWL-CBD-Sec11b	Shuttle vector containing the *cbd* gene encoding the *C.thermocellum* cellulosome CDB (Fine et al., 2006)
Plasmids for regulated gene expression		
PL9	pNPM-tna	pTA409- derived plasmid devoid of *Haloferax* replication origin, with 124 bp fragment of *tna* promoter inserted at *Bcl*I and *Nde*I sites. Constructed by Stuart MacNeill
PL10	pNPM-tna- PolB	pNPM-tna with 300 bp fragment of *polB* inserted at *Nde*I and *Eco*RV sites. Constructed by Stuart MacNeill
PL11	pNPM-tna-PolD1	pNPM-tna with 300 bp fragment of *polD1* inserted at *Nde*I and *Eco*RV sites. Constructed by Stuart MacNeill
PL12	pNPM-tna-PolD2	pNPM-tna with 630 bp fragment of *polD2* inserted at NdeI and *Eco*RV sites. Constructed by Stuart MacNeill
PL13	pNPM-tna-PcnA	pNPM-tna with 525 bp fragment of *pcnA* inserted at *Nde*I and *Eco*RV sites. Constructed by Stuart MacNeill
PL14	pNPM-tna-Mcm	pNPM-tna with 279 bp fragment of *mcm* inserted at *Nde*I and *Eco*RV sites. Constructed by Stuart MacNeill
PL15	pNPM-tna-RfcA	pNPM-tna with 285 bp fragment of *rfcA* inserted at *Nde*I and *Eco*RV sites. Constructed by Stuart MacNeill
PL16	pNPM-tna-RfcB	pNPM-tna with 500 bp fragment of *rfcB* inserted at *Nde*I and *Eco*RVsites. Constructed by Stuart MacNeill
PL17	pNPM-tna-RfcC	pNPM-tna with 630 bp fragment of *rfcC* inserted at *Nde*I and *Evo*RV sites. Constructed by Stuart MacNeill

PL18	pNPM-tna-LigA	pNPM-tna with 398 bp fragment of *ligA* inserted at *Nde*I and *Not*I sites. Constructed by Stuart MacNeill
PL19	pNPM-tna-RpaA	pNPM-tna with 300bp fragment of *rpaA1* inserted at *Nde*I and *EcoRV* sites. Constructed by Stuart MacNeill
PL20	pNPM-tna-RpaB	pNPM-tna with 300bp fragment of *rpaB1* inserted at *Nde*I and *EcoRV* sites. Constructed by Stuart MacNeill
PL21	pNPM-tna-RpaC	pNPM-tna with 300bp fragment of *rpaC* inserted at *Nde*I and *EcoRV* sites. Constructed by Stuart MacNeill
PL22	pNPM-tna-PriS	pNPM-tna with 460 bp fragment of *priS* inserted at *Nde*I and *EcoR*Vsites. Constructed by Stuart MacNeill
PL23	pNPM-tna-PriL	pNPM-tna with 296 bp fragment of *rfcB* inserted at *Nde*I and *EcoR*V sites. Constructed by Stuart MacNeill
PL24	pNPM-tna-Bcl-PolB	pNPM-tna-PolB- derived plasmid with 500 bp fragment upstream of *polB*, inserted at *Bcl*I site
PL25	pNPM-tna-Bcl-RpaC	pNPM-tna-RpaC- derived plasmid with 500 bp fragment upstream of *rpaC*, inserted at *Bcl*I site
PL26	pNPM-tna-Bcl-PriS	pNPM-tna-PriS- derived plasmid with 500 bp fragment upstream of *priS*, inserted at *Bcl*I site
Plasmids for constitutive high-level expression		
PL27	pNPM-fdx-RpaA	pNPM-tna-RpaA1-derived plasmid with *tna* promoter replaced with *fdx* promoter
PL28	pNPM-fdx-RpaB	pNPM-tna-RpaB1-derived plasmid with *tna* promoter replaced with *fdx* promoter
PL29	pNPM-fdx-RpaC	pNPM-tna-RpaC-derived plasmid with *tna* promoter replaced with *fdx* promoter
Plasmids for truncated/mutated *tna* promoter		
PL30	pNPM-tna2	pNPM-tna- derived plasmid with 114 bp fragment of *tna* promoter inserted at *Bcl*I and *Nde*I sites
PL31	pNPM-tna3	pNPM-tna- derived plasmid with 104 bp fragment of *tna* promoter inserted at *Bcl*I and *Nde*I sites
PL32	pNPM-tna4	pNPM-tna- derived plasmid with 94 bp fragment of *tna* promoter inserted at *Bcl*I and *Nde*I sites
PL33	pNPM-tna5	pNPM-tna- derived plasmid with 84 bp fragment of *tna* promoter inserted at *Bcl*I and *Nde*I sites
PL34	pNPM-tna2-PolB	pNPM02- derived plasmid with 300 bp fragment of *polB* inserted at *Nde*I and *EcoRV* sites
PL35	pNPM-tna2-PolD1	pNPM02 with 300 bp fragment of *polD1* inserted at *Nde*I and *EcoRV* sites
PL36	pNPM-tna2-Mcm	pNPM02- derived with 279 bp fragment of *mcm* inserted at *Nde*I and *EcoRV* sites
PL37	pNPM-tna3-PolB	pNPM03- derived plasmid with 300 bp fragment of *polB* inserted at *Nde*I and *EcoRV* sites
PL38	pNPM-tna3-PolD1	pNPM03- derived plasmid with 300 bp fragment of *polD1* inserted at *Nde*I and *EcoRV* sites
PL39	pNPM-tna3-Mcm	pNPM03- derived plasmid with 279 bp fragment of *mcm* inserted at *Nde*I and *EcoRV* sites
PL40	pNPM-tna4-PolB	pNPM04- derived plasmid with 300 bp fragment of *polB* inserted at *Nde*I and *EcoRV* sites
PL41	pNPM-tna4-PolD1	pNPM04- derived with 300 bp fragment of *polD1* inserted at *Nde*I and *EcoRV* sites
PL42	pNPM-tna4-Mcm	pNPM04- derived plasmid with 279 bp fragment of *mcm* inserted at *Nde*I and *EcoRV* sites
PL43	pNPM-tna5-PolB	pNPM05- derived plasmid with 300 bp fragment of *polB* inserted at *Nde*I and *EcoRV* sites
PL44	pNPM-tna5-PolD1	pNPM05- derived plasmid with 300 bp fragment of *polD1* inserted at *Nde*I and *EcoRV* sites

PL45	pNPM-tna5-Mcm	pNPM05- derived plasmid with 279 bp fragment of *mcm* inserted at *Nde*I and *EcoRV* sites
PL46	pNPM-L11-PolB	pNPM-tna-PolB- derived plasmid with wild-type *tna* promoter replaced with L11-*tna* fusion
PL47	pNPM0-L11-PolD1	pNPM-tna-PolD1- derived plasmid with wild-type *tna* promoter replaced with L11-*tna* fusion
PL48	pNPM-L11-Mcm	pNPM-tna-Mcm- derived plasmid with wild-type *tna* promoter replaced with L11-*tna* fusion
PL49	pNPM-tnaM1-PolB	pNPM-tna-PolB- derived plasmid with wild-type *tna* promoter replaced with its mutated form introducing a single base substitution (A to G) in *tna* boxA
PL50	pNPM-tnaM2-PolB	pNPM-tna-PolB- derived plasmid with wild-type *tna* promoter replaced with its mutated form introducing a single base substitution (T to C) in *tna* boxA
PL51	pNPM-tnaM3-PolB	pNPM-tna-PolB- derived plasmid with wild-type *tna* promoter replaced with its mutated form introducing a single base substitution (T to G) in *tna* boxA
PL52	pNPM-tnaM1-PolD1	pNPM-tna-PolD1-derived plasmid with wild-type *tna* promoter replaced with its mutated form introducing a single base substitution (A to G) in *tna* boxA
PL53	pNPM-tnaM2-PolD1	pNPM-tna-PolD1-derived plasmid with wild-type *tna* promoter replaced with its mutated form introducing a single base substitution (T to C) in *tna* boxA
PL54	pNPM-tnaM3-PolD1	pNPM-tna-PolD1-derived plasmid with wild-type *tna* promoter replaced with its mutated form introducing a single base substitution (T to G) in *tna* boxA
PL55	pNPM-M1 tna-MCM	pNPM-tna-MCM-derived plasmid with wild-type *tna* promoter replaced with its mutated form introducing a single base substitution (A to G) in *tna* boxA
PL56	pNPM-tnaM2-MCM	pNPM-tna-MCM-derived plasmid with wild-type *tna* promoter replaced with its mutated form introducing a single base substitution (T to C) in *tna* boxA
PL57	pNPM-tnaM3-MCM	pNPM-tna-MCM-derived plasmid with wild-type *tna* promoter replaced with its mutated form introducing a single base substitution (T to G) in *tna* boxA
PL58	pNPM-tnaM3-PolD2	pNPM-tna-PolD2-derived plasmid with wild-type *tna* promoter replaced with its mutated form introducing a single base substitution (T to G) in *tna* boxA
PL59	pNPM-tnaM3-RpaC	pNPM-tna-RpaC-derived plasmid with wild-type *tna* promoter replaced with its mutated form introducing a single base substitution (T to G) in *tna* boxA
Plasmids for gene deletion		
PL60	pTA131-ΔRpaA	pTA131- derived plasmid with 500bp fragments from 5' and 3' to the *rpaA* locus inserted at *Eco*RV/*BamH*I and *BamH*I/*Spe*I restriction sites, respectively
PL61	pTA131-ΔRpaA-trpA	pTA131-ΔRpaA- derived plasmid with *trpA* marker inserted at *BamH*I site
PL62	pTA131-ΔRpaB	pTA131- derived plasmid with 500bp fragments from 5' and 3' to the *rpaB* locus inserted at *Eco*RV/*BamH*I and *BamH*I/*Spe*I restriction sites, respectively
PL63	pTA131-ΔRpaB-hdrB	pTA131-ΔRpaB- derived plasmid with *hdrB* marker inserted at *BamH*I site
PL64	pTA131-ΔRpaC	pTA131- derived plasmid with 500bp fragments from 5' and 3' to the *rpaC* locus inserted at *Eco*RV/*BamH*I and *BamH*I/*Spe*I restriction sites, respectively
PL65	pTA131-ΔRpaC-trpA	pTA131-ΔRpaC- derived plasmid with *trpA* marker inserted at *BamH*I site
Plasmids for expression of full-length RpaC proteins		
PL66	pTA230-HfxRpaC	pTA230- derived plasmid with Hfx *rpaC* gene under *fdx* promoter cloned at *Nde*I/*Hind*III sites
PL67	pTA230-HgmRpaC	pTA230- derived plasmid with Hgm *rpaC* gene under *fdx* promoter cloned at *Nde*I/*Hind*III sites
PL68	pTA230-HqrRpaC	pTA230- derived plasmid with Hqr *rpaC* gene under *fdx* promoter cloned at *Nde*I/*Hind*III sites
Plasmids for expression of mutated RpaC proteins		
PL69	pTA230-HgmRpaC-ΔNTD	pTA230-HgmRpaC- derived plasmid with full-length Hgm *rpaC* gene replaced with Hgm *rpaC*-ΔNTD. Constructed by Stuart MacNeill

PL70	pTA230-HgmRpaC-ΔOB^{A}	pTA230-HgmRpaC- derived plasmid with full-length Hgm *rpaC* gene replaced with Hgm *rpaC*- ΔOB^{A} Constructed by Stuart MacNeill
PL71	pTA230-HgmRpaC-ΔOB^{B}	pTA230-HgmRpaC- derived plasmid with full-length Hgm *rpaC* gene replaced with Hgm *rpaC*- ΔOB^{B} Constructed by Stuart MacNeill
PL72	pTA230-HgmRpaC-ΔOB^{C}	pTA230-HgmRpaC- derived plasmid with full-length Hgm *rpaC* gene replaced with Hgm *rpaC*- ΔOB^{C} Constructed by Stuart MacNeill
PL73	pTA230-HgmRpaC-ΔOB^{AB}	pTA230-HgmRpaC- derived plasmid with full-length Hgm *rpaC* gene replaced with Hgm *rpaC*- ΔOB^{AB} Constructed by Stuart MacNeill
PL74	pTA230-HgmRpaC-ΔOB^{BC}	pTA230-HgmRpaC- derived plasmid with full-length Hgm *rpaC* gene replaced with Hgm *rpaC*- ΔOB^{BC} Constructed by Stuart MacNeill
PL75	pTA230-HgmRpaC-ΔOB^{AC}	pTA230-HgmRpaC- derived plasmid with full-length HgmRpaC gene replaced with HgmRpaC- ΔOB^{AC} Constructed by Stuart MacNeill
PL76	pTA230-HgmRpaC-ΔOB^{ABC}	pTA230-HgmRpaC- derived plasmid with full-length Hgm *rpaC* gene replaced with Hgm *rpaC*- ΔOB^{ABC} Constructed by Stuart MacNeill
PL77	pTA230-HgmRpaC-ΔCTD	pTA230-HgmRpaC- derived plasmid with full-length Hgm *rpaC* gene replaced with Hgm *rpaC*-ΔCTD
Plasmids for expression of Flag epitope-tagged RpaC proteins		
PL78	pTA230-Flag-HgmRpaC	pTA230-HgmRpaC-derived plasmid made by digesting with *Nde*I and ligating in a short dsDNA coding a Flag tag with *Nde*I cohesive ends, placing Flag tag on 5' terminus of the gene
PL79	pTA230-Flag-HgmRpaC-ΔNTD	pTA230-HgmRpaC-ΔNTD -derived plasmid made by digesting with *Nde*I and ligating in a short dsDNA coding a Flag tag with *Nde*I cohesive ends, placing Flag tag on 5' of the gene
PL80	pTA230-Flag-HgmRpaC-ΔOB^{A}	pTA230-HgmRpaC-ΔOB^{A} -derived plasmid made by digesting with *Nde*I and ligating in a short dsDNA coding a Flag tag with *Nde*I cohesive ends, placing Flag tag on 5' of the gene
PL81	pTA230-Flag-HgmRpaC-ΔOB^{B}	pTA230-HgmRpaC-ΔOB^{B} -derived plasmid made by digesting with *Nde*I and ligating in a short dsDNA coding a Flag tag with *Nde*I cohesive ends, placing Flag tag on 5' of the gene
PL82	pTA230-Flag-HgmRpaC-ΔOB^{C}	pTA230-HgmRpaC-ΔOB^{C} -derived plasmid made by digesting with *Nde*I and ligating in a short dsDNA coding a Flag tag with *Nde*I cohesive ends, placing Flag tag on 5' of the gene
PL83	pTA230-Flag-HgmRpaC-ΔOB^{AB}	pTA230-HgmRpaC-ΔOB^{AB} -derived plasmid made by digesting with *Nde*I and ligating in a short dsDNA coding a Flag tag with *Nde*I cohesive ends, placing Flag tag on 5' of the gene
PL84	pTA230-Flag-HgmRpaC-ΔOB^{BC}	pTA230-HgmRpaC-ΔOB^{BC} derived plasmid made by digesting with *Nde*I and ligating in a short dsDNA coding a Flag tag with *Nde*I cohesive ends, placing Flag tag on 5' of the gene
PL85	pTA230-Flag-HgmRpaC-ΔOB^{AC}	pTA230-HgmRpaC-ΔOB^{AC} -derived plasmid made by digesting with *Nde*I and ligating in a short dsDNA coding a Flag tag with *Nde*I cohesive ends, placing Flag tag on 5' of the gene
PL86	pTA230-Flag-HgmRpaC-ΔOB^{ABC}	pTA230-HgmRpaC-ΔOB^{ABC} -derived plasmid made by digesting with *Nde*I and ligating in a short dsDNA coding a Flag tag with *Nde*I cohesive ends, placing Flag tag on 5' of the gene
PL87	pTA230-Flag-HgmRpaC-ΔCTD	pTA230-HgmRpaC-ΔCTD -derived plasmid made by digesting with *Nde*I and ligating in a short dsDNA coding a Flag tag with *Nde*I cohesive ends, placing Flag tag on 5' of the gene
Plasmids for expression of CBD-fusion proteins		
PL88	pTA233-CCBD	pTA233- derived plasmid carrying 500bp fragment of CBD PCR amplified form pWL-CBD-Sec11b, cloned under *fdx* promoter at *Nde*I/*Kpn*I sites
PL89	pTA233-HfxRpaC-CCBD	pTA233-CCBD- derived plasmid with HfxRpaC NTD inserted at *Nde*I/*Nhe*I sites, giving CDB fused to RpaC at C-terminus
PL90	pTA233-HfxRpaC-NTD-CCBD	pTA233-CCBD- derived plasmid with HfxRpaC NTD inserted at *Nde*I/*Nhe*I sites, giving CDB fused to RpaC at C-terminus
PL91	pTA233-NCBD	pTA233- derived plasmid carrying 500bp fragment of CBD PCR amplified form pWL-CBD-Sec11b, cloned under fdx promoter at *Nde*I/*Eco*RV sites
PL92	pTA233-NCBD-HfxPriS CTD	pTA233-NCBD- derived plasmid with Hfx PriS CTD inserted at *Nde*I/*Nhe*I sites, giving CDB fused to PriS CTD at N-terminus
Plasmids for expression of PriS and GinS		

PL93	pTA230-HfxPriS-GinS	pTA230- derived plasmid with Hfx *priS-ginS* operon under *fdx* promoter cloned at *Nde*I/*Hind*III
PL94	pTA230-HfxGinS	pTA230- derived plasmid with Hfx *ginS* gene under *fdx* promoter cloned at *Nde*I/*Hind*III sites
PL95	pTA230-HgmPriSGinS	pTA230- derived plasmid with Hgm *priS-ginS* operon under *fdx* promoter cloned at *Nde*I/*Hind*III sites
PL96	pTA230-HgmPriS	pTA230- derived plasmid with Hgm *priS* gene under *fdx* promoter cloned at *Nde*I/*Hind*III sites
PL97	pTA230-HfxGinS	pTA230- derived plasmid with Hgm *ginS* gene under *fdx* promoter cloned at *Nde*I/*Hind*III sites

2.1.5 Chemicals and reagents

All reagents used in this study were purchased from Sigma-Aldrich, unless otherwise stated. Agar and yeast extract used for *H.volcanii* media were purchased from BD Biosciences and peptone used for *H.volcanii* media was purchased from Oxoid. Anti-Flag monoclonal antibodies were purchased from Sigma. HRP-linked sheep anti-mouse IgG secondary antibodies were purchased from GE Healthcare. Aphidicolin was purchased from EMD Chemicals. [Methyl-^{3}H] thymidine (1 mCi; 37 MBq) was purchased from MP Biomedicals.

2.1.5.1.1 Growth media

2.1.5.2 *Haloferax volcanii* growth media

H.volcanii cells were routinely grown on Hv-YPC, Hv-Ca or Hv-Min media prepared as described in the Halohandbook (Dyall-Smith, 2009). For selection procedures, supplements were selectively added as described below. All supplement solutions were sterilized by filtration through a 0.2 μm filter and added to media when cooled to 55^{o}C.

Table 2.4 Supplements added to different type of media

Genotype	**Growth on:**		
	Hv-YPC	**Hv-Ca**	**HV-Min**
Δ*pyrE2*	-	Uracil	Uracil
Δ*trpA*	-	Tryptophan	Tryptophan
Δ*hdrB*	Thymidine*	Thymidine*	Thymidine*

In addition to thymidine, Δ*hdrB* strains required supplementation with hypoxanthine to a final concentration of 50 μg/ml

Table 2.5 *H.volcanii* media supplements

Supplement	Abbreviation	Stock conc.	Final conc.
Uracil	ura	50 μg/ml	50 mg/ml
Tryptophan	trp	50 μg/ml	10 mg/ml
Thymidine*	thy	50 μg/ml	50 mg/ml
5-Fluoroorotic acid	5FOA	50 μg/ml	4 mg/ml (+uracil 10 μg/ml)

* In addition to thymidine, Δ*hdrB* strains required supplementation with hypoxanthine to a final concentration of 50 μg/ml

2.1.5.3 *Escherichia coli* growth media

E.coli cells were grown in LB medium (Formedium) supplemented when necessary with ampicillin to a final concentration of 100 μg/ml.

2.1.5.4 *Haloferax volcanii* buffers and solutions

30% Salt Water (SW): 240 g NaCl, 30 g $MgCl_2$ x $6H_2O$, 35 g $MgSO_4$ x $7H_2O$, 7 g KCl, 20 ml 1 M Tris-HCl pH 7.5, dH_2O to 1 litre

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

Spheroplast Dilution Solution: 23% SW, 15% sucrose, 37.5 mM $CaCl_2$

Regeneration Solution: 18% SW, 1xYPC, 15% sucrose, 30 mM $CaCl_2$

Transformation Dilution Solution: 18% SW, 15% sucrose, 30 mM $CaCl_2$

Solutions used for purification of CBD-fused proteins:

Lysis buffer: 1% Triton X-100, 3 M KCl, 50 mM Tris-HCl pH 7.2

Washing buffer: 3 M KCl, 50 mM Tris-HCl pH 7.2

Cellulose (Sigmacell Cellulose, purchased from Sigma): 10% (w/v) in dH_2O

2.1.5.5 Other buffers and solutions

TE buffer: 1 mM EDTA, 10 mM Tris-HCl pH 7.5

TAE buffer: 40 mM Tris-HCl pH 7.5, 20 mM acetic acid, 1 mM EDTA

TBE buffer: 89 mM Tris, 89 mM boric acid, 2 mM EDTA

DEPC Treated Water: dH_2O treated with 0.1% (v/v) diethylpyrocarbonate for at least 1 hour at 37 °C and then autoclaved

RNA loading buffer (6x): 0.25% Bromophenol Blue, 0.25% Xylene Cyanol, 30% Glycerol, 1.2% SDS, 60 mM sodium phosphate buffer pH 6.8

FIX solution: 20% (v/v) ethanol, 7.5% (v/v) acetic acid

TSB buffer: LB broth containing 10% (w/v) PEG, 5% (v/v) DMSO, 20-50 mM magnesium salts ($MgSO_4$ or $MgCl_2$)

2.2 Methods

2.2.1Manipulation and analysis of nucleic acids

2.2.1.1 PCR amplification

Amplification of DNA was performed using either Long PCR Enzyme Mix (Fermentas) or MyTaq Red Mix (Bioline). Long PCR Enzyme Mix was used for cloning purposes, where higher PCR fidelity was required. This kit contains a unique blend of a highly processive *Taq* DNA polymerase and second thermostable polymerase that exhibits 3'→5' exonuclease (proofreading) activity. Quantity and final concentrations of reagents for each reaction are listed below.

Table 2.6 Reagents for PCR amplification using Long PCR Enzyme Mix

Reagent	Quantity for 20 µl reaction	Final concentration
Water, nuclease-free	13.5 µl	-
10x Long PCR Buffer with $MgCl_2$	2 µl	1X
2mM dNTPs mix	2 µl	0.2mM of ech
Forward primer	0.2 µl	1 µM
Reverse primer	0.2 µl	1 µM
DMSO	1 µl	5%
Template DNA	1 µl	variable
Long PCR Enzyme Mix	0.1 µl	1.25 u / 20 µl

DMSO was added to the final concentration of 5% due to the high percentage of G+C base pairs in *H.volcanii* genome. Primers were supplied by Eurofins MWG Operon. MyTaq Red Mix was used for diagnostic amplification. Quantity and final concentrations of reagents for each reaction for this kit are listed below.

Table 2.7 Reagents for PCR amplification using MyTaq Red Mix

Reagent	Quantity for 10 µl reaction	Final concentration
MyTaq Red Mix	5 µl	1x
Forward and reverse primer mix	4.5 µl	4.5 pmol of each
Template DNA	0.5 µl	variable

All PCR reactions were carried out using a Techne TC-4000 thermocycler and Roche G/C rich programme summarised in Table 2.7. If required, PCR reactions were purified using the MinElute Reaction Cleanup kit (QIAGEN).

Table 2.8 PCR reaction conditions

Cycles		Temperature	Time
Initial Denaturation			
1		95°C	3 min
Amplification cycles 1-10			
10	Denaturation	95°C	30 sec
	Annealing	55°C	30 sec
	Elongation	72°C	1 min 30 sec
Amplification cycles 11-35			
25	Denaturation	95°C	30 sec
	Annealing	55°C	30 sec
	Elongation	1st cycle: 72°C	1 min 30 sec
		Final cycle: 72°C	3 min 30 sec
Final Elongation			
1		72°C	10 min

2.2.1.2 Plasmid DNA extraction

Plasmid DNA from *Escherichia coli* cells was extracted using QIAprep Spin Miniprep Kit (QIAGEN). Extractions were carried out using 10 ml of cell culture. The purity and concentration of DNA was determined using a Nanodrop 1000 (Thermo Scientific) and verified by electrophoresis.

2.2.1.3 Restriction digest, dephosphorylation of vector DNA, ligation of DNA

Most of the restriction enzymes were supplied by New England Biolabs and used following the manufacturer's instructions. *Bcl*I enzyme was supplied by Fermentas. Double digestion was performed using Fermentas Tango buffer. To prevent self-ligation of vector DNA, 5' phosphate groups were removed after restriction digest, using Antarctic Phosphatase. Samples were incubated with phosphatase and suitable buffer for 30 minutes at 37°C. If required, restriction digests were purified using the MinElute Reaction Cleanup kit (QIAGEN). Ligations were performed using Quick Ligase or T4 DNA ligase (Fermentas). For Quick Ligase, reactions were carried out at room temperature for 20 minutes and for T4 ligase- at room temperature for 1h. For vector:insert ligations, reactions contained a molar ratio of ~5:1 insert to vector DNA.

2.2.1.4 DNA sequencing and oligonucleotide synthesis

All cloning constructs were sequenced to ensure the absence of unwanted sequence changes. All sequencing reactions were carried out by DNA Sequencing & Services, Dundee. Oligonucleotides were supplied by Eurofins MWG Operon.

2.2.2 *Escherichia coli* microbiology

2.2.2.1 Growth of *Escherichia coli* and storage

Cultures on solid media were grown at 37^{o}C in a static incubator. Liquid cultures were grown at 37^{o}C with shaking. For short-term storage, plates were stored at 4^{o}C. For long-time storage, 25% (v/v) glycerol was added to cultures, mixed and then stored at -80^{o}C.

2.2.2.2 *Escherichia coli* transformation

Escherichia coli cells were grown in 20 ml LB broth to the early exponential phase, pelleted at 3000 rpm for 10 minutes at room temperature and resuspended in 1 ml of ice-cold TSB buffer. After 30 minutes on ice, 0.1 ml aliquots of competent cells were transfer to round-bottom polypropylene tubes, mixed with 10 µl of a ligation reaction or 3 µl of a plasmid and cells were incubated on ice for 30 minutes. After that time, 0.2 ml of LB broth was added and tubes were incubated at 37^{o}C for 1 hour with shaking prior to plating onto selective LB plates. Plates were incubated at 37^{o}C overnight.

2.2.2.3 Generation of unmethylated (*dam*$^{-}$) plasmid DNA

H.volcanii transformation requires unmethylated DNA as this organism possesses a restriction endonuclease that cleaves DNA at 5'-GA$_{M}$TC-3' sequences (Blaseio and Pfeifer, 1990). Consequently, unmethylated plasmid DNA is necessary to prevent degradation. All plasmids were transformed into *E.coli* SCS110, a strain deficient for two methylases (Dam and Dcm). Transformation was carried out as described above and DNA was then extracted as usual, ready for transformation of *H.volcanii*.

2.2.3 *Haloferax volcanii* microbiology

2.2.3.1 Growth of *Haloferax volcanii* and storage

Depending on the strain and assay, cultures were grown in YPC, Hv-Ca or Hv-Min media. Cultures on solid media were grown at 45°C in a static incubator in a sealed plastic bag to prevent desiccation. Liquid cultures were grown at 45°C with shaking.

For short-term storage, plates and cultures were stored at room temperature. For long-term storage, 20% glycerol (v/v) was added to cultures, mixed and then stored at -80°C.

2.2.3.2 Extraction of genomic DNA from *Haloferax volcanii*

For PCR analysis, chromosomal DNA was extracted as follows. Single colonies were grown in 200 µl of suitable medium for 24-48 h. 5 µl was transferred to 500 µl sterile water to lyse the cells by osmotic shock. The DNA solution was then heated to 70^{o}C for 10 min. 1 µl of the DNA solution was used as the template in PCR reactions. When pure chromosomal DNA at concentration 10 ng/ µl was needed to validate RT-PCR assays, the Dnase Blood and Tissue kit (QIAGEN) was used.

2.2.3.3 Extraction of RNA from *Haloferax volcanii*

RNA was isolated using RNeasy kit (QIAGEN), following the manufacturer's instructions. The purity and concentration of the RNA was checked using a Nanodrop 1000 (Thermo Scientific) and verified by electrophoresis (1 % agarose gel in TBE buffer with 1M guanidine thiocyanate as denaturating agent). 10 ng of RNA was used as a template for reverse-transciptase RT-PCR.

2.2.3.4 *Haloferax volcanii* transformation

Transformation was carried out by using polyethylene glycol 600, according to the Halohandbook (Dyall-Smith, 2009)

2.2.3.5 Mating assay

Mating of *H.volcanii* strains was performed essentially as described previously (Zhao et al., 2006). Briefly, strains to be mated were grown to mid-exponential phase (OD_{650nm} of 0.4) in Hv-YPC medium before being combined and filtered onto a 0.45 µm filter. The filter was then placed face-up on an Hv-YPC plate. After overnight incubation at 45^{o}C, the cells were washed from the filter using 1 ml of sterile 18% SW and 100 µl aliquots plated onto Hv-Ca plates with necessary supplements. Colonies formed after 5 days incubation at 45^{o}C were then analyzed by PCR.

2.2.4 *Haloferax volcanii* phenotyping assays

2.2.4.1 Spotting technique

Single colonies were used to inoculate 10 ml of Hv-Min supplemented with 0.075 mM tryptophan (and uracil and thymidine if required) and incubated at 45^{o}C with

shaking until the OD_{650nm} reached 0.5. The cells were pelleted by centrifugation at 2000 rpm for 8 minutes. The supernatant was removed and the cells were washed in 10 ml Hv-Min medium to remove the remaining tryptophan. The cells were pelleted again and the pellet was resuspended in Hv-Min medium to adjust the OD_{650} to 0.3. Ten-fold serial dilutions were made (from 10^{-1} to 10^{-6}) using Hv-Min as a diluent and 5 µl each dilution was spotted onto Hv-Min plates either containing and lacking tryptophan. Plates were incubated at 45°C for at least 5 days. Experiments were always performed for mutant strains with wild-type (DS70 or H98) as a control.

2.2.4.2 Growth curves

Single colonies were used to inoculate 10 ml of Hv-Min supplemented with 0.075 mM tryptophan (and uracil and thymidine) and incubated at 45°C with shaking until OD_{650} reached 0.5. Those cultures were used to inoculate fresh Hv-Min medium supplemented with 0.075mM tryptophan and adjusted to an OD_{650nm} to 0.1. Cells were grown for 7 hours and the OD was measured hourly. At this stage, half of the cells were pelleted by centrifugation at 2000 rpm for 8 minutes. The supernatant was removed and the cells were washed in 10 ml Hv-Min medium to removed remaining tryptophan. The cells were pelleted again and the pellet was resuspended in Hv-Min medium to adjust the OD_{650} of 0.1. The cells were then grown 8 hours and the OD was measured hourly. Experiments were always performed for mutant strains with wild-type (DS70 or H98) as a control.

2.2.4.3 Viability curves

2.2.4.4 DNA damage sensitivity assays

DNA damage sensitivity assays were performed using following agents: UV, 4NQO, MMS and phleomycin. Strains were washed in 10 ml Hv-Min medium to remove the remaining tryptophan and resuspended at an OD_{650} 0.3 prior to exposure to the DNA damaging agents.

In all experiments, the cells were grown in Hv-Min medium supplemented with 0.075mM tryptophan to an OD_{650nm} of 0.3. For the UV sensitivity assay, the cells were 10-fold serial diluted in 18% SW and, depending on technique used, 5 µl was spotted or 100 µl was plated (to generate survival curves) onto Hv-Min plates supplemented or lacking tryptophan, as noted above. The plates were allowed to dry and immediately irradiated with doses from 50 to 200 J/m^2 of UV-C (254 nm) using a Stratalinker UV1800 (Agilent). The plates were incubated in dark at 45°C until

colonies were formed. For 4NQO, MMS and phleomycin sensitivity assays, cells were grown in Hv-Min medium supplemented with 0.075mM tryptophan to an OD_{650nm} of 0.3, divided into several aliquots of 1 ml and incubated with 20 µl of appropriated concentration of drug for 1 h with shaking at 45^{o}C. The cells were 10-fold serial diluted in 18% SW and spotted or plated on appropriate Hv-Min plates. The plates were incubated at 45^{o}C until colonies were formed. Final concentrations for 4NQO, MMS and phleomycin are listed below.

Table 2.9 Final concentrations for DNA damaging agents used is sensitivity assays

DNA damaging agent	Final concentrations	Stock
4NQO	0.2, 0.4, 0.6, 0.8 µg/ml	1 mg/ml in DMSO
MMS	0.02, 0.04, 0.06, 0.08 %	4% in 18% SW
phleomycin	5 and 10 mg/ml	20 mg/ml in 18% SW

2.2.5 Gene expression analysis by RT-PCR

Analysis of mRNA levels was carried out using One-Step RT-PCR with SYBR Green (BioRad). A SYBR Green assay uses a pair of PCR primers that amplify a specific region within the target sequence of interest and includes SYBR Green (double- stranded DNA binding dye) for detecting the amplified product. Amplicons were designed to be 75- 200 bp. To avoid secondary structure the mfold Web Server (http://www.bioinfo.rpi.edu/applications/mfold/) was used. Assay validation and optimization with SYBR Green Supermix (BioRad) was performed using serial dilutions of DNA of known concentration (from 10 ng/µl to 0.001 ng/µl) as a template. The optimal annealing temperature was chosen by performing a temperature gradient, testing a range of different temperatures above and below melting temperature (Tm) of the primers. To avoid nonspecific product co-amplification, melt-curve analysis was performed for all experiments. Dilution of a known template serves to construct standards curves for all tested genes and determines the efficiency of the assay. The reaction is well designed when the standard curve is linear (R^2 > 0.998) and efficiency is between 90 and 105 %. Quantity and final concentrations of reagents for each reaction using SYBR Green Supermix are listed below.

Table 2.10 SYBR Green Supermix reagents for RT-PCR

Reagent	Quantity for 20 µl reaction	Final concentration
Water, nuclease- free	7 µl	-
2x SYBR Green Supermix	10 µl	1x
Forward primer	1 µl	1 µM
Reverse primer	1 µl	1 µM
Template DNA	1 µl	10 ng – 0.001 ng

Negative controls without DNA template were included. The small subunit of ribosomal protein gene *HVO_0561* has been chosen as internal standard as expression of this gene is not influenced by tryptophan. For final reverse transcriptase experiments, One-Step RT-PCR with SYBR Green was used, where cDNA synthesis and PCR amplification are carried out in one tube.

Single *H.volcanii* colonies were used to inoculate 15 ml of Hv-Min medium supplemented with 0.075 mM tryptophan (and uracil and thymidine if required) and incubated at 45^{o}C with shaking until the OD_{650nm} reached 0.1. Cells were pelleted by centrifugation at 2000 rpm for 10 minutes. The supernatant was removed and cells were washed in 10 ml Hv-Min medium to removed tryptophan completely. Cells were pelleted again and the pellet was resuspended in 1 ml Hv-Min medium. Those cultures were used to inoculate 20 ml of Hv-Min/ Hv-Min + 0.075mM tryptophan and the OD_{650} was adjusted to 0.1. Cells were grown 8 hours at 45^{o}C with shaking and OD was measured hourly. 2 ml samples were taken immediately after centrifugation at time 0 and after 2, 4, 6, and 8 hours. RNA was isolated using RNeasy kit as described previously (Materials and Methods, 2.2.3.3). The quantity and final concentrations of reagents for each reaction for One-Step RT-PCR with SYBR Green (BioRad) are listed below. Relative tryptophan- induced transcipt levels were calculated according to the Pfaffl method (Pfaffl, 2001)

Table 2.11 One-Step RT-PCR with SYBR Green reagents for reverse transcriptase RT-PCR

Reagent	Quantity for 20.5 µl reaction	Final concentration
Water, nuclease- free	7.2 µl	-
2x SYBR Green Supermix	10 µl	1x
Reverse transcriptase	0.4 µl	
Forward primer	0.8 µl	0.8 µM
Reverse primer	0.8 µl	0.8 µM
Template RNA	1 µl	10 ng

All reverse transcriptase RT-PCR reactions were carried out using a BioRad iQ5 thermocycler and programme described in Table 2.12.

Table 2.12 Reverse transcriptase RT-PCR reactions

Cycles	Repeats	Temperature	Time
1	1	50°C	10 min
		95°C	5 min
2	35	95°C	10 sec
		60.5°C	30 sec
3	1	95°C	1 min
		55°C	1 min
4	80	55°C	10 sek

2.2.6 Monitoring DNA synthesis

DNA synthesis was monitored by labelling cells with [methyl-^{3}H] thymidine and measuring the incorporation of radionucleotides into DNA for 30 hours. Single colonies were used to inoculate 25 ml of Hv-Min supplemented with 0.075 mM tryptophan and incubated at 45°C with shaking until cells reached the mid-exponential phase. The cells were pelleted by centrifugation at 2000 rpm for 8 minutes. The supernatant was removed and the cells were resuspended in 20 ml Hv-Min medium to remove the remaining tryptophan. The cells were pelleted again and resuspended in 1 ml of Hv-Min medium. These cultures were used to inoculate fresh Hv-Min/Hv-Min supplemented with 0.075 mM tryptophan and the OD_{650nm} was adjusted to 0.1. The cells were grown for 16 hours and OD_{650nm} was adjusted again to 0.01. Cells were grown for 2 hours and 5 ml portions of cultures were labelled with 20 μCi [methyl-^{3}H] thymidine (mixed with a cold thymidine in 1:1000 ratio). The cells were grown at 45°C with shaking and at the indicated time points 200 μl of labelled cells were mixed with 10 ml ice-cold 10% (w/v) trichloroacetic acid containing 1% (w/v) sodium pyrophosphate. Samples were incubated on ice for 30 minutes and filtered through 0.7 μm filters (Whatman grade GF/F). The filters were washed three times with 10 ml of 10% TCA and 1% sodium pyrophosphate and once with 10 ml of 100 % ice-cold ethanol. Filters were transferred to 4 ml liquid scintillation vials and the radioactivity was measured using MicroBeta^2 counter (Perkin Elmer).

2.2.7 *Haloferax volcanii* reverse genetics

2.2.7.1 Construction of conditional mutants of replication genes

2.2.7.1.1 Construction of *tna*-plasmid integrant strains

Tna-plasmid integrant strains were generated by construction of pNPM-*tna* plasmid and its derivatives and transforming those plasmids into *H.volcanii* H98 (Δ*pyrE2* Δ*hdrB*) (Table 2.1). Plasmid pNPM-tna was constructed by amplifying the 124 bp *tna*A promoter from *H.volcanii* DS70 genomic DNA using oligonucleotides P1 and P2 (Appendix, Table A1), digesting the PCR product with *Bcl*I and *Not*I and ligating into *Bcl*I- and *Not*I-digested pTA409 (Holzle et al., 2008). Oligonucleotide P2 was design to introduce unique *Nde*I and *Pac*I restriction sites downstream of the *tna* promoter so the resulting plasmid (pNPM-tna, PL9, Table 2.3) contains a multiple cloning site with those unique sites along with additional unique sites that were present in pTA409. pNPM-*tna* plasmid was sequence to ensure the absence of unwanted sequence changes in the *Bcl*I-*Not*I region. Next, oligonucleotides P3-20 (Appendix, Table A1) were used to amplify 5' regions of selected ORFs form *H.volcanii* DS70 genomic DNA. The PCR products were digested with appropriate restriction enzymes and cloned separately into plasmid pNPM-tna digested with corresponding enzymes, giving plasmids PL9-22. Details of the genes of interest, size of the inserts and restriction sites are summarised in Table 2.13.

Plasmids PL9-22 were subsequently passaged through *E.coli* SCS110 and transformed onto *H.volcanii* H98. Integrative transformants were obtained on Hv-Ca medium without additional supplements. Plasmid integration at the correct chromosomal loci was confirmed by PCR using oligonucleotides P1 with P21-P34 (Appendix, Table A1). The validated strains were then purified by streaking to the single colonies on Hv-Min plates supplemented with 0.075 mM tryptophan.

Table 2.13 Details about construction of pNPM-tna-derived plasmids

No.	Plasmid name	Gene	Insert size	Restriction site
PL10	pNPM-tna- PolB	*polB*	300 bp	*Nde*I-*EcoR*V[1]
PL11	pNPM-tna-PolD1	*polD1*	300 bp	*Nde*I-*EcoR*V[1]
PL12	pNPM-tna-PolD2	*polD2*	300 bp	*Nde*I-*EcoR*V[1]
PL13	pNPM-tna-PcnA	*pcnA*	525 bp	*Nde*I-*BmgB*I[2]
PL14	pNPM-tna-Mcm	*mcm*	279 bp	*Nde*I-*EcoR*V[2]
PL15	pNPM-tna-RfcA	*rfcA*	285 bp	*Nde*I-*BmgB*I [2]
PL16	pNPM-tna-RfcB	*rfcB*	500 bp	*Nde*I-*BsrB*I[2]
PL17	pNPM-tna-RfcC	*rfcC*	630 bp	*Nde*I-*BmgB*I [2]
PL18	pNPM-tna-LigA	*ligA*	398 bp	*Nde*I-*Not*I[2]
PL19	pNPM-tna-RpaA	*rpaA*	300 bp	*Nde*I-*EcoR*V[1]
PL20	pNPM-tna-RpaB	*rpaB*	300 bp	*Nde*I-*EcoR*V[1]
PL21	pNPM-tna-RpaC	*rpaC*	300 bp	*Nde*I-*EcoR*V[1]
PL22	pNPM-tna-PriS	*priS*	460 bp	*Nde*I-*Eco*721[2]
PL23	pNPM-tna-PriL	*priL*	296 bp	*Nde*I-*BmgB*I [2]

1-Restriction site introduced into 3' oligonucleotide sequence
2-Restriction site in genomic DNA

2.2.7.1.2 Construction of *tna*-promoter replacement strains

For *polB*, *rpaC* and *priS* genes, genetically stable *tna*-promoter replacement strains were constructed as follows. A 500 bp *Bcl*I fragment carrying the upstream flanking region of the appropriate gene was amplified by PCR using oligonucleotides P37-42 (Appendix, Table A1) and DS70 genomic DNA as a template and cloned into corresponding pNPM-tna plasmids at unique *Bcl*I site upstream of *tna* promoter. The resulting plasmids (PL24-26, Table 2.3) were sequenced in order to determine the correct orientation of inserts and to ensure the absence of unwanted sequence changes and passaged through *E.coli* SCS110 prior to transformation into *H.volcanii*. Integrative transformants were obtained on Hv-Min medium supplemented with 0.075 mM tryptophan. Plasmid integration at the correct chromosomal loci was confirmed by PCR using oligonucleotides P43-P46 (Appendix, Table A1). For each promoter replacement strain, three integrant colonies were picked, resuspended in 18% SW and plated on Hv-Min plus tryptophan plates supplemented with uracil, thymidine and 5FOA. Candidate *tna*-promoter replacement colonies were screened by PCR using oligonucleotides P43-50 (Appendix, Table A1).

2.2.7.2 Gene deletion in *Haloferax volcanii*

2.2.7.2.1 Construction of plasmids for gene deletion

Deletion strategy involved construction of pTA131-derived plasmids carrying the *trpA* or *hdrB* marker genes flanked by the sequences corresponding to about 500 bp from 5' and 3' flanking regions of genes to be deleted. The flanking regions were amplified by PCR using *H.volcanii* DS70 genomic DNA as a template, external oligonucleotides introducing restriction sites present in the polylinker region of pTA131 and internal oligonucleotides introducing unique restriction site, not present in pTA131. PCR products and pTA131 were digested with corresponding restriction enzymes and flanking regions were cloned into pTA131. Next, marker gene was inserted in the centre on the 5' and 3' flanking regions using unique restriction site introduced by the internal oligonucleotides. The resulting plasmids were passage through *E.coli* SCS110 to be demethylated. The strategy used to construct plasmids for gene deletion is summarised in Figure 2.1.

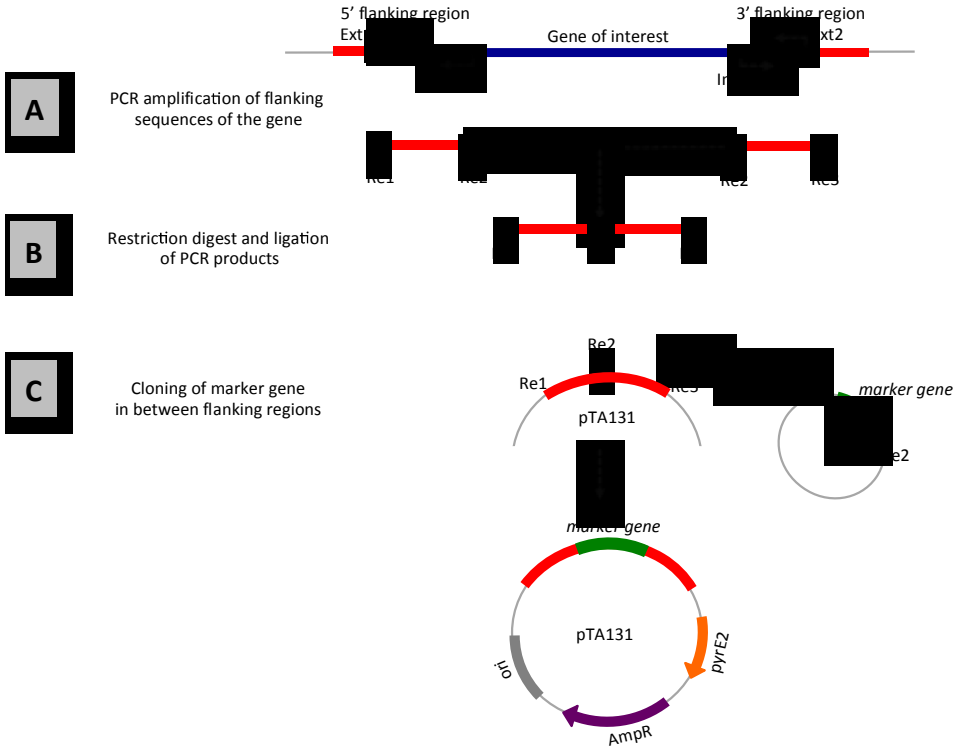


Figure 2.1 Construction of plasmids for gene deletion

Plasmids for gene deletion were constructed by PCR amplification 5' and 3' flanking regions of the genes and ligating them into plasmid pTA131. Next, restriction fragments carrying selected marker gene were cloned into the plasmid in the centre of flanking regions.

2.2.7.2.2 Gene deletion

Plasmids for gene deletion were transformed into *H.volcanii* H99 (Δ*pyrE2* Δ*hdrB* Δ*trpA*). Transformant colonies with deletion constructs integrated at the native locus of gene of interest were obtained on Hv-Ca medium followed by plating on selective and non-selective Hv-Ca plates supplemented with both uracil and 5-FOA. Adding 5-FOA to the medium allowed counter selection of recombinants that have lost the plasmid while this compound is converted to toxic 5-fluorouracil in *pyrE2*$^{+}$ cells but not in *pyrE2*$^{-}$ cells. Depending upon the location of the recombinant event either wild-type or the desired deletion cells can be recovered (if viable). Candidate gene deletion colonies were screened by PCR. Deletion strains were then purified by streaking to single colonies on Hv-YPC medium, re-tested by PCR and stored at -80^{o}C in 20% glycerol.

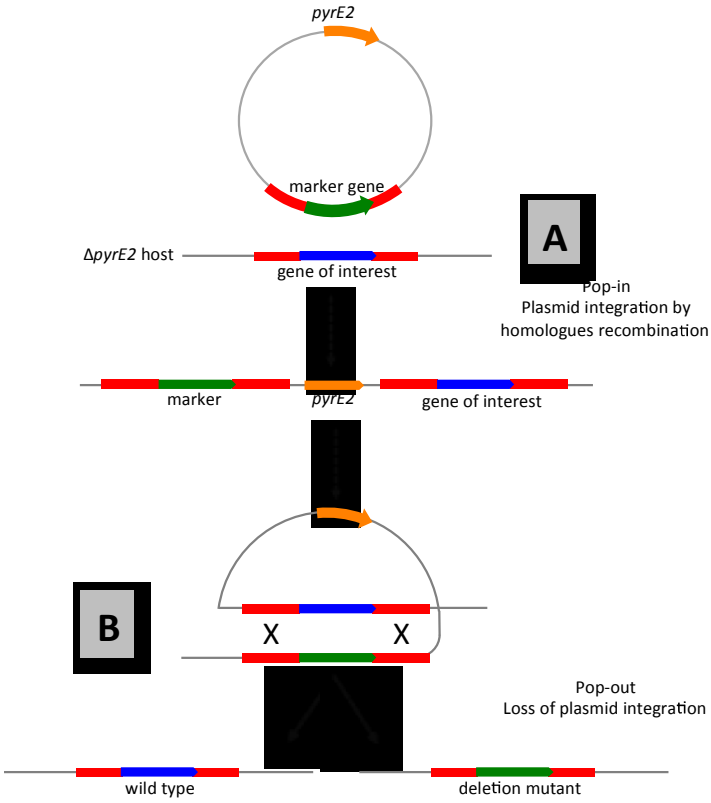


Figure 2.2 Pop-in/pop-out gene deletion strategy

A plasmid carrying *pyrE2* gene (orange box) and deletion construct containing the auxotrophic marker (green box) flanked by the 5'- and 3'-region of the gene of interest (red boxes) is transformed to Δ*pyrE2* *H.volcanii* strain. The plasmid is integrated into the chromosome (pop-in step, **A**). Subsequent loss of the plasmid (pop-out step, **B**) by intrachromosomal recombination may acur by reversion to the wild-type or replacing the gene of interest with marker.

2.2.8 Western blot of Flag epitope-tagged RpaC proteins

Expression of the proteins was performed in Hv-Min medium (with appropriate supplements). Protein extracts were prepared from cell culture grown to mid-log phase (OD_{650nm} of 0.8) in Hv-Min medium (with supplements) and analysed by sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE). Then, samples were transferred to PVDF membrane and subsequently incubated in blocking solution (5% milk, 0.05% Tween in PBS) for 1h at room temperature. The

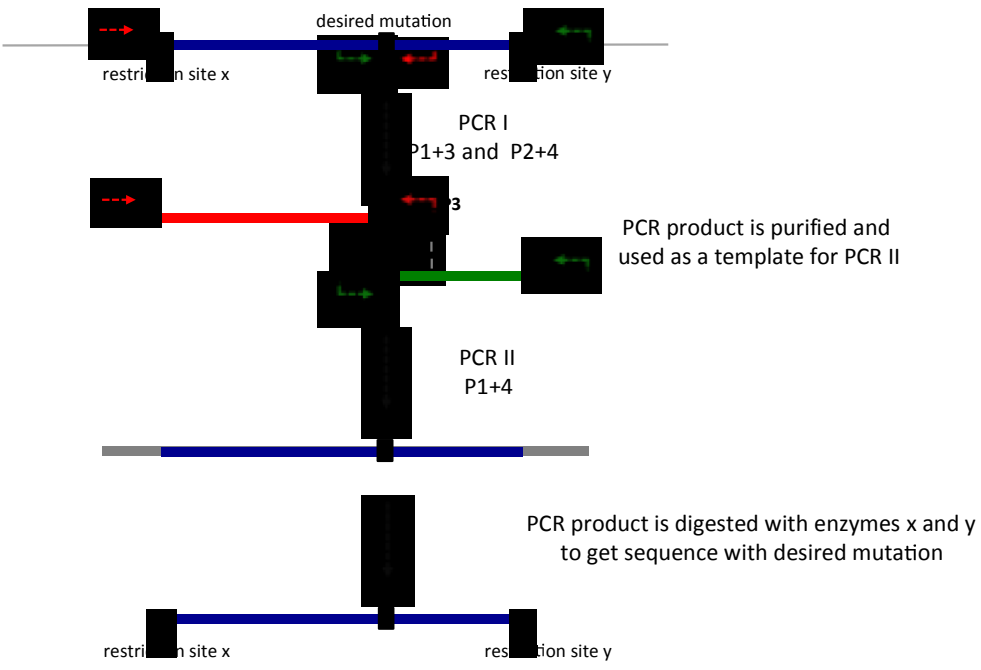
membrane was rinsed in PBS and incubated for 1h at room temperature with anti-Flag monoclonal antibodies (diluted 1:1000). Next, the membrane was washed 5 times in PBS and incubated for 1h at room temperature with HRP-linked sheep anti-mouse secondary antibodies (diluted 1:20000). After washing membrane in PBS, chemiluminescence detection was performed using Thermo Scientific SuperSignal West Pico Chemiluminescent Substrate (Thermo Scientific). The signal was visualized using an LAS 3000 imaging system (Fujifilm).

2.2.9 Site-directed mutagenesis of *tna* promoter

Single point mutation in the p*tna* TATA box consensus were generated with site-directed mutagenesis using PCR overlap extension (Figure 2.3). Two pairs of primers were desired, external and internal. Internal primers contained a mutated nucleotide. Primers were design to create an overlap region in the two PCR fragments. Those two pairs of primers were used in the first round on PCR. Two PCR fragments were then purified and served as a template for the second round of PCR. In this PCR reaction, the two external primers were used to amplify the full-length sequence containing desired mutation. The final PCR product was purified and digested with appropriate restriction enzymes to release the desired fragment.

Figure 2.3 Site-directed mutagenesis using PCR overlap extension

Two pairs of primers are design, external (P1 and P4) and internal (P2 and P3). Internal primers contain desired mutation. In the first round of PCR (PCR I), two PCR products having overlapping ends are generated (showed as red and green rectangles). These fragments are combined in a second round of PCR (PCR II). The complementary ends anneal, allowing the 3' overlap of each strand to serve as a primer for the 3' extension of the complementary strand. The final PCR product is then purified and digested with appropriate restriction enzymes (x and y) to release wanted sequence.



2.2.10 Expression of *Haloferax volcanii* RpaC

For purification purposes, selected proteins were fused to a cellulose-binding domain of *Clostridium thermocellum* cellulosome (CBD) and expressed in *H.volcanii*. Vectors for the expression of the chimeric proteins were based on the pTA233 shuttle vector (Allers et al., 2004) containing *E.coli* and *H.volcanii* replication origins and *hdrB* marker gene to allow selection of transformed *H.volcanii* cells in Hv-YPC rich medium. The *cbd* gene encoding *C.thermocellum* cellulose binding domain was PCR amplified from plasmid pWL-CBD-Sec11b (Fine et al., 2006) as described below. Expression was carried out in *H.volcanii* H98 (Δ*pyrE2* Δ*hdrB*) strain in Hv-YPC medium.

2.2.10.1 Construction of plasmids for the *H.volcanii* RpaC expression

The full-length RpaC protein and RpaC N-terminal domain (NTD) were expressed with C-terminal CBD. The *cbd* gene was PCR amplified from plasmid pWL-CBD-Sec11b using oligonucleotides P123 and P124 (Appendix, Table A1). The forward oligonucleotide P123 was design to introduce unique *Nde*I, *EcoR*V and *Nhe*I sites to ease future sub-cloning. The reverse oligonucleotide P124 was design to introduce two stop codons (TAG and TAA) and *Kpn*I site at the 3' end of the CBD. The resulting PCR product was digested with *Nde*I and *Kpn*I and ligated together with the *Bam*HI-*Nde*I restriction fragment of *fdx* promoter, described previously (Material and Methods, 2.2.7.1), into plasmid pTA233 that had been digested with *BamH*I and *Kpn*I. The resulting plasmid pTA233-CCBD (PL88, Table 2.3) was sequenced to ensure the absence of unwanted sequence changes in the *Bam*HI-*Kpn*I region. Next, oligonucleotides P125-P127 (Appendix, Table A1) were used to amplified full-length *rpaC* ORF (lacking the native stop codon) and *rpaC* NTD from *H.volcanii* DS70 genomic DNA. The PCR products were digested with *Nde*I-*Nhe*I and clone separately into pTA233-CCBD restricted with the same enzymes. The resulting plasmids (PL89 and PL90, Table 2.3) were sequence again and passaged through *E.coli* SCS110 prior to transformation into *H.volcanii*.

2.2.10.2 Protein expression

For protein expression, *H.volcanii* cells were grown in Hv-YPC medium at 45^{o}C until an OD_{650nm} of 0.8 was achieved. Cells were then harvested by centrifugation at 8500 rpm for 15 minutes and pellet was resuspended in lysis buffer (1% Triton X-100, 3 M KCl, 50 mM Tris-HCl pH 7.2) prior to sonication (2 x 30 seconds) on ice. The cell lysate was clarified by centrifugation at 10000 rpm for 15 minutes followed

by filtration through a 0.45 μm syringe filter. 200 μl of cellulose beads (10% w/v solution in dH_2O) was added to the lysate, mixed and rotated on wheel for 1h at room temperature. The beads were next washed twice with washing buffer (3 M KCl, 50 mM Tris-HCl pH 7.2), resuspended in 2xSB, boiled for 5 minutes and analysed on 8-12% (v/v) SDS-PAGE gels with PageRuler protein ladder (Promega). After electrophoresis, protein bands were visualized by Coomassie Blue staining.

2.2.11 Expression of *Sulfolobus solfataricus* PriS CTD

Plasmids for protein expression were transformed into *E.coli* Rosetta 2 (DEX3) [plusS] (Table 2.1.2) and transformant cells were obtained on Luria-Bertani (LB) medium supplemented with appropriate antibiotics (100 μg/ml of ampicillin and 34 μg/ml of chloramphenicol). Starter cultures were grown overnight and used to inoculate 250 ml of fresh LB. These were grown until an OD_{600nm} reached O.8, when protein expression was induced by adding 0.1 mM of IPTG. After three hours, cells were harvested at 3000 rpm for 10 minutes at 4^oC, pellet was resuspended in 1.8 ml ice-cold lysis buffer (depending on experiment 150-500 mM NaCl, 20mM sodium phosphate buffer with complete proteases inhibitor and 1 mM PMSF) and lysed by sonication 3 times for 15 seconds. Next, Triton was added to a final concentration of 1% and lysate was incubated on wheel for 30 minutes at 4^oC. After that time, lysate was clarified by centrifugation at 13000 rpm for 15 minutes at 4^oC. Supernatant was mixed with appropriate volume of Glutathione Sepharose 4B beads (GE Healthcare) and samples were incubated on wheel for 30 minutes at 4^oC. Beads were then centrifuged and washed 4 times with lysis buffer. Beads were resupended in 30 μl of sample buffer and boiled for 4 minutes followed by analysis on 10% SDS-PAGE gel. Protein bands were visualized by Coomassie blue staining.

2.2.12 Bioinformatics

All DNA sequence analysis was carried out using EnzymeX 3.1 (http://www.mekentosj.com/science/enzymex). Sequence alignments were performed using Clustal X 2.0.12 (Larkin et al., 2007). Amplicons for RT-PCR were chosen with the mfold Web Server (http://www.bioinfo.rpi.edu/applications/mfold/) that was used for predicting the secondary structure of RNA and DNA. Relative gene expression was analysed using Rest 2008 (Pfaffl et al., 2002). The structure modelling was performed using Phyre2 server (Bennett-Lovsey et al., 2008, Kelley and Sternberg, 2009).

Chapter 3

Conditional inactivation of replication genes

3.1 Introduction

As chromosomal DNA replication is a vital process for all living organisms, genes involved in that process are essential for cell viability. Therefore, construction of conditional lethal mutants of replication genes would be a very useful tool for functional and structural analysis of the corresponding proteins. This goal may be achieved by using a tightly controlled gene promoter. An ideal promoter should be induced in some simple, reversible manner, such as manipulation of the growth condition, and be inactive, or show very poor activity, when uninduced. At the same time, whatever causes promoter activation should not influence the expression of any other gene in the cell. Development of such a promoter will be a valuable tool to study essential genes in *H.volcanii* in addition to the tools already available for molecular genetics in this archaeon.

Transcription analysis using DNA microarrays of cells grown in the presence and absence of several compounds, such as different carbon sources, drugs and heavy metals, lead to identification of the tryptophanase gene, HVO_0009, or *tnaA*, which is strongly expressed in the presence of tryptophan (Large et al., 2007). *H.volcanii* is a tryptophan prototrophic organism but tryptophan is the most costly amino acid to make so it is advantageous for the cell to regulate production of the tryptophanase enzyme (Lam et al., 1992, Lam et al., 1990). Regulated activity of the *tna* promoter was confirmed by studying expression of three independent reporter genes (*pyrE2*, *bgaH* and *cct1*) (Large et al., 2007).

The initial aim of this project was to assess the feasibility of using the *tna* promoter to generate conditional-lethal (promoter shut-off) mutants of several replication genes (listed in Table 3.1).

Table 3.1 Selected components of *H.volcanii* chromosomal DNA replication machinery

ORF i.d.	Gene	Chromosome location	Protein	Role in DNA replication/Notes
HVO_0858	*polB1*	770425- 774432F	DNA polymerase B	Essential gene in *Halobacterium* sp. NRC-1 (Berquist et al., 2007). Exact role in DNA replication unknown
HVO_0003	*polD1*	2942 – 4543F	DNA polymerase D, catalytic subunit	Essential gene in *Halobacterium* sp. NRC-1 (Berquist et al., 2007). Exact role in DNA replication unknown
HVO_0065	*polD2*	64044 – 67652F	DNA polymerase D, non- catalytic subunit	Essential gene in *Halobacterium* sp. NRC-1 (Berquist et al., 2007). Exact role in DNA replication unknown
HVO_0175	*pcna*	157928 – 158671R	PCNA, the sliding clamp	Essential gene in *Halobacterium* sp. NRC-1 (Berquist et al., 2007).
HVO_0220	*mcm*	198868 – 201976F	MCM, DNA helicase	Essential gene in *Halobacterium* sp. NRC-1 (Berquist et al., 2007).
HVO_0203	*rfcA*	181313 – 182296F	Replication factor C, small subunit	
HVO_2427	*rfcB*	2295031- 2296497F	Replication factor C, large subunit	
HVO_0145	*rfcC*	135168 - 136193F	Replication factor C, small subunit	
HVO_2697	*priS*	2544015- 2545172F	DNA primase, small subunit	Essential gene in *Halobacterium* sp. NRC-1 (Berquist et al., 2007).

HVO_0173	*priL*	155952 - 157052R	DNA primase, large subunit	Essential gene in *Halobacterium* sp. NRC-1 (Berquist et al., 2007).
HVO_1338	*rpaA1*	1218023- 1219306R	Replication protein A, putative ssDNA binding-protein	Non-essential gene in *H.volcanii* in the presence of RpaB, as shown in this study. Exact role in DNA replication unknown
HVO_0292	*rpaB1*	261191 - 262126R	Replication protein B, putative ssDNA binding-protein	Non-essential gene in *H.volcanii* in the presence of RpaA, as shown in this study. Exact role in DNA replication unknown
HVO_0519	*rpaC*	453269 - 454720F	Replication protein C, putative ssDNA binding-protein	An essential gene in *H.volcanii* as shown in this study. Exact role in DNA replication unknown.
HVO_2698	*ginS*	2545169- 2546164F	DNA replication factor GINS	
HVO_1565	*ligA*	1432856 – 1435613F	DNA ligase, ATP dependent	Non-essential gene in *H.volcanii* in the presence of LigN (Zhao et al., 2006)

3.2 Optimizing conditions to use *tna* promoter system

Among the three types of media that are commonly used for *Haloferax* cultivation, only minimal medium (HvMin) does not contain tryptophan. We tested growth of wild-type *H.volcanii* DS70 in minimal medium supplemented with different concentrations of tryptophan (0.075 mM, 0.1 mM, 0.25 mM and 2.5 mM) (Figure 3.1). We found that tryptophan at concentration above 0.25 mM reduces the growth rate. This effect is most clearly seen when cells were grown in HvMin medium supplemented with 2.5 mM tryptophan. We also tested growth of *H.volcanii* in HvCa medium. HvCa medium, which is based on casamino acids, contains a low level of tryptophan, enough to activate p*tna* (Large et al., 2007). Growth of *Haloferax* cells was also reduced when additional tryptophan was added to a final concentration of 2.5 mM (data not shown).

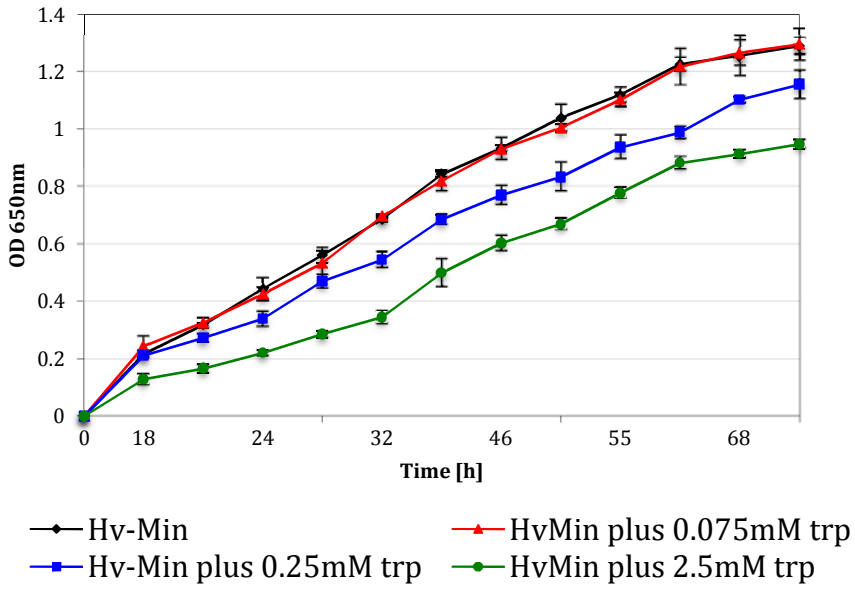


Figure 3.1 Growth of *H.volcanii* DS70 on HvMin medium supplemented with different concentrations of tryptophan

H.volcanii DS70 strain was grown in Hv-Min with different concentrations of tryptophan and growth was monitored by measuring absorbance at 650 nm. Mean and standard deviation of three independent experiments are shown.

3.3 *tna*-plasmid integrant strains

3.3.1 Construction of *tna*-plasmid integrant strains

In order to test whether the *tna* promoter could be used for conditional replication arrest, a promoter replacement strategy was used. This strategy leads to generating plasmid-integrant strains with a 5'-truncated gene with its native promoter intact and a full-length gene under p*tna*. This was achieved by constructing an integrating plasmid pNPM01 (PL9, Table 2.3) carrying the 124 bp *tna* promoter between *Bcl*I and *Nde*I restriction sites and a synthetic operon consisting of *H.volcanii pyrE*2 and *hdr*B genes under control of the *H.salinarum* ferredoxin promoter (p*fdx*). *pyrE2* and *hdrB* are the auxotrophic markers for uracil and thymidine/hypoxanthine biosynthesis, respectively. Next, the 5' ends (285-630 bp fragments) of the selected replication genes were inserted downstream of p*tna*, giving plasmids PL9-PL22 (pNPM01-tna-polB, etc, Table 2.3)(Figure 3.2). These plasmids were constructed by Dr Stuart MacNeill, prior to the commencement of this PhD study. Details about size of the inserts and restriction sites are given in Materials and Methods chapter (Material and Methods, 2.2.7.1.1).

It should be noted that gene encoding small subunit of DNA primase, HVO_2697, overlaps with gene encoding replication factor GINS, HVO_2698, and both genes are likely to be co-transcribed. For that reason placing the *tna* promoter upstream of *priS* and growing cells in medium lacking tryptophan will also interrupt expression of *ginS*.

Routine cloning steps were performed in *Escherichia coli* DH5α. Subsequently, the plasmids were passaged through *Escherichia coli* SCS110 (dam^- dcm^-) mutant for preparing unmethylated DNA as *H.volcanii* has a restriction system that recognizes adenine-methylated GATC sites, resulting in DNA fragmentation, which affects transformation efficiency.

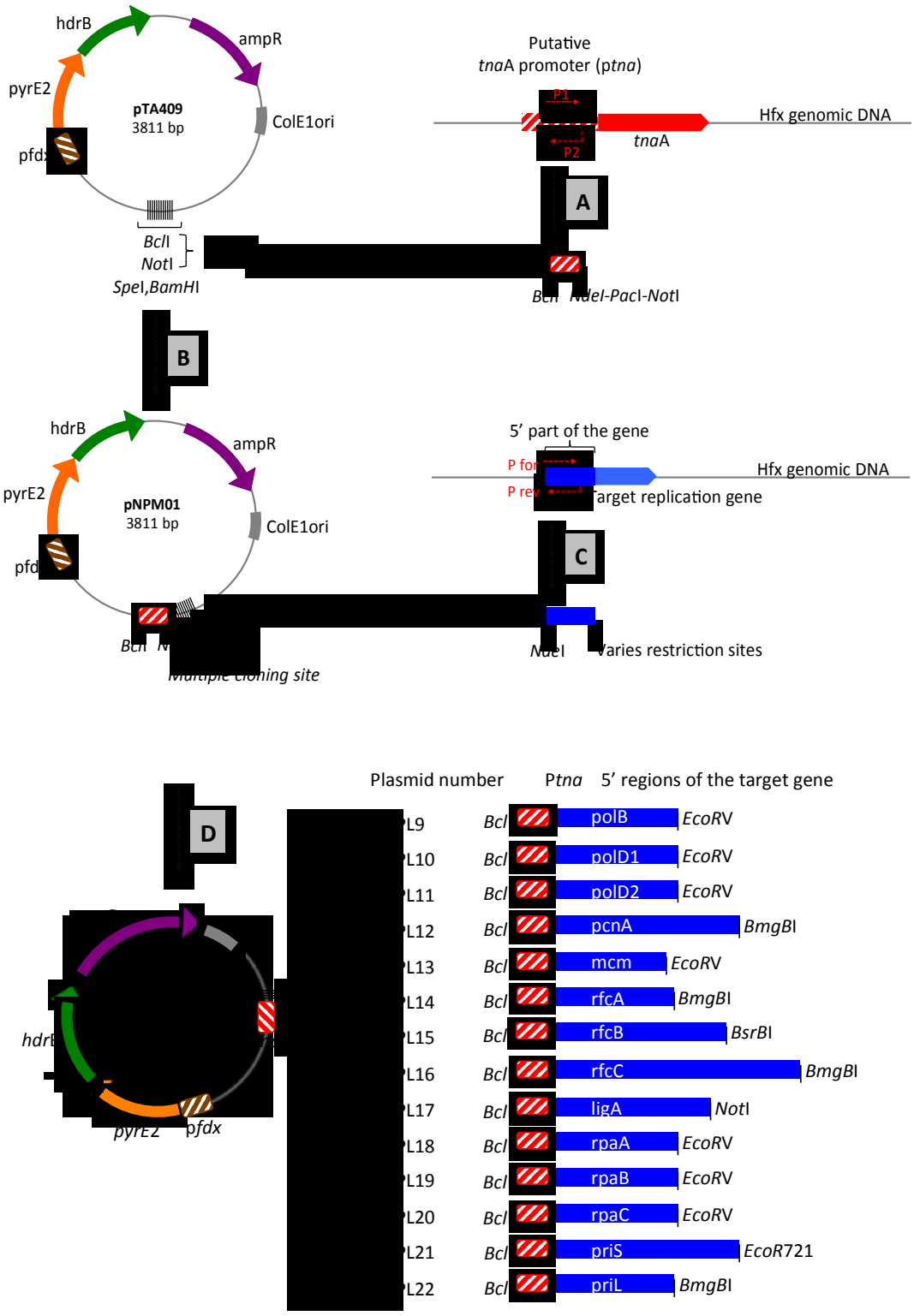


Figure 3.2 Construction of pNPM01-series plasmid for regulated expression

The *tna* promoter was amplified from *H.volcanii* DS70 genomic DNA (**A**) and inserted into pTA409 plasmid at *Bcl*I-*Nde*I sites, giving plasmid pNPM01 (**B**). 5' regions of selected replication genes were PCR amplified with introduced restriction sites (**C**) and inserted into pNPM01, immediately upstream *tna* promoter (**D**), giving plasmids PL9-PL22. See Materials and Methods for further details.

pNPM01-series plasmids carrying the *tna* promoter and 5' regions of selected replication genes were transformed into *H.volcanii* H98 strain (Δ*pyrE2* Δ*hdrB*) and transformant colonies obtained on HV-Ca medium without additional supplements to select for *pyrE2*$^{+}$ *hdrB*$^{+}$ cells (strains *polB*::[pNPM-tna-PolB], etc, SMH719-SMH732, Table 2.1). HvCa medium contains low level of tryptophan, enough to activate *tna* promoter (Large et al., 2007). Plasmid integration at correct locus was confirmed by PCR, using a forward primer located on p*tna* and reverse primers located on target genes on chromosome (Figure 3.3).

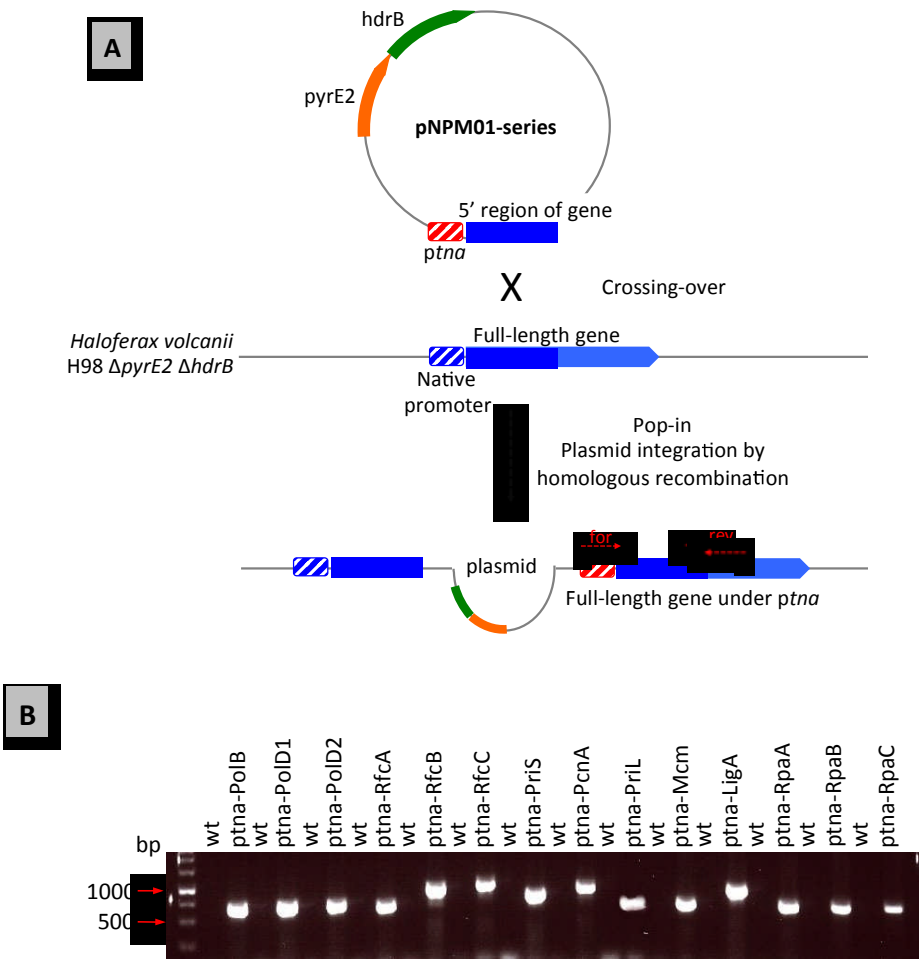


Figure 3.3 Construction of *tna* plasmid-integrant strains

A Schematic diagram of construction *tna* plasmid-integrant strains. *H.volcanii* H98 Δ*pyr* Δ*hdrB* was transformed with pNMP01-series plasmids carrying *tna* promoter, 5' region of selected replication genes and two selectable markers *pyrE2* and *hdrB*. Plasmids were integrated into chromosome via homologous recombination between the truncated gene on the plasmid and the full-length gene on the chromosome. Plasmid integration at the correct locus was confirmed by PCR. Red arrows indicate position of screening oligonucleotides.
B PCR screening of *tna*-plasmid integrant strains. Forward oligonucleotide is located within *tna* promoter (P1) and reverse oligonucleotide is located within target gene, at the position 500 or 750, depending on the gene (P21-P34).

3.3.2 Phenotypic analysis of *tna* plasmid-integrant strains

The ability of the *tna* plasmid-integrant strains to grow in the absence of tryptophan was tested using the spotting technique. Strains were grown in liquid Hv-Min medium supplemented with 0.075 mM tryptophan to mid-exponential phase (OD_{650} 0.5). After pelleting and washing, serial diluted cells were spotted onto solid Hv-Min agar plates lacking tryptophan and, as a control, onto Hv-Min plates supplemented with 0.075 mM tryptophan (for details, see Materials and Methods). This tryptophan concentration was chosen based on previous growth analysis (see section 3.1), as it was shown to not influence growth of the wild-type *H.volcanii* and presumably is high enough to induce *tna* promoter. On plates lacking tryptophan, little or no growth was seen for *tna*-PolB-, *tna*-RpaC- and *tna*-PriS-plasmid integrant strains. *Tna*-PCNA-, *tna*-Mcm, *tna*-RfcA-, *tna*-RfcB- and *tna*-RfcC plasmid-integrant strains showed a reduction in growth compared to the wild-type but less than what was seen with *tna*-PolB, *tna*-RpaC and *tna*-PriS, whereas *tna*-PolD1, *tna*-PolD2, *tna*-PriL, *tna*-LigA, *tna*-RpaA and *tna*-RpaB strains grew equally well in the presence and absence of tryptophan (it should be noted that LigA, in the presence of LigN, is non- essential protein, so no reduction in growth is expected) (Figure 3.4). Interestingly, *tna* PCNA plasmid-integrant strain showed a reduction in growth on plate supplemented with tryptophan suggesting that overexpression of PCNA is toxic to cells (see section 3.2.4).

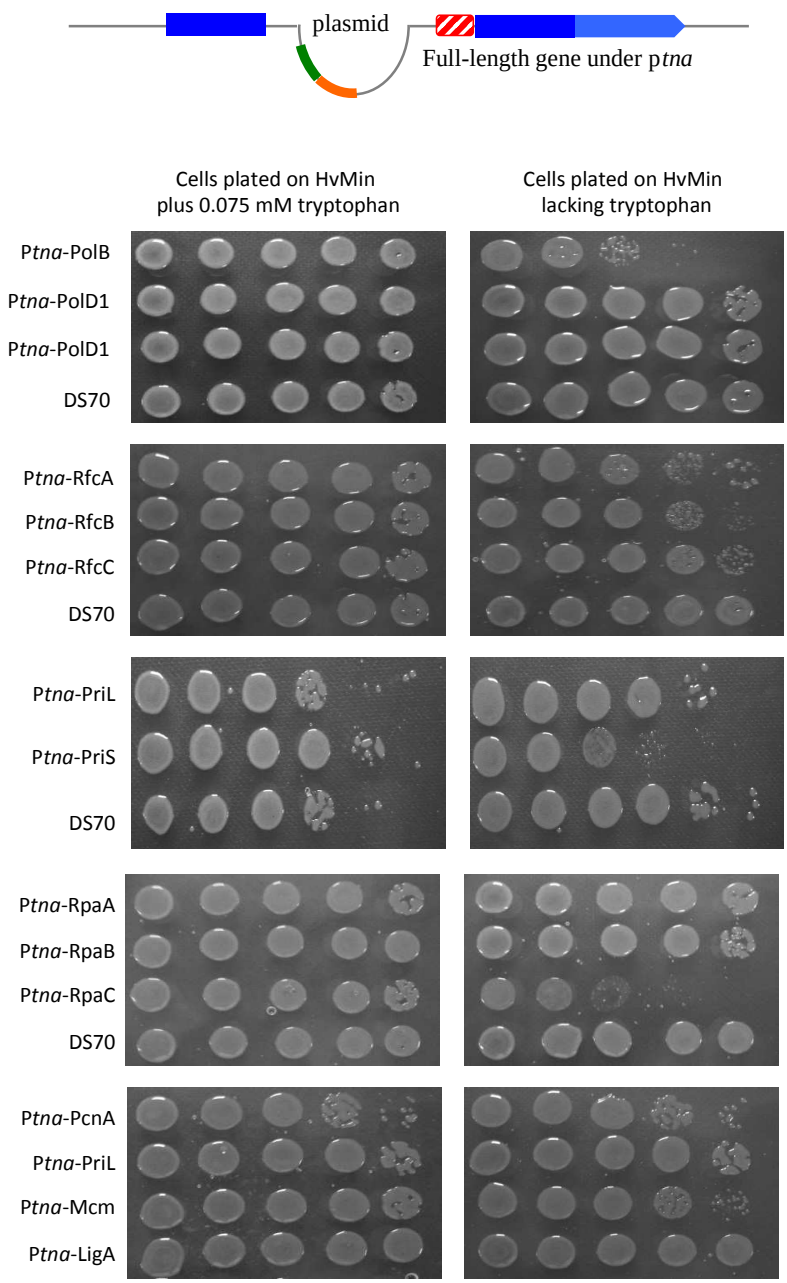


Figure 3.4 Growth of the *tna* plasmid-integrant strains in the presence and absence of tryptophan

Strains were grown to mid-exponential phase in HvMin medium supplemented with 0.075 mM tryptophan. The cells were then washed and 10-fold serial dilutions spotted into HvMin plates with (right panels) or without (left panels) tryptophan and incubated at 45^{o}C for 5 days. Strain DS70 is a wild-type control.

For the three mutant strains that showed inhibited growth on plates lacking tryptophan, growth curves were determined (Figure 3.5). Cells were grown in minimal liquid medium supplemented with 0.075 mM tryptophan to mid-exponential phase (OD_{650nm} 0.5) and then cells were washed to remove tryptophan completely, split into Hv-Min media with and without tryptophan and grown for seven hours. Growth rate was monitored by measuring the absorbance at 650 nm. All three strains grew equally well compared to each other and to the wild-type in Hv-Min with tryptophan, whereas growth in medium without tryptophan was significantly inhibited. Interestingly, the *tna*-PolB-plasmid integrant strain was unable to grow in the absence of tryptophan and *tna*-RpaC strain showed growth inhibition only after 5 hours. Similar analysis has been performed for *tna*-PolD1, *tna*-PolD2, *tna*-RpaA1, *tna*-RpaB1 and *tna*-PriL strains but all were indistinguishable from the wild-type (data not shown).

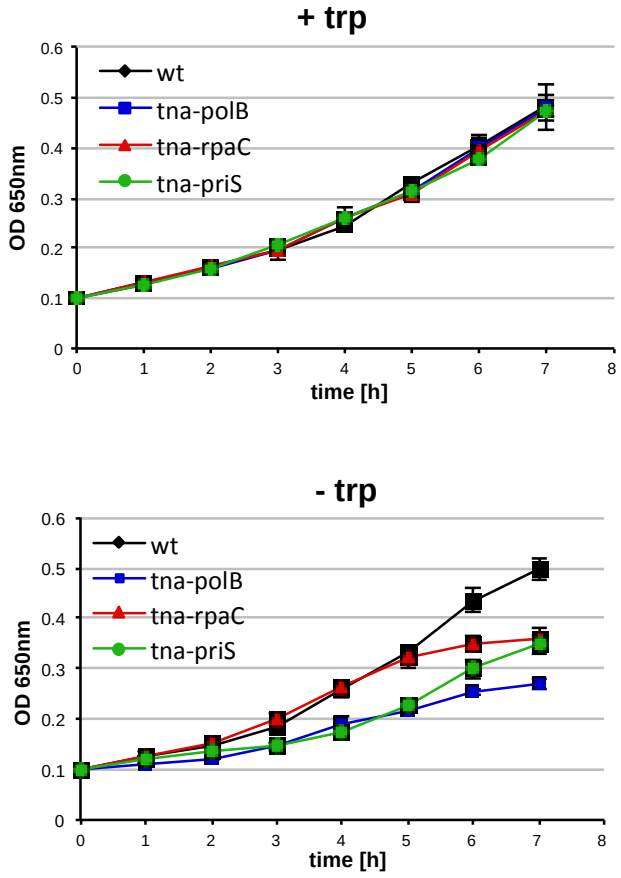


Figure 3.5 Growth curves of selected *tna* plasmid-integrant strains

H.volcanii strains were grown in Hv-Min with or without 0.075 mM tryptophan (trp) and growth was monitored by measuring absorbance at 650 nm. Mean and standard deviation of three independent experiments are shown

3.3.3 Analysis of gene expression by qRT-PCR

To confirm that the observed growth inhibition of *tna* plasmid-integrant strains is caused by turning-off expression of the targeted genes, quantitative reverse transcription real-time PCR (qRT-PCR) was used. All experimental procedures are described in details in Materials and Methods section (Materials and Methods 2.2.5). Amplicons were designed, using the mfold Web Server (http://mfold.rna.albany.edu/?q=mfold), to be 75-200 bp long. Because the genome of *H.volcanii* is characterised by a high G+C content (65%) that might affect the efficiency of the real-time PCR reactions, care was taken to choose optimal template and primer locations for all genes of interest (see Table 3.2). To validate and optimise reactions, standard curves were constructed using 10- fold dilutions of *H.volcanii* chromosomal DNA and a temperature gradient feature was used to find the optimal annealing temperatures (Table 3.2).

Gene expression analysis was performed for *polB*, *polD1*, *polD2* and *rpaC* genes in the *tna* plasmid-integrant strains and wild-type DS70 background as a control. Despite testing two of the most promising template regions (161-280 and 841-974), six primers and several annealing temperatures, it was not successful to optimized assay for *priS* gene (data not shown). This might be caused by very high G+C content of the gene (68.7%).

Table 3.2 Experimental design for gene expression analysis

Gene	Template region	Primers*	Annealing temperature	Efficiency parameters	
				E	R^2
polB	2390-2530	PL68-69	60.5 ^{o}C	102.8%	0.996
polD1	610-720	PL70-71	60.5^{o}C	90.1%	0.998
polD2	2841-2940	PL72-73	60.5^{o}C	99.8%	1.000
RpaC	490-590	PL74-75	64.2^{o}C	90.0%	0.997
rps$^{\#}$	231-360	PL76-77	60.0^{o}C	91.1%	0.990

***** Oligonucleotide table, Apendix

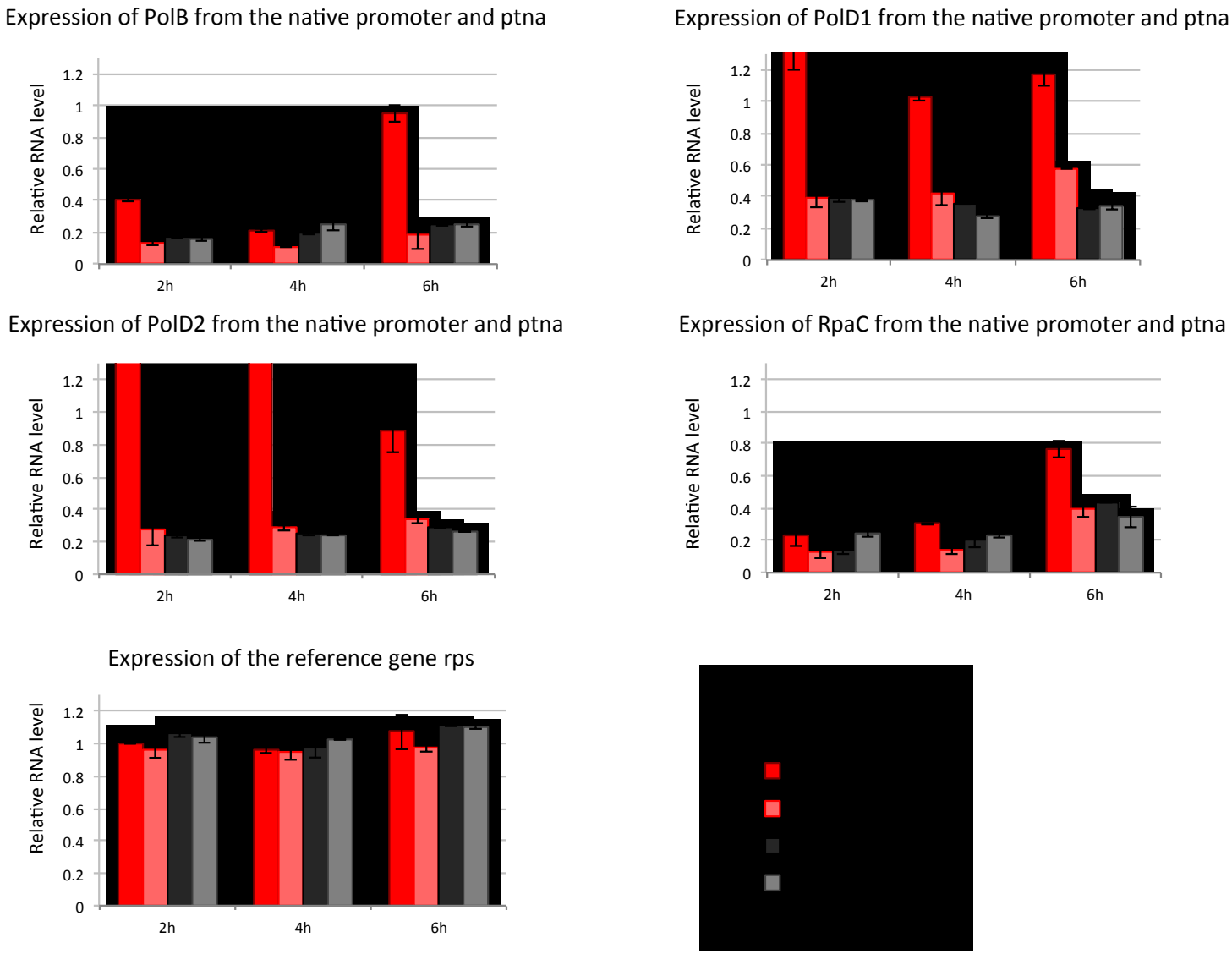
$^{\#}$ internal standard

In all qRT-PCR experiments cells were grown in Hv-Min supplemented with 0.075mM tryptophan to mid-exponential phase. Next, the cells were washed to

remove tryptophan and used to inoculate fresh cultures in Hv-Min and HvMin supplemented with 0.075mM tryptophan. Cells were grown for eight hours and samples for RNA isolation were taken every two hours. RNA isolation was carried out with RNase kit (QIAGEN) and RNA at a concentration of 10 ng/µl was used as a template. The gene encoding the small subunit of ribosomal protein HVO_0561 (*rps*), as a housekeeping gene constitutively expressed in the cells, has been chosen as internal standard. The expression level of that gene was analysed in *H.volcanii* DS70 and *polB* tna-integrant strain in media either containing and lacking tryptophan and, as expected, no differences in expression level were seen under these two conditions (Figure 3.6, bottom panel). Data analysis was conducted with REST 2008 software (Pfaffl et al., 2002), calculating Ct values from the intersection of a threshold line with the early exponential interval of the fluorescence curve. Relative expression levels were calculated according to the Pfaffl method (Pfaffl, 2001). Expression levels were normalized to overnight culture and to the corresponding internal standard. Figure 3.6 presents relative expression levels of *polB*, *polD1*, *polD2* and *rpaC* genes in the *tna* plasmid-integrant strains and *H.volcanii* wild-type DS70 in the presence and absence of tryptophan. For all genes tested, 2-5-fold decrease in transcript level was seen when cells were grown in Hv-Min medium lacking tryptophan. After eight hours transcript levels were reduced by 80%, whereas in cells grown in the presence of tryptophan genes were transcribed at nearly constant level (data not shown). In Hv-Min medium with no tryptophan supplementation, the *polB* and *rpaC* mRNA levels were reduced to below the level seen in the wild-type cells, consistent with the view that growth arrest after tryptophan removal is caused by reduction in gene expression. In opposition, the *polD1* and *polD2* mRNA levels were still above the levels seen in wild-type cells what might explain why both, polD1 and polD2 tna-plasmid integrant strains showed no phenotype. At the same time we observed that expression levels of the *tna*-plasmid integrant strains in medium with 0.075 mM tryptophan were elevated above the normal wild-type levels (in the *tna-polB* and *tna-rpaC* strains elevated mRNA levels were seen after 6h of growth). Since there in no detectable difference in growth rate between the wild-type and the integrated *tna* strains overexpression does not appear to have a detrimental effect on cells.

Figure 3.6 Analysis of transcript level of selected *tna* plasmid-integrant strains

Cells were grown in Hv-Min ± 0.075 mM tryptophan. The course of expression was followed for 8 hours by measuring the OD and collecting samples for RNA isolation and qRT-PCR. All relative expression levels were normalized to overnight culture on Hv-Min + tryptophan.



3.3.4 Phenotypic analysis of *pcna tna* plasmid-integrant strain

Phenotypic analysis of original fifteen *tna* plasmid-integrant strains indicated reduction in growth rate of the strain expressing PCNA from *tna* promoter on plates supplemented with 0.075 mM tryptophan. Transcriptomics studies of *tna* polB-, *tna* polD1-, *tna* polD2 and *tna* rpaC plasmid-integrant strains showed that growing cells in given concentration of tryptophan results in gene overexpression. If the same is true for *tna* PCNA, it might explain the observed phenotype. It was shown in yeast and mammalian cells, that overexpression of PCNA displays a toxic effect; it blocks cell cycle progression (Waseem et al., 1992, Piard et al., 1998) and inhibits DNA repair (Shan et al., 2003).

To order to find conditions that will activate *tna* promoter but not cause PCNA overproduction and the resulting inhibition of cell growth, we spotted *tna-pcnA* cells on HvMin medium supplemented with different concentration of tryptophan: 0.025 mM and 0.05 mM. As seen in Figure 3.7, cells grown on medium containing lower levels of tryptophan grew better in comparison to plates with 0.075 mM tryptophan. That experiment shows that expression level from *tna* promoter can be modified simply be changing tryptophan supplementation. This might be useful in situations, such as for the *pcna* gene, when overproduction of the protein has detrimental effect on the cells.

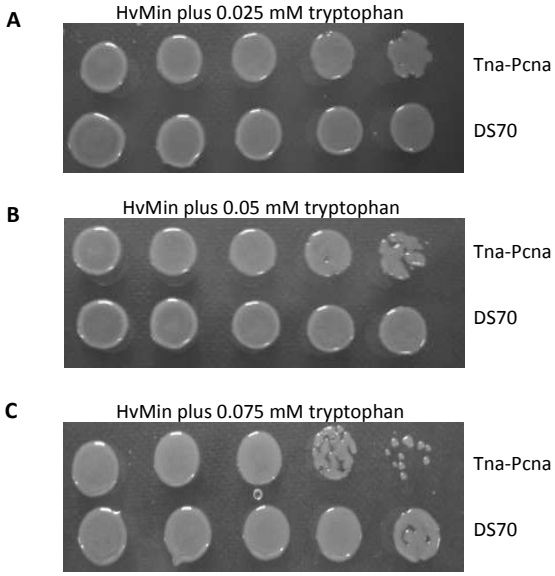


Figure 3.7 Growth of the *tna* pcnA plasmid-integrant strains in the presence of different tryptophan concentration

Cells were grown to mid-exponential phase in HvMin medium supplemented with 0.025 mM, 0.025 mM or 0.075 mM tryptophan. The cells were then washed and 10-fold serial dilutions spotted onto HvMin plates supplemented with corresponding concentration of tryptophan (**A**, **B** and **C**, respectively). DS70 strain is a wild-type control.

3.4 *tna* promoter replacement strains

Results shown in the previous section indicate that *tna* promoter can be successfully used for down-regulation of some, but not all, replication genes in *H.volcanii*. However, the plasmid integration strategy creates genetically unstable strains in which homologous recombination between the 5' truncated ORF and the full-length gene on the chromosome will lead to reversion to wild-type. In order to prevent this, for three genes that showed strong phenotype with plasmid integration strategy (*polB*, *rpaC* and *priS*), genetically stable *tna* promoter replacement strains were constructed (*polB*::*tna-polB*, *rpaC*::*tna-rpaC*, *priS*::*tna-priS*, SMH737-SMH739, Table 2.1). The pop-in/pop-out method was used, in which the native promoter was replaced with p*tna* (see Materials and Methods 2.2.7.1.2 for details of strain construction). Briefly, 500 bp fragments carrying the upstream flanking region of an appropriate gene were amplified by PCR and inserted into corresponding pNPM-*tna* plasmids at *Bcl*I site upstream of *tna* promoter, giving pNPM-Bcl series plasmids PL24-PL26 (Table 2.3)(Figure 3.8).

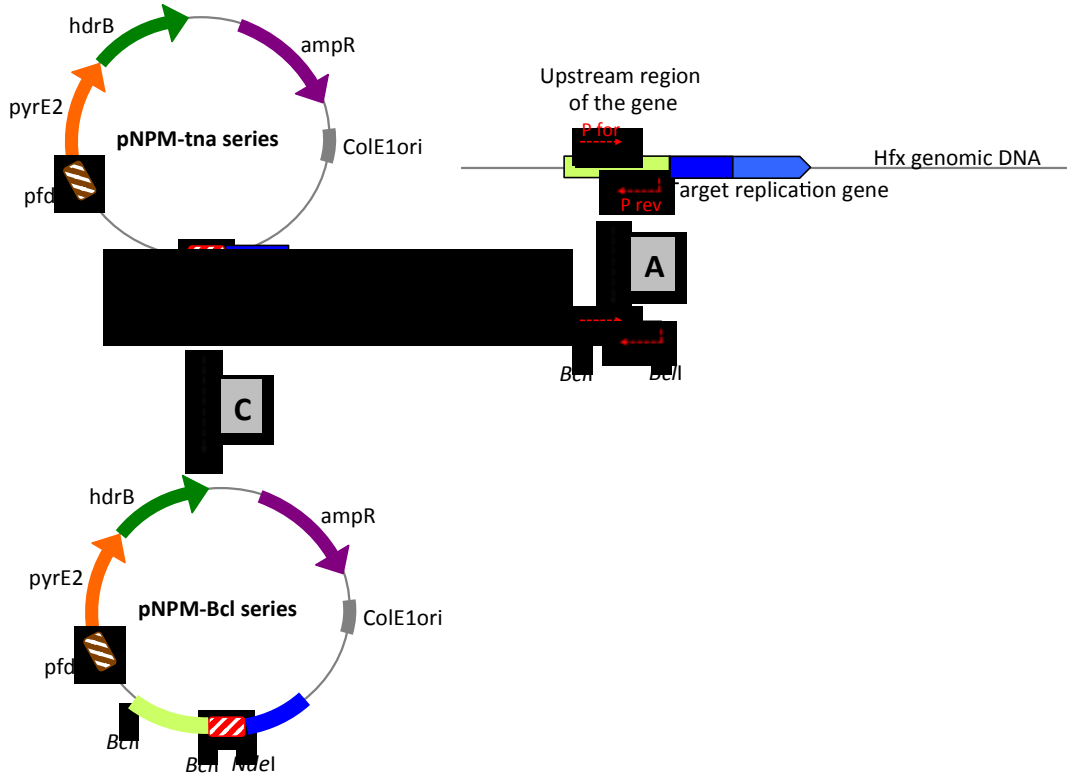


Figure 3.8 Construction of pNPM-Bcl series plasmids for regulated expression

A 500 bp regions from upstream of the genes were amplified from *H.volcanii* DS70 genomic DNA as *Bcl*I fragments (**A**) and inserted into pNPM-tna plasmid series at *Bcl*I site (**B**), giving pNPM-Bcl series plasmid PL24-PL26 (**C**).

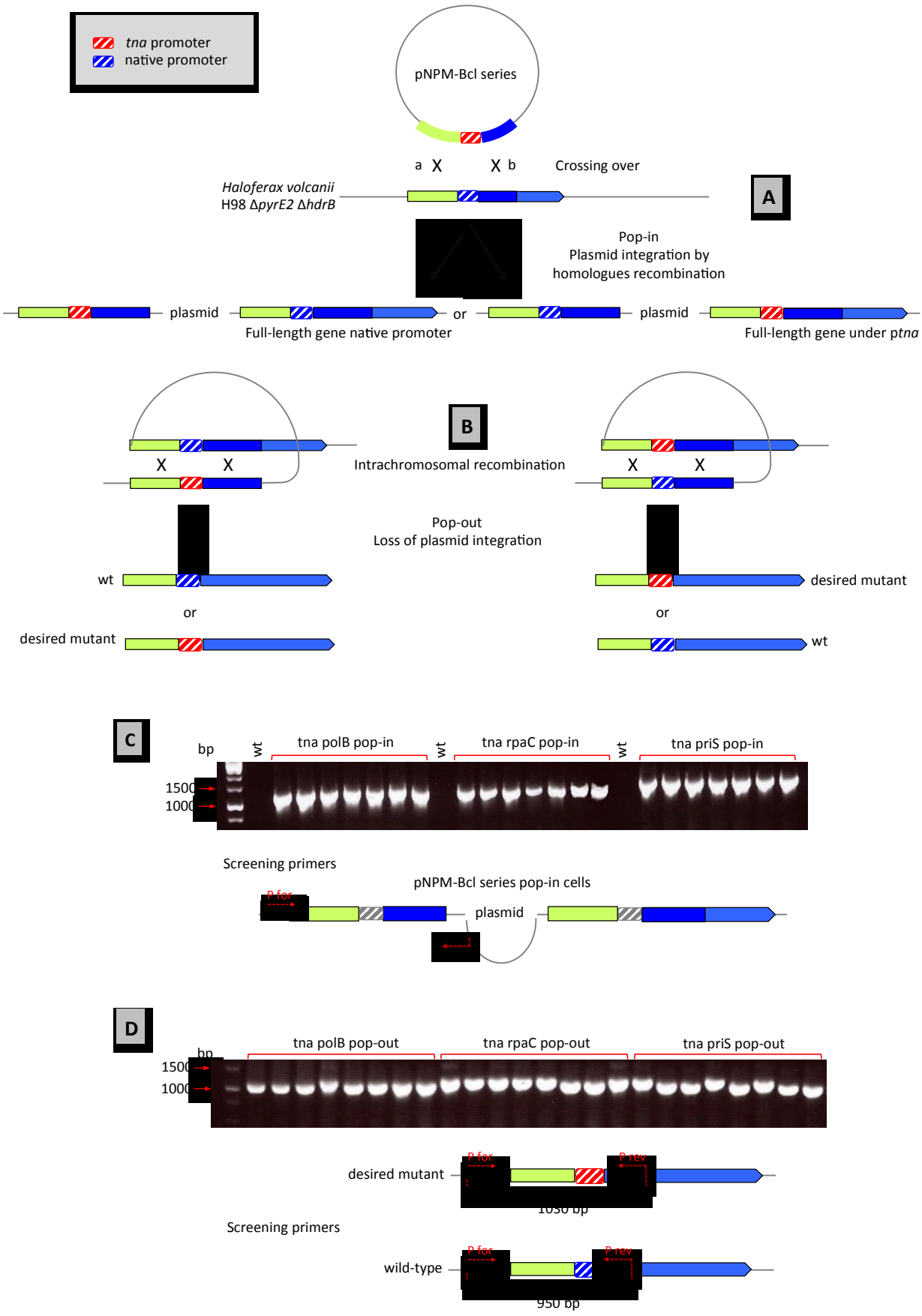
Plasmids PL24-PL26 were transformed into *H.volcanii* H98 strain (Δ*pyrE2* Δ*hdrB*) (this is the pop-in step, Figure 3.9A) and *pyrE2*$^{+}$ *hdrB*$^{+}$ transformants were obtained on Hv-Min plates supplemented with 0.075 mM tryptophan. Plasmid integration was verified by PCR using a forward primer located on the chromosome 650 bp upstream on the target gene and a reverse primer located on plasmid (Figure 3.9C). Once plasmid integration had been confirmed, selection was relieved by plating cells on Hv-Min supplemented with uracil and thymidine/hypoxanthine (this is the pop-out step, Figure 3.9B). Intramolecular recombinants that have lost the plasmids were counter-selected using 5-fluoroorotic acid, which is converted to a toxic 5-fluorouracil in ura^{+} but not ura^{-} cells. Depending on the orientation of the recombination event, the target gene is now located 3' to its native promoter (reversion to the wild-type) or 3' to the *tna* promoter. To distinguish the desired strains from the wild-type, pop-out cells were screened by PCR using a forward primer located 650 bp downstream of the target gene and a reverse primer located within the gene, at the position 300 bp from the start codon. This screening is based on the difference in size between the *tna* promoter (124 bp) and the native promoters (50 bp) (Figure 3.9D).

Figure 3.9 Construction of *tna* promoter replacement strains

A and **B** Schematic diagram of construction of *tna* promoter replacement strains by using pop-in/pop-out method. For simplicity, marker genes and selection approach were omitted. *H.volcanii* H98 was transformed with pNMP-Bcl series plasmids carrying 500 bp upstream region of an appropriate gene, *tna* promoter and 5' region of the gene. The plasmids may be integrated into chromosome in two possible orientations (A a and b). Loss of the plasmid by intrachromosomal recombination may result in desired mutant or reversion to the wild-type (B).

B PCR screening of pop-in cells. Integration of the plasmid at correct chromosomal loci was confirmed using forward primer located 650 bp upstream on the target gene and reverse primer located on plasmid (P43-P46). Wild-type DNA is a negative control.

C PCR screening of pop-out cells. Wild-type and desired *tna* promoter replacement strains were distinguished using a forward primer located 650 bp upstream on the target gene and a reverse primer located within target gene, at the position 300 bp from the start codon, which give a PCR product approximately 80 bp bigger for the mutant than for the wild-type.


tna promoter
native promoter
pNPM-Bcl series
a X X b
Crossing over
Haloferax volcanii
H98 ΔpyrE2 ΔhdrB
A
Pop-in
Plasmid integration by
homologues recombination
plasmid
or
plasmid
Full-length gene native promoter
Full-length gene under ptna
B
X X
Intrachromosomal recombination
X X
Pop-out
Loss of plasmid integration
wt
or
desired mutant
desired mutant
or
wt
C
bp
wt
tna polB pop-in
wt
tna rpaC pop-in
wt
tna priS pop-in
1500
1000
Screening primers
pNPM-Bcl series pop-in cells
P for
plasmid
D
bp
tna polB pop-out
tna rpaC pop-out
tna priS pop-out
1500
1000
desired mutant
P for
P rev
1050 bp
Screening primers
wild-type
P for
P rev
950 bp

The properties of the *tna* promoter replacement strains were re-tested by the spotting technique. As expected, the stable *tna-polB*, *tna-rpaC* and *tna-priS* strains also showed reduced growth rate on medium lacking tryptophan while growth on medium supplemented with tryptophan was indistinguishable from the wild-type (Figure 3.10).

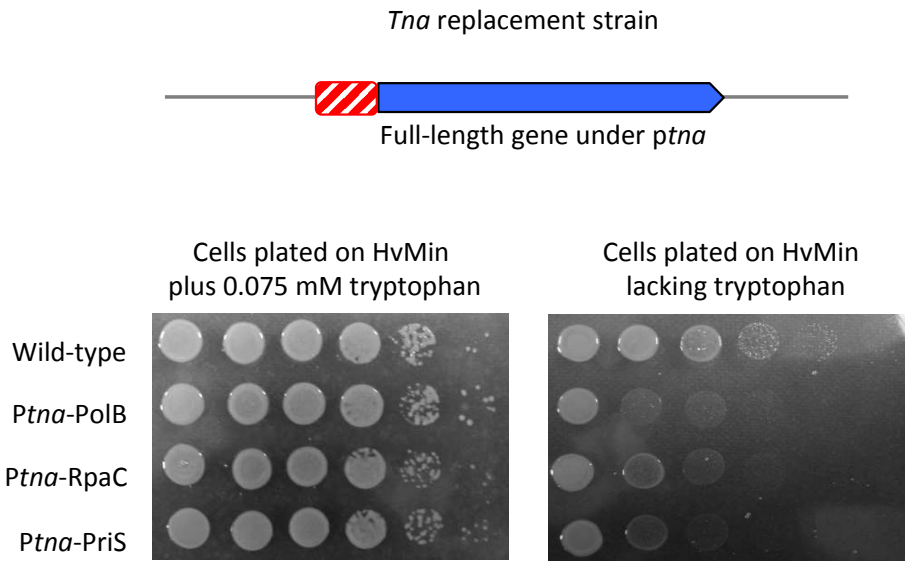


Figure 3.10 Growth of the *tna* promoter replacement strains in the presence and absence of tryptophan

Strains were grown to mid-exponential phase in HvMin medium supplemented with tryptophan (0.075 mM), uracil and thymidine/hypoxanthine. The cells were then washed and 10-fold serial dilutions spotted HvMin plates with (right panels) or without (left panels) tryptophan and incubated at 45^{o}C for 5 days. Strain H98 is a wild-type control.

3.5 Improving the *tna* promoter system

As shown in sections above, a 124 bp *tna* promoter originating from the 5' region of the *H.volcanii* tryptophanase gene is capable of directing conditional inactivation of three replication genes (*polB*, *rpaC* and *priS*), when cells are grown in medium lacking tryptophan. The *tna* plasmid-integrant strains were constructed for twelve genes believed to be essential for cell viability in the archaeon and for remaining nine genes no phenotype was observed on plates without tryptophan supplementation. Gene expression analysis by qRT-PCR performed for *tna*-PolB-, *tna*-PolD1-, *tna*-PolD2- and *tna*-RpaC plasmid-integrant strains revealed that transcription from *tna* promoter is indeed regulated by the presence of tryptophan in medium. However, even without tryptophan p*tna* is not completely inactivated and it provides low transcription level that for some genes (*polD1* and *polD2*) was above the level seen

in the wild-type cells. This might explain why a clear phenotype was not observed for all twelve replication genes in *tna* plasmid-integrant strains. To extend the usage of *tna* promoter system to more replication genes, attempts were made to construct mutant forms of the *tna* promoter that would still show induction by tryptophan but which would display reduced minimal/maximal expression levels. These attempts included producing truncated versions of p*tna*, placing the transcriptional terminator L11 upstream of *tna* promoter and mutating the TATA box consensus. Mutant *tna* promoters listed below were constructed as described in the following sections.

Table 3.3 Mutant *tna* promoter constructs

Promoter	Sequence changes
tna2	Truncation of 10 bp from 5' sequence
tna3	Truncation of 20 bp from 5' sequence
tna4	Truncation of 30 bp from 5' sequence
tna5	Truncation of 40 bp from 5' sequence
tnaL11	L11 terminator positioned upstream of *tna*
tnaM1	Single point mutation A to G at position -28
tnaM2	Single point mutation T to C at position -27
tnaM3	Single point mutation T to G at position -26

polB, *polD1* and *mcm* were chosen as a reporter genes for all experiments aimed at improving the *tna* promoter system. *tna*-polB plasmid-integrant strain showed a strong reduction in growth rate on plates lacking tryptophan and in these experiments serves as a control to ensure that newly constructed *tna* promoters are still activated by tryptophan. *tna*-mcm plasmid-integrant strain showed a moderate phenotype and therefore *mcm* is a promising candidate to show improvement in the system. *polD1* represents all the genes that did not show any phenotype in earlier experiments.

3.5.1 Construction of truncated forms of the *tna* promoter

Four truncated *tna* promoters, *tna*2-5, weere designed where the first is 10 nucleotides shorter at the 5' than the original and the next is 10 nucleotides shorter again, and so on. Promoters *tna*2-5 were constructed by PCR amplifying the appropriate fragments of *tna* promoter using oligonucleotide P51-P54 with P55 (Appendix, Table A1) and plasmid pNPM-tna-PolB as a template, digesting PCR products with *Bcl*I and *Nde*I and ligating into *Bcl*I- and *Nde*I-digested plasmid pNPM-tna, giving plasmids pNPM-tna2, etc (PL30-PL33, Table 2.3). Next, plasmids PL30-33 were digested with appropriate enzymes and ligated with 5' regions of *polB*,

polD1 and *mcm* ORFs excised from plasmids PL10, PL11 and PL14, respectively (Figure 3.11).

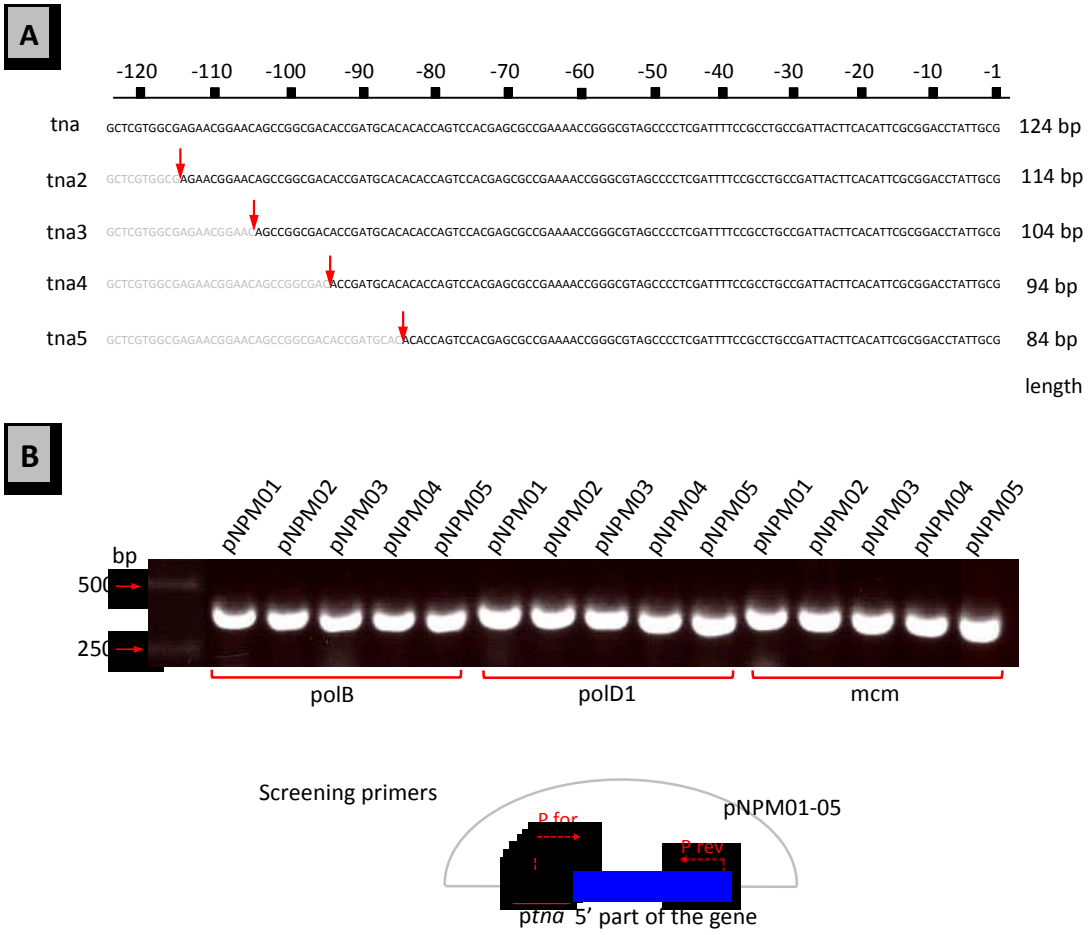


Figure 3.11 Construction of *tna2-5* truncated promoters

A The red arrows correspond to the map position of the first nucleotide of truncated promoters. Sequences shown in grey have been deleted from the wild-type promoter (first sequence).
B PCR screening of the plasmids carrying truncated *tna* promoters. Red arrows indicate position of the screening oligonucleotides. Plasmid names correspond to truncated promoter (pNPM02 carries *tna2* promoter, etc). Plasmid pNPM01 carries wild-type *tna* promoter.

Modifications made in p*tna* sequence were not sufficient to reduce overall transcription levels below the wild-type levels that would be seen as inhibition of cells growth.

3.5.2 Placing transcriptional terminator L11 upstream of *tna* promoter

The next approach to improve the *tna* promoter system involved placing a transcriptional terminator 5' to the *tna* promoter to preclude read-through

transcription from other promoters upstream of *tna*, which might produce a low level of transcript even when p*tna* is completely inactivated. A terminator sequence was previously identified upstream of the L11e ribosomal gene of *H.volcanii* (Shimmin and Dennis, 1996). Plasmids carrying the L11e transcriptional terminator positioned 5' to the *tna* promoter (pNPM-L11-PolB, -PolD1 and –MCM, PL46-48, Table 2.3) were constructed by PCR amplifying a L11e-*tna* fusion as a *Bcl*I-*Nde*I fragment using oligonucleotides P58 and P59 (Appendix, Table A1) and plasmid pTA927 as a template and cloning it into plasmids PL10, PL11 and PL14 that have been digested with *Bcl*I and *Nde*I.

Plasmids with truncated versions of *tna* promoter and L11e terminator were transformed onto *H.volcanii* H98 and growth of the transformant cells was re-tested on plates lacking and containing tryptophan (Figure 3.12).

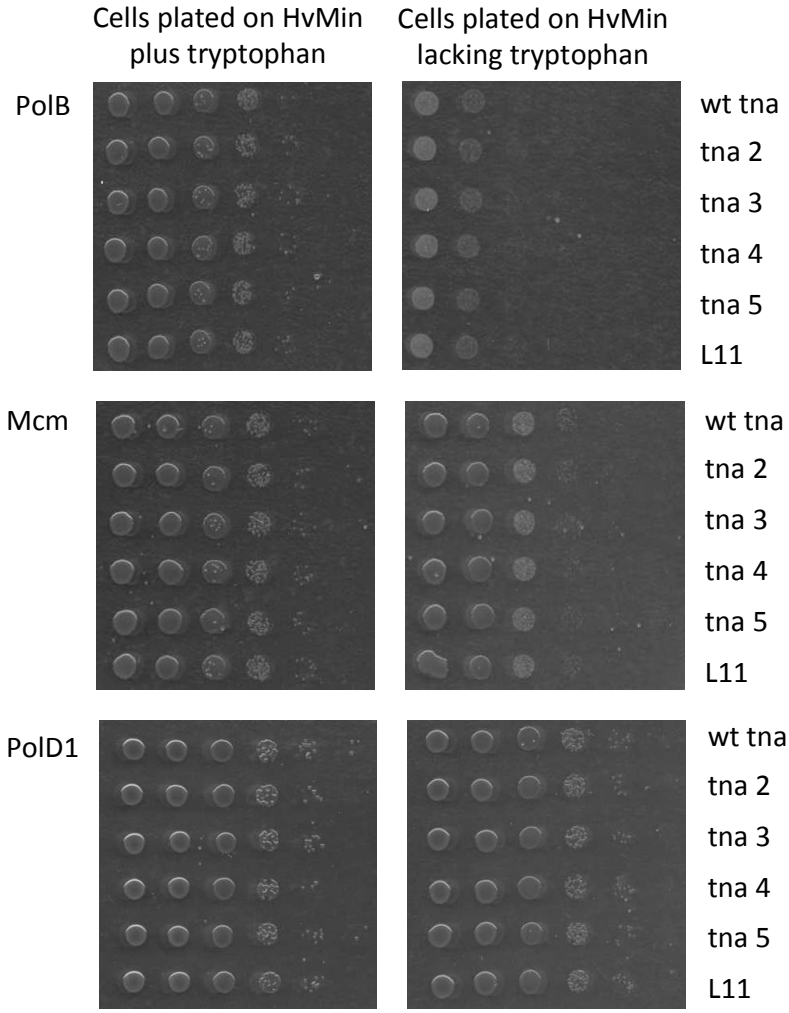


Figure 3.12 Growth of the strains with truncated *tna* promoters in the presence and absence of tryptophan

Strains were grown to mid-exponential phase in HvMin medium supplemented with 0.075 mM tryptophan. The cells were then washed and 10-fold serial dilutions spotted onto HvMin plates with (right panels) or without (left panels) tryptophan and incubated at 45°C for 5 days.

As shown on the spots panel above, none of the *tna* promoter modifications were successful. All the promoters still show induction by tryptophan, as *tna*-polB strains grow well on plates supplemented with 0.075 mM tryptophan but not on plates lacking tryptophan, but for two other genes no effect is observed in comparison to the wild-type *tna* promoter. Because placing the terminator sequence upstream of p*tna* did not improve results obtained for *mcm* and *polD2* genes, the activity of additional promoters is not responsible for read-through transcription. Modifications made in p*tna* sequence were not sufficient to reduce overall transcription levels below the wild-type levels that would be seen as inhibition of cells growth.

3.5.3 Mutating the p*tna* TATA box consensus

As described in the Introduction, the archaeal transcriptional apparatus resembles the eukaryotic one, including RNA polymerase and general transcription factors. Similarity is also found in the composition of promoter elements. A comparison of sequences located upstream of archaeal genes revealed the presence of two conserved elements: the TATA box (also called box A) and the B-factor Recognition Element (BRE or box B) (Peng et al., 2011). The TATA box is located approximately 25 bp upstream of the transcription start site and is rich in A+T residues. The BRE is a purine-rich region located immediately upstream of TATA-box. Mutational analysis of the TATA box of *Sulfolobus* sp. B12 16S/23S rRNA promoter (Reiter et al., 1990) and *H.volcanii* $tRNA^{Lys}$ promoter (Palmer and Daniels, 1995) showed that sequence is important in determining promoter efficiency *in vivo*.
Consequently, in order to construct improved versions of *tna* promoter, we decided to examine the transcriptional effects of single point mutations in the TATA box of this promoter.
First, we identified TATA-box within *tna* promoter sequence in accordance with the criteria that the element should be approximately 25 bp upstream of the transcription start site, it should be A+T rich, begin with a pyrimidine, and guanine should be excluded from the four adjacent 3' positions (Palmer and Daniels, 1995). The most likely p*tna* TATA-box is sequence 5'-ATTACTTC-3' located at position -28 to -21, with respect to the transcriptional initiation site (Figure 3.13).

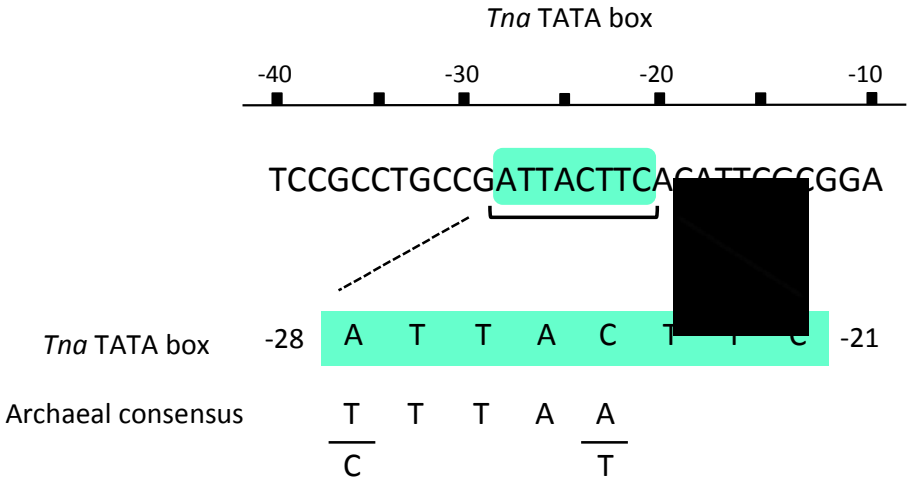


Figure 3.13 Location of the TATA box element in *tna* promoter

Three single point mutations in p*tna* TATA box element were generated: A to G at the position -28, T to C at the position -27 and T to G at the position -26. The new promoters carrying mutations were named *tna*M1, *tna*M2 and *tna*M3, respectively (Figure 3.14). These particular mutations were chosen according to the study on *H.volcanii* $tRNA^{Lys}$ promoter, where replacement of the wild-type nucleotides at the first three 5' sequence of the TATA box with G, C and G, respectively, resulted in moderate decrease in transcription efficiency (Palmer and Daniels, 1995).

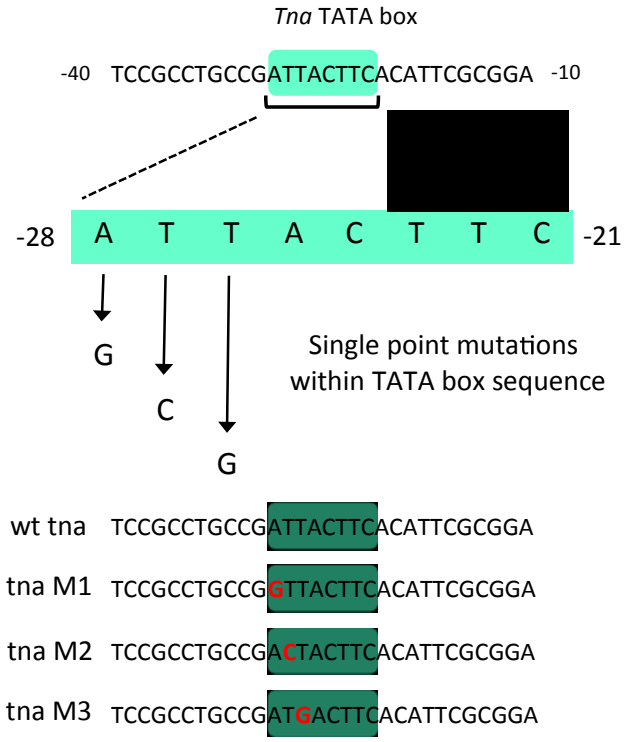


Figure 3.14 Mutations generated in *tna* promoter TATA box element

Constructs carrying the TATA box-mutated *tna* promoter versions were constructed by PCR overlap extension, using plasmid PL9 (Table 2.3) as a template, P60-P61 as an external oligonucleotides and P62-P67 as an internal oligonucleotides (Appendix, Table). See Material and Methods 2.2.9 for details. The external primers were located approximately 400 bp from 5' and 3' to the *tna* TATA box sequence. PCR products carrying desired mutations were inserted, as a *Bcl*I-*Nde*I fragments, into plasmid PL10 (pNPM-tna-PolB, Table 2.3), replacing the wild-type *tna* promoter. The resulting plasmids (pNPM-M1tna-PolB, etc., PL49-PL51, Table 2.3) were sequenced to ensure the absence of unwanted sequence changes. Next, plasmids PL52-59 (pNPM-M1tna-PolD1, etc., Table 2.3) were constructed by ligating *Nde*I-*Nco*I fragments carrying mutated *tna* promoter excised from plasmid PL49-PL51 and *Nde*I-*Nco*I fragments carrying 5' region of appropriate ORF. All newly constructed plasmids were transformed onto *H.volcanii* H98 and growth of the transformant cells was tested on plates containing and lacking tryptophan (Figure 3.15).

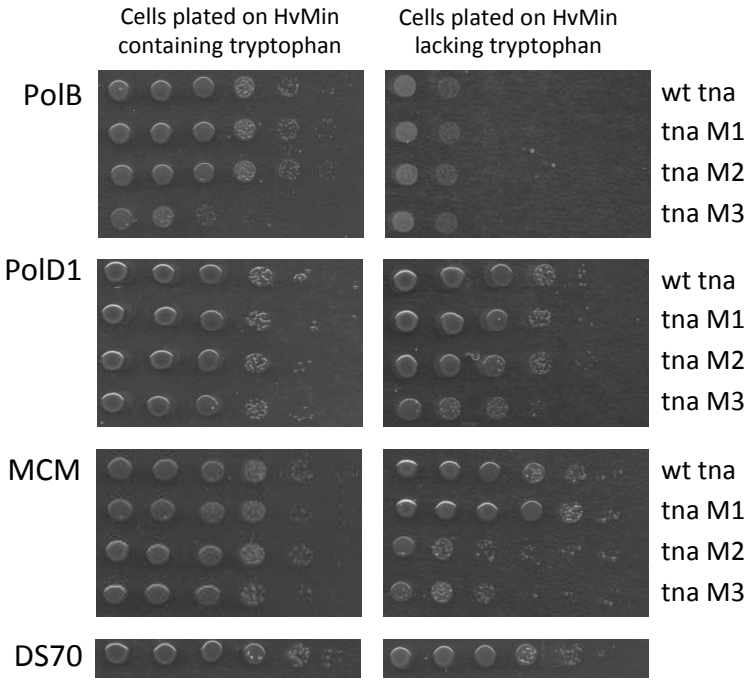


Figure 3.15 Growth of the cells expressing PolB, PolD1 and Mcm from mutated *tna* promoters in the presence and absence of tryptophan

Strains were grown to mid-exponential phase in HvMin medium supplemented with 0.075 mM tryptophan. The cells were then washed and 10-fold serial dilutions spotted onto HvMin plates with (right panels) or without (left panels) tryptophan and incubated at 45^{o}C for 5 days. DS70 is a wild-type control.

The results were as follows. On plates lacking tryptophan, little growth was seen for all *tna*-PolB plasmid-integrant strains. There was no significant difference in growth

rate between cells expressing *polB* from the wild-type *tna* and its mutated versions. Difference was seen on plates supplemented with tryptophan: *tna*M3-*polB* strain showed a reduction in growth in comparison to others *polB* strains and the wild-type. *tna-polD1-*, *tna*M1-*polD1-* and *tna*M2-*polD1-* plasmid-integrant strains grew equally well on plates containing and lacking tryptophan, while *tna*M3-*polD1* strain showed the desired growth inhibition on plates without tryptophan. The growth inhibition was not as strong as for *tna-polB* strains but a clear difference was seen between plates where *tna* promoter is activated and inactivated. Among *tna-mcm* plasmid-integrant strains, *tna*M2-*mcm* and *tna*M3-*mcm* strains showed a reduction in growth when cells were grown on plates lacking tryptophan.

Taken together, these results show that the *tna* promoter with a single substitution T to G at the position -26, designated *tnaM3,* can be successfully used for conditional down-regulation of *polD1* and *mcm* expression, extending the usage of *tna* promoter system to two other replication genes, alongside *polB, rpaC* and *priS*.

The promising results obtained for *tna*M3-*polD1-* and *tna*M3-*mcm* plasmid-integrant strains courage us to construct a strain where *polD2* was expressed from the *tnaM3* promoter. We also constructed *tna*M3-*rpaC* plasmid-integrant strain to compare its growth properties with *tna*M3-*polB* strain, which showed reduction in growth rate on plates supplemented with tryptophan. Plasmids carrying *tna*M3 promoter upstream to the 5' region of *polD2* and *rpaC* (PL58 and PL59, respectively), were constructed by ligating *Nde*I-*Nco*I fragments carrying mutated *tna* promoter excised form plasmid PL51 and *Nde*I-*Nco*I fragments carrying 5' region of appropriate ORF. Plasmids PL58-59 were transformed into *H.volcanii* H98 and growth of the transformant cells was tested on plates lacking and containing tryptophan (Figure 3.16).

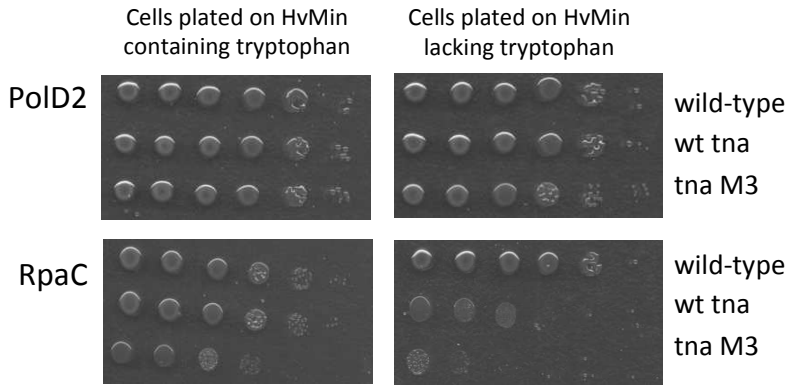


Figure 3.16 Growth of the cells expressing PolD2 and RpaC from mutated *tna* promoters in the presence and absence of tryptophan

Strains were grown to mid-exponential phase in HvMin medium supplemented with 0.075 mM tryptophan. The cells were then washed and 10-fold serial dilutions spotted onto HvMin plates with (right panels) or without (left panels) tryptophan and incubated at 45^{o}C for 5 days.

As the promoter TATA box element is important for transcription efficiency, altering nucleotides corresponding to that sequence in the *tna* promoter is likely to decrease the overall promoter strength. This explains the weak growth of cells expressing *polD1*, *polD2* and *mcm* from *tna*M3 promoter in the absence of tryptophan, when cells expressing those genes from the wild-type *tna* promoter grew well in the same medium. At the same time, all strains grew equally well in the presence of tryptophan indicating that mutated version of p*tna* is still activated by tryptophan, as the wild-type promoter. However, for genes which expression was down-regulated below the wild-type level from wt *tna* promoter (*polB* and *rpaC*), the further decrease seen using *tnaM3* causes reduction in growth rate even on medium supplemented with tryptophan.

Quantification of the wild-type *tna* promoter and *tnaM3* promoter strength was perform using qRT-PCR. Comparison of relative mRNA levels of *polB*, *polD1* and *polD2* expressed from either wt *tna* or *tnaM3* after six hours of growing cells with and without tryptophan proved reduction in overall transcript level, independently of the presence of tryptophan, from *tnaM3* (Figure 3.17).

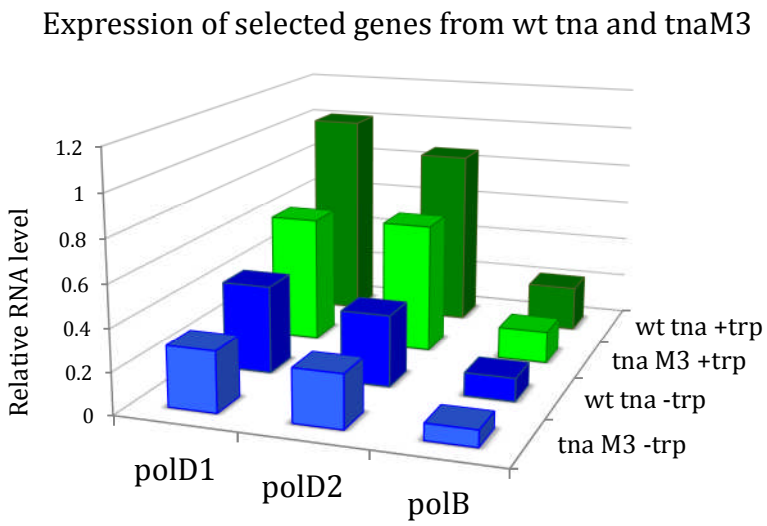


Figure 3.17 Comparison analysis of transcript level of selected genes expressed from the wt *tna*- and *tnaM3* promoter
Cells were grown in Hv-Min ± 0.075 mM tryptophan. The graphs show the relative genes expression after 6 hours of growth in medium supplemented and lacking tryptophan. All relative expression levels were normalized to overnight culture on Hv-Min + tryptophan.

3.6 Discussion

Conditional lethal mutants are powerful tools to study the function and structure of proteins encoded by essential genes. These mutants would be particularly advantageous for studying genes involved in DNA replication. Therefore, the initial

aim of this project was establish a system for inactivation of replication genes using the tryptophan-regulated promoter, p*tna*. The *tna* promoter, derived from the region upstream of the *H.volcanii* tryptophanase gene, is strongly induced in the presence of tryptophan and shows tight repression in medium lacking tryptophan (Large et al., 2007). P*tna* was previously used for down-regulation of three reporter genes, *pyrE2, bgaH* and *cct1*, in *H.volcanii* (Large et al., 2007). It is also used for conditional overexpression of halophilic proteins in *H. volcanii* in medium supplemented with tryptophan (Allers et al., 2010). We decided that *tna* promoter is a promising candidate for directing conditional inactivation of replication genes in *H.volcanii*.

First, we constructed fourteen plasmid-integrant strains carrying a 5'-truncated gene with its native promoter intact and a full-length gene under p*tna* (Figures 3.2 and 3.3). A complete list of genes used in this study is given in Table 3.1, but briefly, it includes the major components of the replication machinery in the Archaea. Among the genes, eleven were shown to be essential to cell viability in *H.volcanii* or the closely related halophile, *Halobacterium* sp. NRC-1 (Berquist et al., 2007, Blaby et al., 2010, Zhao et al., 2006). Growing the *tna* plasmid-integrant strains in the presence and absence of tryptophan showed that three strains, expressing *polB, rpaC* and *priS* from p*tna*, grew poorly in medium lacking tryptophan (Figure 3.4). Analysis of expression levels by quantitative reverse transcription real-time PCR confirmed that the observed phenotype was due to reduction of the mRNA levels of genes to below the levels seen in the wild-type cells (Figure 3.6). For *polB, rpaC* and *priS* genes, genetically stable *tna*-replacement strains were generated which, as expected, also showed growth arrest in the medium lacking tryptophan (Figures 3.9 and 3.10). For the essential genes that did not show any growth defect on medium without tryptophan, qRT-PCR indicated that the mRNA level was above the level seen in the wild-type. This might be caused by the trace amounts of tryptophan remaining in medium that were sufficient to activate p*tna*, by incomplete shut-off of the promoter in the absence of tryptophan or by read-through transcription initiated upstream of p*tna*. In order to improve and extend the usefulness of the *tna* promoter system to more replication genes, attempts were made to construct altered *tna* promoter versions that would still be activated by tryptophan but which would display reduced minimal/maximal expression levels. In total, eight p*tna* derivatives were generated: *tna2-5* with truncations at 5' region; *tnaL11*, with the expression cassette flanked by transcriptional terminator L11e to ensure that gene of interest is insulated from read-through transcription and *tna*M1-M3, carrying single point

mutations in TATA box element. These promoters were used to drive expression of three replication genes, *polB*, *polD1* and *mcm* and the growth properties of corresponding strains were re-tested on plates lacking and supplemented with tryptophan. Within the group of altered *tna* promoters, *tna*M3 showed a strong decrease in transcriptional efficiency, observed in qRT-PCR data and in spotting assays as a growth defect of strains expressing *polD1*, *polD2* and *mcm* genes from this promoter.

The TATA box is an element found in all eukaryotic and archaeal promoters that is important for binding of TATA box binding protein and subsequent DNA distortion and transcription initiation (Jun et al., 2011, Peng et al., 2011). We identified the putative *tna* promoter TATA box element as the sequence 5'-ATTACTTC-3', located at position -28 to -21 relative to transcriptional start site. Analysis of point mutations in this region of *H.volcanii* $tRNA^{Lys}$ promoter (Palmer and Daniels, 1995) and *S.shibatae* 16S rRNA promoter (Reiter et al., 1990) showed that efficient transcription initiation requires a pyrimidine residue at the 5' end of the TATA box (position -28 of the p*tna* TATA box) and an adjacent 3- or 4-nucleotide region rich in AT base pairs. Cytosine and guanine at this position reduces promoter strength. In agreement with that observation, mutating T to G at position -26 in the *tna* promoter decreases transcriptional activity to about 30%. It should be noted that the two other changes introduced to *tna* promoter, A to G at position -28 and T to C at position -27, may also influence transcriptional efficiency but this was not further investigated. Mutational analysis of *tna* promoter was done in order to identify altered versions of the promoter capable of directing down-regulation of replication genes for which wild-type p*tna* failed: *tna*M3 showed those properties.

The results presented in this chapter showed that the *tna* promoter and its mutated version *tna*M3 could be successfully used for down-regulation of replication genes in the halophilic archaeon *H.volcanii*. The *tna*M3 displays a reduced overall expression level in comparison to its wild-type version and can be used for genes with low expression level, which do not show the desired phenotype with wt *tna*. Selected plasmid-integrant strains and genetically stable promoter replacement strains, whose construction was described in this chapter, were used to study the structure and *in vivo* function of proteins believed to be involved in DNA replication in *H.volcanii*.

Chapter 4
Identification of putative single-stranded DNA-binding proteins in *Haloferax volcanii*

4.1 Introduction

As described in the Introduction chapter, single-stranded DNA-binding proteins are abundant, OB fold-containing proteins, playing a key role in DNA replication and other cellular processes involving ssDNA. Unlike in Bacteria and Eukaryotes, SSBs present in Archaea display a variety of domain and subunit organisations; in Euryarchaea multiple proteins with various number of OB folds were identified. The presence of multiple orthologues raises questions of interplay between them, why different SSBs are needed, do they have different physiological functions or do they overlap, what particular pathways they participate in?

A detailed understanding of the function of the archaeal RPA proteins requires a combination of *in vivo* genetic and *in vitro* biochemical and biophysical approaches. To date, most of the work on archaeal SSBs was done for crenarchaeal proteins; very little is known about RPA proteins from euryarchaea. Bioinformatic and biochemical studies have been performed for three RPA homologous from the methanogenic archaeon *Methanosarcina acetivorans* MacRPA1-3 (Lin et al., 2008, Robbins et al., 2005, Robbins et al., 2004a) but no genetic study was done for any archaeal SSB. In order to better understand the function of RPA proteins, we decided to identify RPA proteins in *H.volcanii* and perform genetic analysis of their function taking advantage of the genetic tools available for *Haloferax.*

4.2 *H.volcanii* encodes three putative RPA homologous

In order to identify genes encoding putative single-stranded DNA-binding proteins in *H.volcanii,* the sequences of the three *M.acetivorans* RPA homologous, MacRPA1 (MA_4645, GI:19918797), MacRPA2 (MA_3019, GI:20091837) and MacRPA3 (MA_0590, GI:19914385) were used as a query in a BLAST search (Altschul et al., 1990). *Methanosarcinales* are a sister clade of the *Halobacteriales* and *M.acetivorans* RPA proteins were previously characterized biochemically (Lin et al., 2008, Robbins et al., 2005, Robbins et al., 2004a). MacRPA1 encodes a protein

containing four OB folds whereas MacRPA2 and MacRPA3 encode proteins with two OB folds and a $CX_2CX_8CX_2H$ zinc finger motif (Figure 4.1A). BLAST search identified orthologues of these proteins in all haloarchaeal species with sequenced genomes available in then current databases (accession numbers of haloarchaeal RPA proteins and identity/similarity values with *M.acetivorans* RPAs are given in Appendix, Table A2 and A3).

In *H.volcanii*, the MacRPA1, MacRPA2 and MacRPA3 orthologues were identified as HVO_0519, HVO_1338 and HVO_0292 genes, respectively (summarised in Figure 4.1A). The products of HVO_1338 and HVO_0292 have a similar domain organisation to their methanogen counterparts: HVO_1338 encodes protein consisting of two putative OB folds and the product of HVO_0292 possesses one OB fold. In addition, both proteins contain a zinc finger motif with general consensus $CX_4CX_{8-12}CX_2H$ (the same motif in MacRPA2 orthologue has sequence $CX_2CX_8CX_2H$). In *H.volcanii*, ORFs encoding HVO_1338 and HVO_0292 ORFs are likely to be co-transcibed with the downstream ORFs (HVO_1337 and HVO_0291, respectively). For simplicity, we designed the protein encoded by the HVO_1338 gene as RpaA1 and by HVO_1337 as RpaA2. Similarly, the proteins encoded by HVO_0292 and HVO_0291 were designated as RpaB1 and RpaB2, respectively. The *rpaA1* and *rpaA2* ORFs overlap by three nucleotides, whereas *rpaB1* and *rpaB2* are separated by one nucleotide only. In both cases, the products of the downstream genes are related to each other and possess conserved COG3390 domain, which has some weak sequence similarities to OB fold (indicated by interactive database searching using PSI-BLAST (Altschul et al., 1990). Five nucleotides downstream of the *rpaA2* in *H.volcanii* is an ORF encoding a putative metallophosphoesterase, Rpe (HVO_1336), of unknown function, that might be also co-transcribed with *rpaA1* and *rpaA2*. Downstream of *rpaB2* is an ORF encoding the haloarchaeal-specific protein named Ral. In this case, the gap between *rpaB2* and *ral* is 165 nucleotides, suggesting that Ral is not part of the same transcription unit. Similar gene organisation is seen in all other haloarchaeal organism with sequenced genomes (see Appendix, Table A2) and in some Methanogens: in *M.acetivorans* for example MacRPA2 and MacRPA3 are located adjacent to the genes encoding homologous of RpaA2 and RpaB2, respectively, but the biochemical properties of these proteins have not been tested. An Rpe-related ORF is also found in *M.acetivorans*, but its location is not linked with MacRPA2.

The third ORF encoding putative RPA protein in *H.volcanii*, HVO_0519 (orthologue of MacRPA1) was designated as RpaC. The corresponding protein consists of three OB fold domains and conserved N- and C-terminal domains of unknown function. *rpaC* is separated from its downstream and upstream ORFs by almost 200 nucleotides, strongly suggesting that it is transcribed as a monocistronic mRNA.

Figure 4.1 RPA proteins in *H.volcanii* and *M.acetivorans*.

A Tables summarizing RPA homologous in two euryarchaeal species *M.acetivorans* (left) and *H.volcanii*

B Gene organization of RPA proteins in *M. acetivorans* (left) and *H. volcanii*. The arrows represent individual open reading frames with systematic gene designations shown above (for clarity the MA and HVO prefixes are omitted). The shading groups ORFs encoding OB fold-containing proteins (white), putative OB fold-containing COG3390 proteins (light grey), the phosphoesterase Rpe (dark grey) and the haloarchaeal-specific Ral protein (black).

C Schematic of SSB protein structures showing OB folds (blue hatched boxes), COG3390 domains (orange speckled boxes) and the zinc finger motifs (red triangle). Protein lengths are indicated.

A

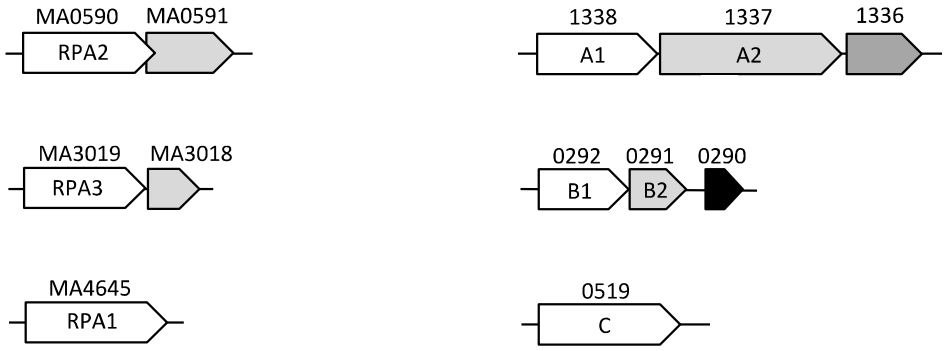
Methanosarcina acetivorans

	MacRPA2	MacRPA3	MacRPA1
Gene ID	20091837	19914385	19918797
Locus tag	MA_3019	MA_0590	MA_4645
Protein size (aa)	450	470	484
Domain organisation	2 OB folds + zinc finger	2 OB folds + zinc finger	4 OB folds

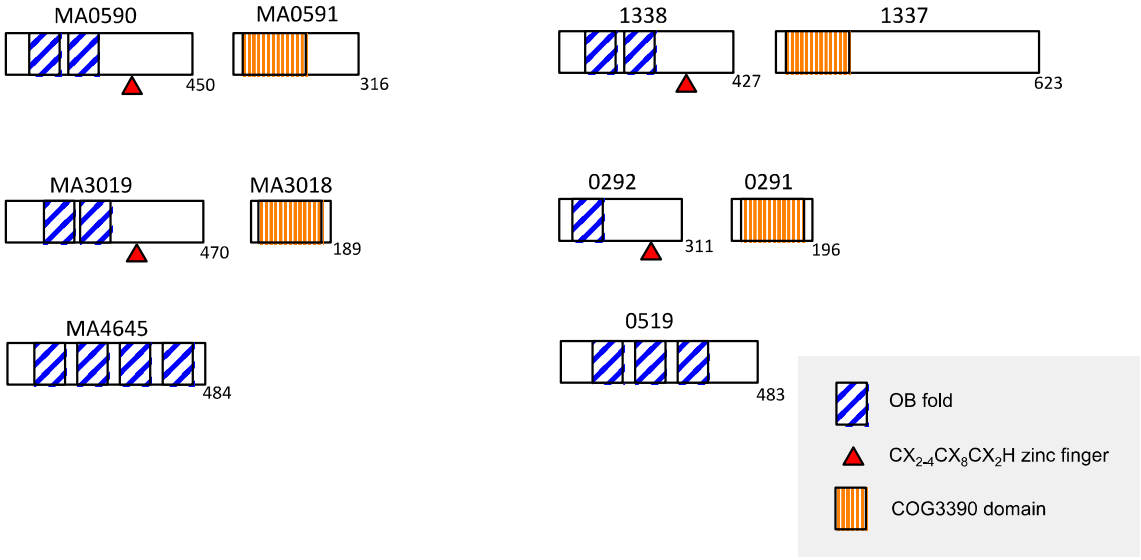
Haloferax volcanii

	RpaA1	RpaB1	RpaC
Gene ID	8924887	8924234	8926823
Locus tag	HVO_1338	HVO_0292	HVO_0519
Protein size (aa)	427	311	483
Domain organisation	2 OB folds + zinc finger	1 OB fold + zinc finger	3 OB folds

B



C



4.3 Deletion analysis of RPA homologous in *H.volcanii*

4.3.1 Construction of plasmids for gene deletion

To construct plasmids for gene deletion, regions of about 500 bp from 5' and 3' to the region to be deleted were amplified by PCR, using oligonucleotide P78-P89 (Appendix, Table A1) and *H.volcanii* DS70 genomic DNA as a template. The PCR products were then digested with *Eco*RI and *BamH*I (5' flanking regions of the genes) or *BamH*I and *Spe*I (3' flanking regions of the genes) and ligated into plasmid pTA131 that had been digested with *Eco*RI and *Spe*I. All sites that was introduced were unique and not present anywhere else in the PCR product. Plasmids pTA131 carries *pyrE2* selectable marker, allowing selection for *pyrE2*$^+$ integrant cells (pop-in cells) and counter- selection for intramolecular recombinants that have lost the plasmid (pop-out cells) on medium supplemented with uracil and 5-fluoroorotic acid (5- FOA) as 5-FOA is converted into toxic 5- fluorouracil in *pyrE2*$^+$ cells.

The resulting plasmids (PL60, PL62 and PL64, Table 2.3) were sequenced to ensure the absence of unwanted sequence changes. Next, *trpA* and *hdrB* selectable markers were cloned into the *BamH*I site located in the centre of the 5' and 3' flanking regions. The *trpA* marker was excised as a *BamH*I fragment from plasmid pTA298 and *hdrB* was excisted as a *BamH*I-*Bgl*II fragment from plasmid pBBHdrB. Plasmid pBBHdrB is a derivative of pTA187 in which a *Bgl*I site has been introduced downstream of the *hdrB* gene to ease subsequent sub-cloning. The resulting plasmids (PL61, PL63 and PL65, Table 2.3) were passage through *E.coli* SCS110 to be demethylated. The strategy used to construct plasmids for gene deletion is summarised in Figure 4.2.

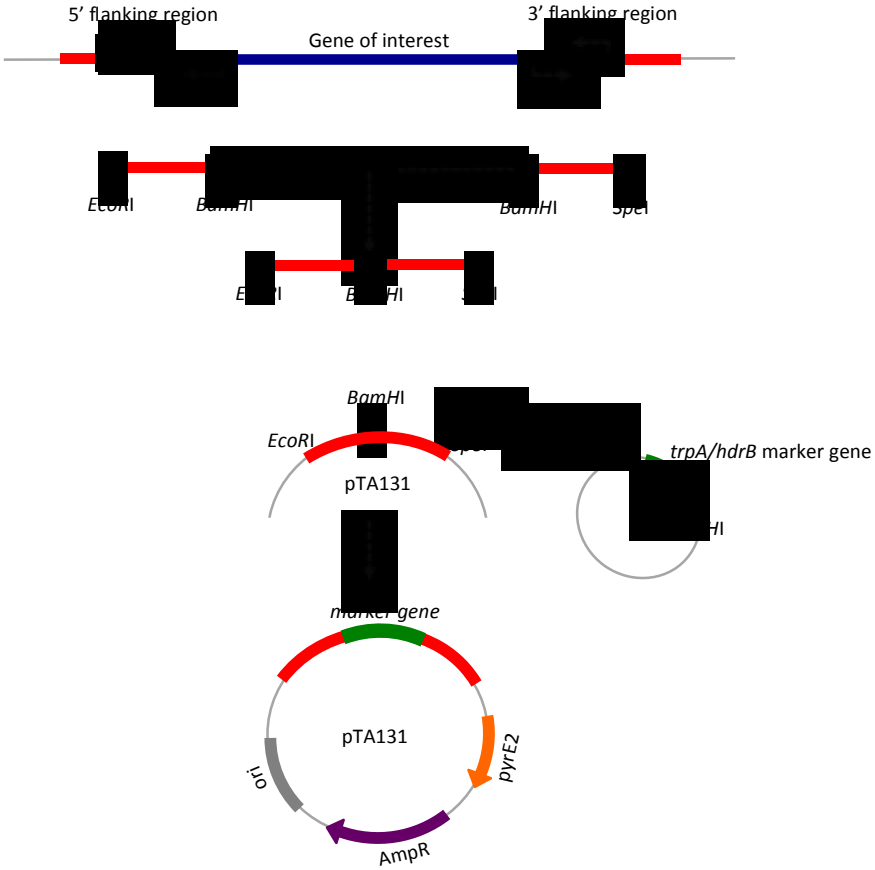


Figure 4.2 Construction of plasmids for *rpa* gene deletions

Plasmids for gene deletion were constructed by PCR amplification of 5' and 3' flanking regions of the genes to be deleted and ligating them into plasmid pTA131. Next, restriction fragments carrying selected marker gene were cloned into the plasmid in the centre of flanking regions.

4.3.2 Construction of a single *rpa* deletions

Plasmids PL61, PL63 and PL65 described in the previous section were transformed into *H.volcanii* H99 (Δ*pyrE2* Δ*hdrB* Δ*trpA*) and transformants obtained on Hv-Ca medium with appropriate supplements. For each gene deletion, three integrant colonies were picked, serial diluted in 18% SW from 10^0 to 10^{-3} and plated on Hv-Ca plates, supplemented with both uracil and 5-FOA, with and without tryptophan (for selection of deletions marked with *trpA*) or with or without thymidine and hypoxanthine (for selection of deletions marked with *hdrB*). Adding 5-FOA to the medium allowed counter selection of recombinants that have lost the plasmid while this compound is converted to toxic 5-fluorouracil in *pyrE2*$^+$ cells but not in *pyrE2*$^-$ cells. If gene is essential, the number of colonies should be significantly higher on non- selective plates, when loss of the plasmid had restored the chromosome to the wild- type, losing the marker gene. Candidate gene deletion colonies (typically

sixteen colonies of each type) were screened by PCR. Deletion strains were then purified by streaking to single colonies on Hv-YPC medium and re-tested by PCR.

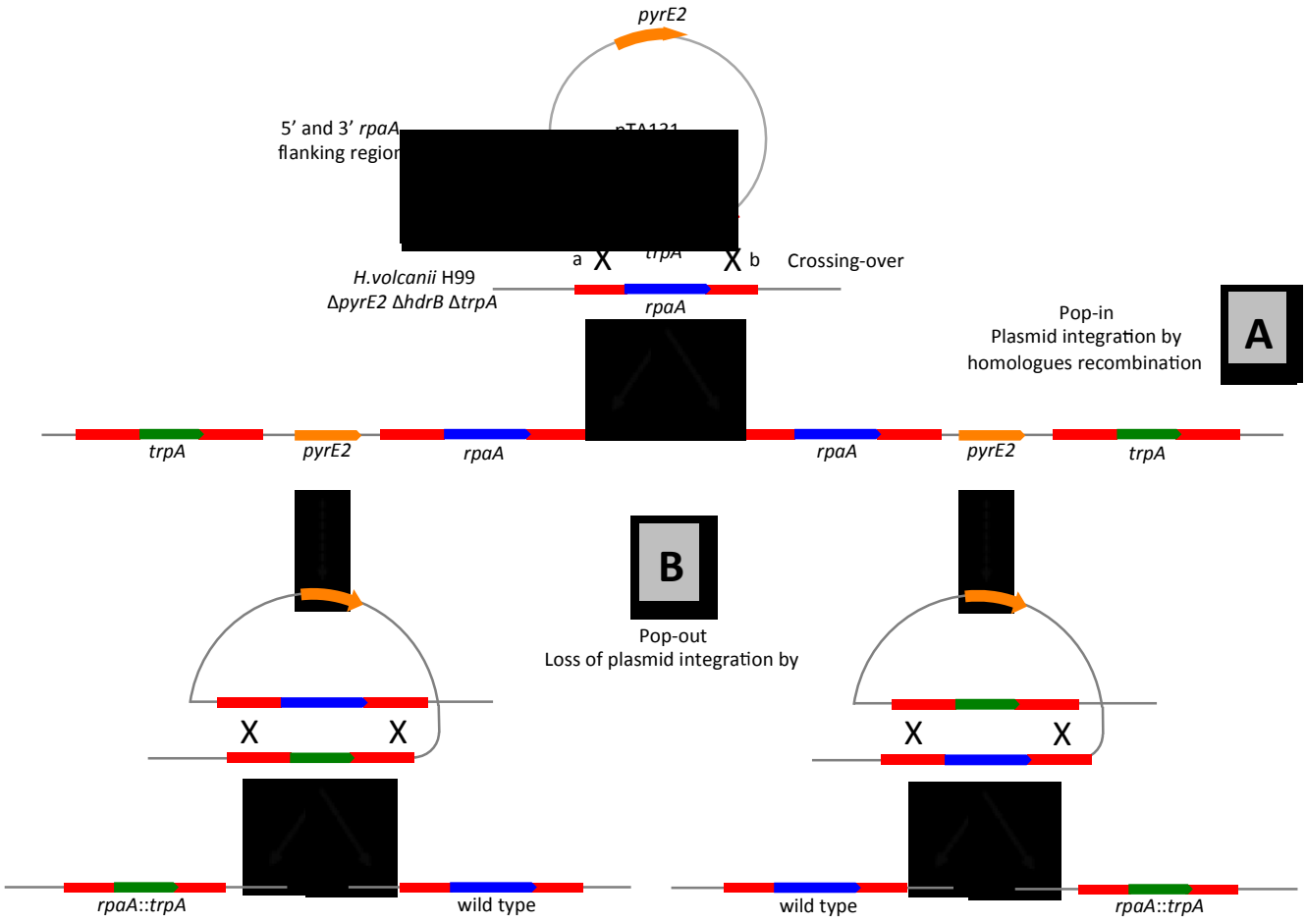


Figure 4.3 Schematic diagram of pop-in/pop-out gene deletion strategy exemplified for *rpaA*.
H.volcanii H99 was transformed with pTA131-derived plasmid carrying *tpaA* marker flanked by sequences 5' and 3' to the *rpaA* (**A**). The *pyrE2*$^{+}$ pop-in cells that have integrated the plasmid were selected on medium lacking uracil and tryptophan (the plasmid might be integrated into the chromosome in two possible orientations represented by a and b). Next, transformant cells were plated on medium supplemented with both uracil and 5-FOA, with or without tryptophan (non-selective and selective medium, respectively (**B**). Candidate gene deletion colonies were analysed by PCR screening for the presence of *rpaA* and *trpA* genes.

The results of two independent experiments were as follows. For deletion of *rpaA1-rpaA2-rpe* operon, we found comparable number of colonies on selective and non-selective plates (Table 4.1, lines 1) and all sixteen colonies picked from selective medium and analysed by PCR were confirmed to be deleted for the *rpaA1-rpaA2-rpe* genes, indicating that the products of *rpaA1*, *rpaA2* and *rpe* are non-essential for cell viability in *H.volcanii*.

For deletion of *rpaB1-rpaB2* operon, although the number of colonies on non-selective medium was significanlly higher in comparison to selective medium

following pop-out (Table 4.1, lines 3), all sixteen colonies from the selective plates screened by PCR showed loss of the *rpaB1-rpaB2* from their native locus, implying that the products of *rpaB1* and *rpaB2* genes are also non-essential in *H.volcanii*. A similar disparity between observed and expected colony number was seen in the lab when creating *hdrB*-marked (but not *tpaA*-marked) deletions of other non-essential genes in *H.volcanii* on HvCa medium (unpublished results). The *rpaB1-rpaB2* deletion experiment was repeated using HvMin medium instead of HvCa and we found that the number of colonies on selective and non-selective medium was comparable (Table 4.1, lines 4) and all the pop-outs screened from selective plates were Δ*rpaB1*-Δ*rpaB2* (data not shown). That indicates that for the future experiments, HvMin might be a more appropriate medium for gene deletion using the *hdrB* marker that HvCa.

For deletion of *rpaC* with *trpA* marker, we found from 10^4 to 10^5 more colonies on non-selective plates than on selective plates following pop-out (Table 4.1, lines 2). All colonies from selective plates analysed by PCR (thirty two colonies in total) still contained the gene, strongly suggesting that *rpaC* is an essential gene in *H.volcanii*. The colonies obtained on the selective medium are presumably either heterozygotic (they possess both *rpaC* and *rpaC::trpA* at *rpaC* locus on different chromosomes after pop-out as *H.volcanii* is a polyploid species or carry mutations in *pyrE2* gene that give resistance to 5-FOA that allow cells that did not lose the plasmid to grow on plates supplemented with this compound.

Table 4-1 Number of colonies obtained in gene knockout experiments

Cells were grown as described in Materials and Methods chapter and plated onto non-selective and selective medium. Colony-forming units per ml of cell culture are shown, determined by plating serial dilutions.

	Strain background	Attempted deletion	Cfu/ml		Ratio
			Non-selective medium FOA+ura+t\h+trp	Selective medium FOA+ura+t/h	
1	H99	*rpaA::trpA*	3.7×10^6	2.4×10^6	1.5:1
2	H99	*rpaC::trpA*	1.8×10^5	20	9000:1
			Non-selective medium FOA+ura+t\h+trp	**Selective medium FOA+ura+trp**	
3	H99	*rpaB::hdrB*	3×10^4	1.2×10^2	250:1
			HvMin non-selective medium FOA+ura+t\h+trp	**HvMin selective medium FOA+ura+t/h**	
4	H99	*rpaB::hdrB*	2.5×10^4	1.7×10^4	1.5:1

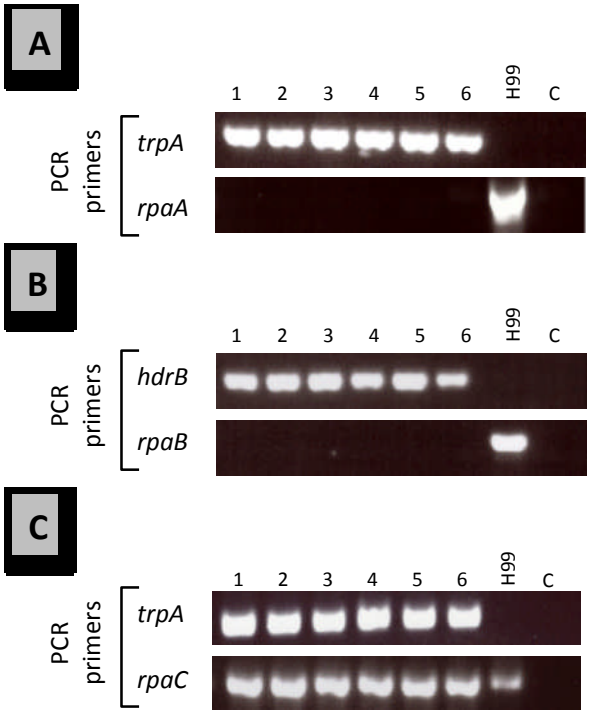


Figure 4.4 Screening of *rpa* gene deletions

A and **B** PCR analysis of chromosomal DNA prepared from six *rpaA*::*trpA* (**A**) and *rpaB*::*hdrB* (**B**) colonies isolated on selective medium demonstrating successful deletion of the *rpaA1-rpaA2-rpe* locus and *rpaB1-rpaB2* locus, respectively.
C PCR analysis of six representatives of the rare colonies isolated on selective medium following attempted replacement of *rpaC* with *trpA*. The primers used for PCR were located within the ORFs targeted for deletion and marker genes.

4.3.3 Phenotypic analysis of *rpaA* and *rpaB* deletion strains

Deleting the entire *rpaA1-rpaA2-rpe* and *rpaB1-rpaB2* loci did not cause any obvious phenotype or growth defects indicating that the products of these genes are non-essential for cell viability under normal growth conditions. An increase in transcript level of *rpaB1-rpaB2-ral* genes has been reported in *Halobacterium* sp.

NRC-1 after UV irradiation (McCready et al., 2005) as well as in two extremely radiation-resistant mutants of this archaeon (DeVeaux et al., 2007), suggesting the role of RpaB proteins in the repair of DNA damage caused by UV irradiation. Therefore, we have tested sensitivity of Δ*rpaA1-rpaA2-rpe* and Δ*rpaB1-rpaB2* to UV treatment. We found comparable response to UV irradiation between *rpaA::trpA* strain the wild-type whereas loss of the *rpaB* operon increases sensitivity to UV (Figure 4.5).

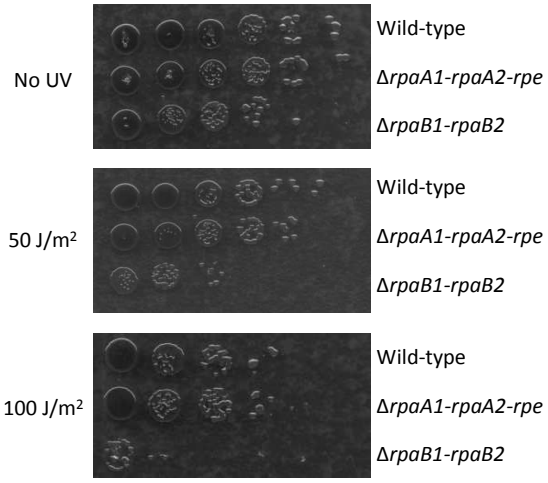


Figure 4.5 Response to UV treatment of *rpaA* and *rpaB* deletion strains

The *rpaA::trpA*) and *rpaB::hdrB* strains were grown up to mid-exponential phase at 45°C in YPC medium and ten fold serial dilutions spotted onto YPC plates following irradiation with the indicated doses of UVC. Plates were incubated in dark at 45°C for three days.

4.3.4 Construction of a double Δ*rpaA* Δ*rpaB* deletion

As shown in previous section, the individual deletions of the *rpaA1-rpaA2-rpe* and *rpaB1-rpaB2* operons are viable in *H.volcanii* and deletion strains do not have any growth defect. The relatedness of the products of *rpaA1* and *rpaB1* and also *rpaA2* and *rpaB2* raises the possibility that these genes might share an essential function. Therefore, to test this, we attempted to construct an Δ*rpaA* Δ*rpaB* double deletion strain. This was done is three ways.

First, we tried to delete *rpaA* operon in *rpaB*::*hdrB* background and *rpaB* operon in *rpaA*::*trpA* background using the pop-in/pop-out approach as described for construction of single *rpa* deletions. Plasmid PL61 was transformed into the *rpaB*::*hdrB* strain and integrant colonies were plated onto HvCa medium supplemented with uracil and 5-FOA, with or without tryptophan to identify the cells in which plasmid has been lost leaving the *trpA* marker gene replacing *rpaA1-rpaA2-rpe* genes at their native loci. Colonies from selective plates (lacking tryptophan) were screened by PCR to detect the presence or absence of the *rpaA* operon. All colonies tested still contained *rpaA1-rpaA2-rpe* operon suggesting that an Δ*rpaA* Δ*rpaB* double deletion is not viable in *H.volcanii* (Figure 4.6, A). Consistent with

this result, our attempt to replace the *rpaB* operon with *hdrB* marker in *rpaA*::*trpA* background was also unsuccessful. While deletion of *rpaB1-rpaB2* operon in the wild-type using plasmid PL63 was straightforward, we were unable to delete the genes in a *rpaA*::*trpA* strain; all pop-out colonies picked from selective medium and analysed by PCR showed the presence of *rpaB* genes (Figure 4.6, B).

Finally, we attempted to construct an Δ*rpaA* Δ*rpaB* double deletion strain by mating *rpaA*::*trpA* and *rpaB*::*hdrB* strains as this method has been used in the lab (Zhao et al., 2006). Briefly, exponentially growing *rpaA*::*trpA* and *rpaB*::*hdrB* cultures were combined, harvested by filtration and incubated overnight on rich medium before being plated on selective medium (HvCa lacking both tryptophan and thymidine/hypoxanthine). Control matings were performed with *H.volcanii* H53 (Δ*trpA hdrB*$^{+}$) and H98 (*trpA*$^{+}$ Δ*hdrB*) strains. We were unable to obtain any cells carrying Δ*rpaA* Δ*rpaB* double deletion, despite the fact that all control matings resulted in a both *rpaA*::*trpA* and *rpaB*::*hdrB* single deletion.

Taken together, our unsuccessful attempts to create Δ*rpaA* Δ*rpaB* double deletion strain in *H.volcanii* indicates that the products of *rpaA1-rpaA2-rpe* and *rpaB1-rpaB2* operons share at least one essential function.

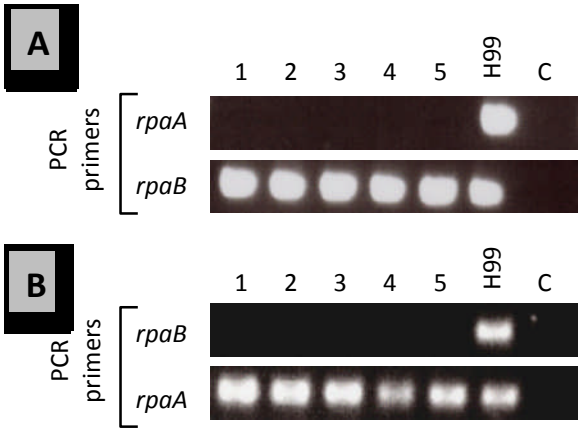


Figure 4.6 Screening of Δ*rpaA* Δ*rpaB* double deletion

PCR analysis of chromosomal DNA prepared from five colonies from selective medium following attempted construction of *rpaB*::*hdrB* strain in Δ*rpaA1-rpaA2-rpe* background (**A**) and *rpaA*::*trpA* strain in Δ*rpaB1-rpaB2* background (**B**). The primers used for PCR were located within the ORFs targeted for deletion and marker genes.

To extend the observation that simultaneous deletion of *rpaA1-rpaA2-rpe* and *rpaB1-rpaB2* loci is not viable in *H.volcanii*, we examined the effect of placing the *rpaB1-rpaB2* ORFs under control of the *tna* promoter in a Δ*rpaA* strain. As shown in Chapter 3, *rpaA*- and *rpaB*- pNPM-*tna*-plasmid integrant strains (SMH728 and SMH729, respectively, Table 2,1), carrying a 5'-truncated genes with their native promoter intact and a full-length genes under p*tna*, did not show any growth defect when cells were grown in medium lacking tryptophan (*tna* promoter shut off) (Figure

3.5). When we generated *rpaB-tna*-plasmid integrant strain in a Δ*rpaA* background and grew cells without tryptophan, to mimic an Δ*rpaA* Δ*rpaB* double deletion, growth was significantly retarded (Figure 4.7). The construction of a *rpaB*-p*tna* Δ*rpaA* strain offers a useful tool for future analysis of physiological functions of non-essential RPA proteins in *H.volcanii*.

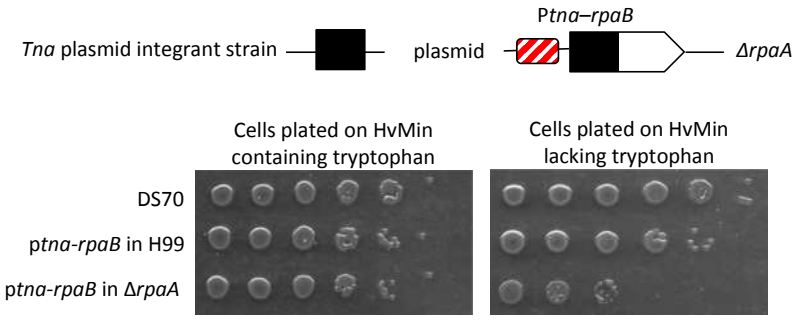


Figure 4.7 Growth of the Δ*rpaA* strain carrying integrated pNPM-tna-RpaB
Cells were grown up to mid-exponential phase at 45°C in HvMin medium containing 0.075 mM tryptophan. The cells were then washed and 10-fold diluted cells spotted onto HvMin plates with and without tryptophan. Plates were incubated for 3 days at 45°C.

4.4 Discussion

Single-stranded DNA-binding proteins (or RPA as they are called in Eukaryotes and Euryarchaea) are abundant proteins found in all three domains of life that play roles in many aspects of DNA metabolism including replication, repair and recombination (Flynn and Zou, 2010, Pestryakov and Lavrik, 2008, Broderick et al., 2010). They use the oligosaccharide-oligonucleotide binding fold (OB fold), a five-stranded β-sheet coiled to form a closed β-barrel, to stabilize exposed ssDNA regions and facilitate ssDNA-protein and protein-protein interactions (Bochkarev and Bochkareva, 2004). As the part of this PhD project the first genetic analysis of RPA in archaea was performed, using the genetically tractable halophilic species *H.volcanii*. As the *H.volcanii* genome encodes three putative RPA homologous, studying RPA proteins in this organism offers a better understanding how multiple SSBs perform their cellular functions.

RPA proteins among Euryarchaea display diverse architecture and subunit organisation. Proteins with multiple OB folds and a zinc finger motif from species belonging to several major lineages have been identified. The best-studied examples are trimeric complex found in *P.furiosus* and three distinct proteins MacRPA1, MacRPA2 and MacRPA3 from the methanogenic archaeon *M.acetivorans*. The

MacRPA2 and MacRPA3, both containing two OB folds and a zinc finger motif, are closely related to each other, indicating a likely descent from a common ancestor.

Using sequences of the MacRPAs as a query in a BLAST search homologous of these proteins in haloarchaea, a sister clade to the methanogens, were identified. In *H.volcanii*, RPA homologous were designated as RpaA1, RpaB2 and RpaC (homologous of MacRPA2, MacRPA3 and MacRPA1, respectively). The RpaA1 and RpaB1 have a similar domain organization to their methanogenic counterparts with one (RpaB1) or two (RpaA1) OB folds and a zinc finger motif. The third RPA homologue, RpaC, contains three clear OB folds and N- and C-terminal domains of unknown function that are also conserved among haloarchaea. Analysis of the genomic context of *rpa* genes in both haloarchaea and methanogens revealed the presence of conserved operons: RpaA1 and RpaB1 (as well as MacRPA2 and MacRPA3) are likely co-transcribed with adjacent downstream ORFs encoding proteins of the COG3390 family (Figure 4.1 and Table A2 and A3 for accession numbers of the proteins). These *rpaA1* and *rpaB1* downstream ORFs were designated as *rpaA2* and *rpaB2*, respectively. It was recently shown that RpaA1 and RpaB1 interact specifically with the proteins encoded by their respective downstream genes (Stroud et al., 2012).

The RpaA operon contains also the third gene, *rpe*, encoding a putative metallophosphoesterase. In Haloarchaea this gene is adjacent to *rpaA1-rpaA2* ORFs in all species with completely sequenced genome and often it overlaps with *rpaA2* (for example in *Halorubrum lacusprofundi*, *Halomicrobium mukohataei* and *Halorhabdus utahensis*). In methanogens, an *rpe*-related gene was also identified but it is not co-localized with genes encoding single-stranded DNA binding proteins.

An additional ORF, *ral*, is also found downstream of *rpaB1-rpaB2* locus. *ral* encodes a haloarchaeal-specific protein of unknown function. Although *ral* is close to the *rpaB* operon the distance from *rpaB2* ORF varies in different species; in *Halobacterium* sp. NRC-1 the gap is 45 nucleotides and *rpaB1*, *rpaB2* and *ral* appear to be co-transcribed (McCready et al., 2005), in *H.volcanii* the gap is 165 nucleotides and in several other species *rpaB2* and *ral* are physically separated by ORFs encoding (or predicted to encode) unrelated proteins indicating that the co-expression of *rpaB-rpaB2* and *ral* is not obligatory. It is also important to emphasize that Ral does not contain an OB fold and is perhaps unlikely to have any role in ssDNA binding.

To investigate which of the *H.volcanii* RPA proteins is essential for cell viability, attempts were made to delete the corresponding genes from their native chromosomal loci. The pop-in/pop-out method was used in which the regions of interest are replaced by an auxotrophic marker. For RPAs that are likely to be transcribed as a polycistronic mRNA we decided to delete the entire operons as a precursor to more detailed analysis. The individual *rpaA1-rpaA2-rpe* and *rpaB1-rpaB2* deletions were straightforward to generate and deletion strains showed no phenotype in terms of growth rate under normal growth conditions. However, *rpaB*::*hdrB* showed increased sensitivity to UV. Recently, sensitivity of an Δ*rpaB1* Δ*rpaB2* strain to mitomycin C (MMC) was also reported (Stroud et al., 2012). UV irradiation introduces bulky adducts such as cyclobutane-pyrimidine dimers (CPDs) and 6-4 photoproducts (6-4PPs) into DNA that distort the DNA helix and arrest replication forks which in turn might cause double-strand breaks (reviewed by (Yang, 2011). MMC is an alkylating agent that reacts covalently with DNA causing interstrands cross-links that can also potentiate double-strand breaks through replication fork blockage (Sedgwick et al., 2007). The increased sensitivity of the Δ*rpaB* strain to UV and MMC possibly implies a role of *rpaB* in the repair if DNA double-strand breaks and/or restoring stalled replication forks. In agreement with this, a rapid increase in *rpaB1-rpaB2-ral* expression was observed in *Halobacterium* sp. NRC-1 after exposure cells to UV. Under such conditions, RpaB proteins were strongly up-regulated together with the central player of homologous recombination (HR) RadA1, the archaeal homologue of Rad51/RecA (McCready et al., 2005). In humans, the interplay between Rad51 and RPA plays a role in the first step of HR, as RPA binding to exposed ssDNA precedes Rad51 presynaptic filament formation. In addition, RPA promotes recombination by removing secondary structures formed on ssDNA that could impair filament formation (reviewed by (Krejci et al., 2012)). An increased level of *rpaB1-rpaB2-ral* transcript was also seen in two *Halobacterium* sp. NRC-1 mutants that are extremely resistant to ionizing radiation (DeVeaux et al., 2007). The possible role of *H.volcanii* Ral protein in DNA repair is still not confirmed and will require deleting the corresponding gene in the wild-type and Δ*rpaB* strain and comparing the sensitivity pattern of these strains to different DNA-damaging agents.

While the individual *rpaA1-rpaA2-rpe* and *rpaB1-rpaB2* deletions were viable, we were unable to construct the double Δ*rpaA* Δ*rpaB* strain, either by the pop-in/pop-out method or by mating, suggesting that the products of these loci share at least one

essential function. The construction of a p*tna-rpaB* Δ*rpaA* conditional lethal mutant offers a useful tool for future analysis of this function.

In contrast to *rpaA1-rpaA2-rpe* and *rpaB1-rpaB2* genes, the third putative RPA homologue in *H.volcanii*, *rpaC,* is essential for cell viability. This observation encouraged us to perform detailed analysis of the physiological function performed by RPA protein.

Chapter 5

Structure-function analysis of RpaC

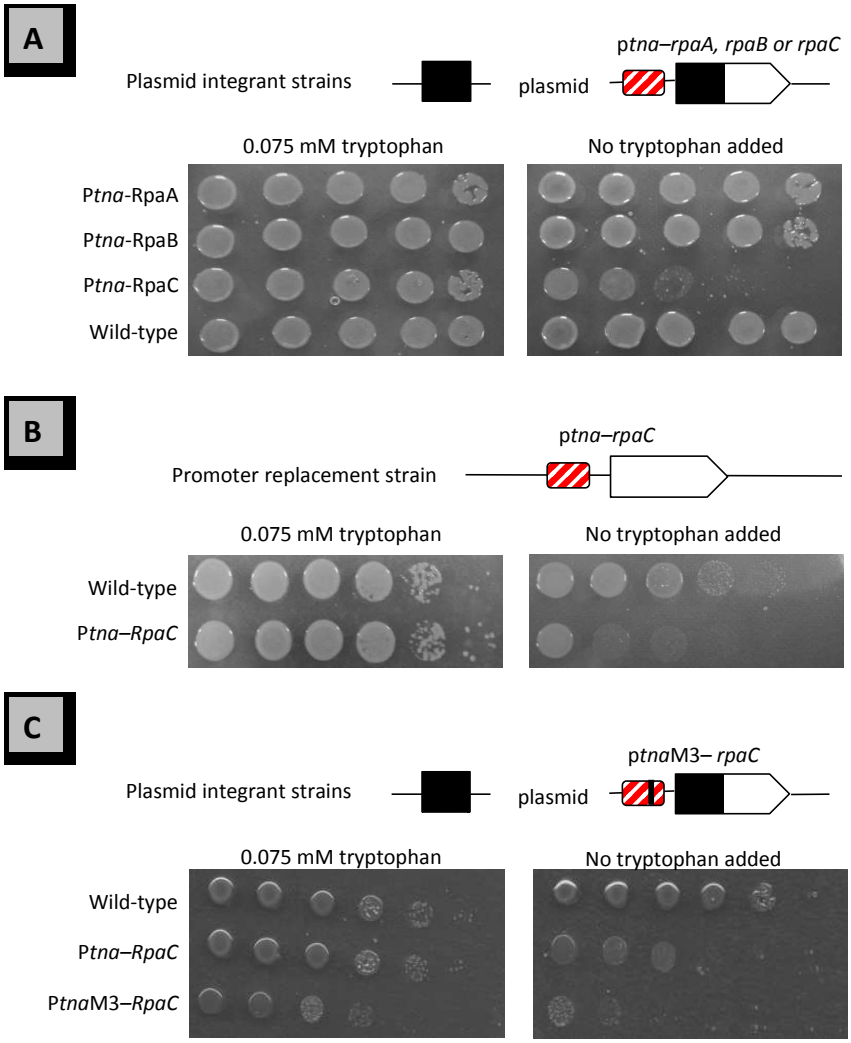
5.1 Conditional down-regulation of *rpaC* gene expression

In order to analyse the *in vivo* function of RpaC protein in *H.volcanii*, we used a tryptophan-inducible promoter system. As discussed in Chapter 3, the *tna* promoter can be successfully used for conditional down-regulation of expression of some replication gene, including *rpaC*. To summarize, we initially engineered strains in which the native promoters of *rpaA*, *rpaB* and *rpaC* were displaced by p*tna* (SMH728-SMH730, p*tna*- plasmid integrant strains). Little or no growth was seen when p*tna-rpaC* strain was grown on medium lacking tryptophan (promoter turned off), consistent with *rpaC* being an essential gene in *H.volcanii*. Transcriptomic analysis by qRT-PCR confirmed that the *rpaC* expression level in cells grown without tryptophan is indeed reduced below the level seen in the wild-type. At the same time, we observed that in cells grown on medium supplemented with 0.075 mM tryptophan the level of *rpaC* mRNA is elevated above the normal wild-type level. No detectable difference in growth rate between the integrated p*tna-rpaC* strain and the wild-type indicated that overexpression does not appear to have a detrimental effect on cells.

Next, we constructed a genetically stable p*tna-rpaC* strain (SMH738) using the pop-in/pop-out method; in this strain the native promoter of *rpaC* was replaced by p*tna*. We also constructed p*tna*M3-*rpaC* strain (SMH225), in which the wild-type *tna* promoter was replaced by its mutated version, *tna*M3, displaying reduced minimal/maximal expression levels (Figure 5.1).

Figure 5.1 Growth of the p*tna-rpaC* conditional lethal mutants

A *H.volcanii* H98 strain carrying integrated pNPM-tna-Rpa plasmids (PL19-PL21 Table 2.3). Strain DS70 is a wild-type control.
B The *H.volcanii* p*tna-rpaC* promoter replacement strain. Strain H98 is a wild-type control.
C *H.volcanii* H98 strain carrying integrated pNPM-tna-RpaC and pNPM-tnaM3-RpaC plasmids (PL59, Table 2.3). Strain DS70 is a wild-type control. The tna-RpaC plasmid integrant strain is shown to emphasize the further reduction of growth rate with mutated version of *tna* promoter. In **A**, **B** and **C** cells were grown up to mid-exponential phase at 45^{o}C in HvMin medium containing 0.075 mM tryptophan. Cells were then washed twice to remove tryptophan and 10-fold serial dilutions were spotted onto HvMin plates with or without tryptophan. Plates were incubated at 45^{o}C for 3 days.



Conditional lethal mutants are a powerful tool to study function and structure of proteins encoded by essential genes. The p*tna-rpaC* promoter replacement strain is particularly useful for complementation assays as this strain is genetically stable and uracil and thymidine/hypoxanthine auxotrophic (Δ*pyrE2* Δ*hbrB*), which allows further transformation with plasmids carrying the *pyrE2* and *hbrB* marker genes.

5.1.1 RpaC involvement in DNA replication

Although single-stranded DNA-binding proteins are known to play important roles in DNA replication, there is no direct evidence which of the multiple haloarchaeal RPAs is an replicative SSB. Based on fact that in *H.volcanii* RpaC is the only individually essential SSB protein, we concluded that RpaC is the best candidate for RPA being involved in DNA replication. To determine whether lack of RpaC expression has an effect on DNA synthesis a series of experiments was performed, in

which we monitored on-going DNA synthesis in p*tna-rpaC* and p*tna*M3-*rpaC* integrant cells grown in medium lacking and containing tryptophan. Both strains and the wild-type *H.volcanii* DS70 were grown for 18 hours in appropriate Hv-Min medium before being labelled with 20 µCi [methyl-^{3}H] thymidine (see Materials and Methods 2.2.6 and Figure 5.2A for details about experimental design). DNA synthesis was monitored at different time points by measuring the incorporation of radio-nucleotides into DNA.

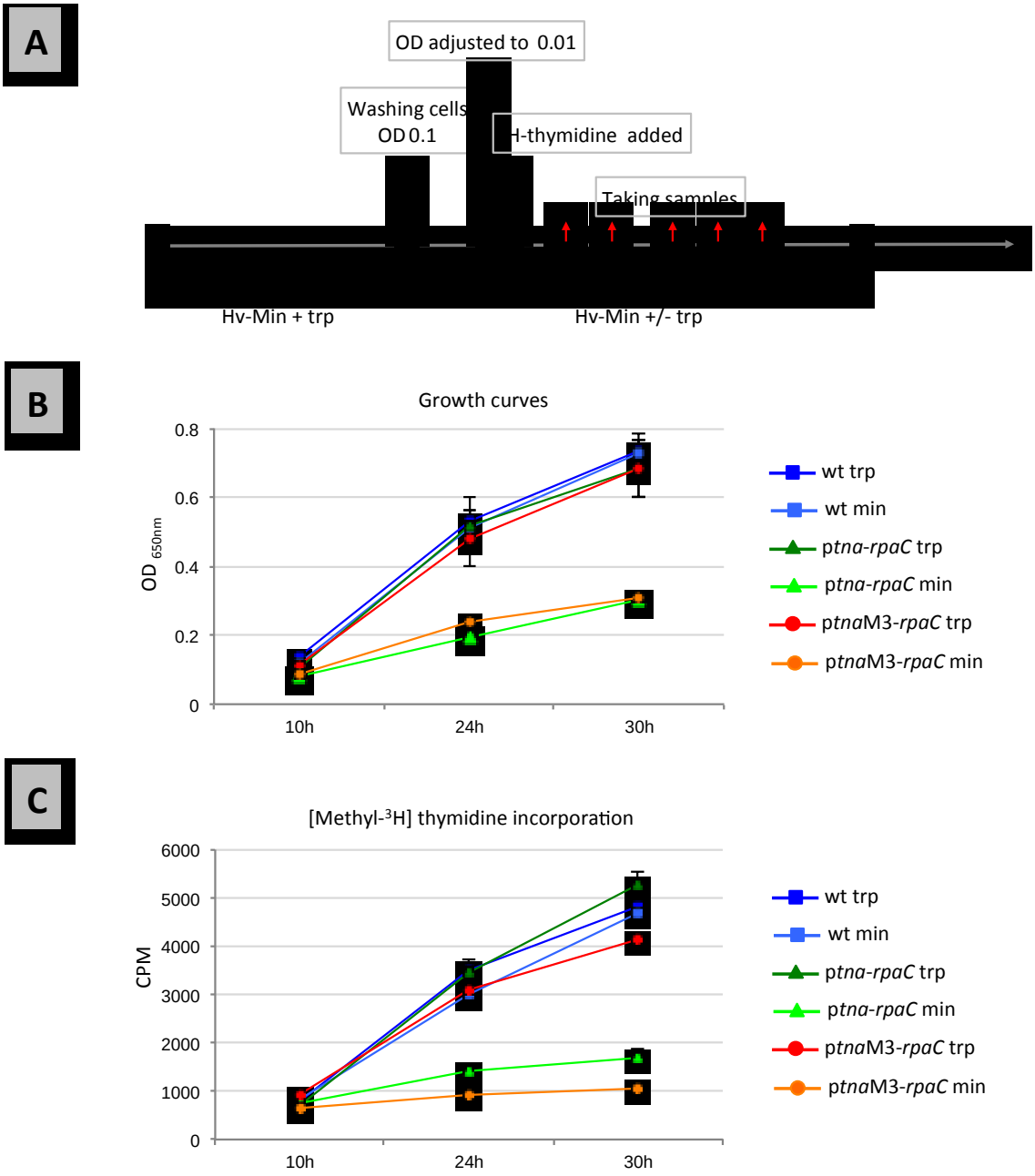
A strong reduction in [methyl-^{3}H] thymidine incorporation in cells expressing RpaC from *tna* and *tna*M3 promoter was observed starting 6 hours after adding the label in medium with no tryptophan (Figure 5.2B). The reduction in ^{3}H-thymidine incorporation is not observed in the wild-type grown under the same conditions nor in p*tna-rpaC* cells grown in the Hv-Min medium supplemented with 0.075 mM tryptophan. In addition, the p*tna-rpaC* strain showed weak growth on medium lacking tryptophan while growth on medium with tryptophan was indistinguishable from the wild-type (Figure 5.2C). Cells expressing RpaC from *tna*M3 promoter showed lower ^{3}H-thymidine incorporation in comparison with p*tna-rpaC* cells. Both ^{3}H-thymidine incorporation and growth were reduced compared to p*tna-rpaC* and the wild-type on Hv-Min medium with tryptophan as the *tna*M3 is a mutated version of p*tna* which displays reduced minimal expression level. Retarded growth of p*tna*M3-*rpaC* integrant strain was previously observed in spotting assay (see Chapter 3.3.4).

Figure 5.2 Monitoring of DNA synthesis in p*tna-rpaC* and p*tna*M3-*rpaC* strains

A Schematic view of experimental design. Cells were grown in Hv-Min medium containing 0.075 mM tryptophan until they reached mid-exponential phase (an OD650 nm 0.5). After pelleting and washing in Hv-Min medium cells were adjusted to an OD_{650nm} of 0.1 in a fresh Hv-Min medium either containing and lacking tryptophan and grown for 16 hours. Next, cells were diluted back to an OD_{650nm} 0.01 in corresponding medium and allowed to grow undisturbed for 2 hours when 20 µCi [methyl-3H] thymidine was added. At the time indicated, 200 µl of cultures was filtered through 0.7 µm filters and precipitated with 10% trichloroacetic acid and 1% (w/v) sodium pyrophosphate.

B Growth of p*tna-rpaC* and p*tna*M3-*rpaC* strains in Hv-Min medium. Mean and standard deviation of three independent experiments are shown. H.volcanii DS70 is a wild-type control.

C [Methyl-3H] thymidine incorporation. Experiment was performed as described above. DNA replication was monitored by using scintillation counting to measure the incorporation of labelled thymidine into DNA after TCA precipitation of whole cells. Mean and standard deviation of two independent experiments are shown.



5.1.1.1 PolB involvement in DNA replication

The experimental approach described above was used to test whether *H.volcanii* DNA polymerase B (PolB) is involved in DNA replication *in vivo*. Cells expressing PolB protein from the wild-type and mutated *tna* promoter (p*tna-polB* and p*tna*M3-*PolB* plasmid integrant strains, respectively) were arrested in Hv-Min medium for 18 hours following labelling with 20 µCi [methyl-^{3}H] thymidine. Both strains showed a strong growth inhibition on medium lacking tryptophan as a result of reducing *polB* transcript below the level seen in wild-type cells. Note that in the p*tna*M3-*PolB* plasmid integrant strain *polB* expression was further reduced in comparison to p*tna*-

polB and cells showed a growth defect also on medium supplemented with tryptophan. The diterpene antibiotic aphidicolin, a specific inhibitor of eukaryotic family B DNA polymerases (Krokan et al., 1981), was used in these experiments as a control of DNA replication arrest. It was previously shown that addition of aphidicolin results in cell division arrest and formation of elongated cells in *H.salinarum* (Forterre et al., 1984), however, the exact drug target was not examined in Euryarchaea.

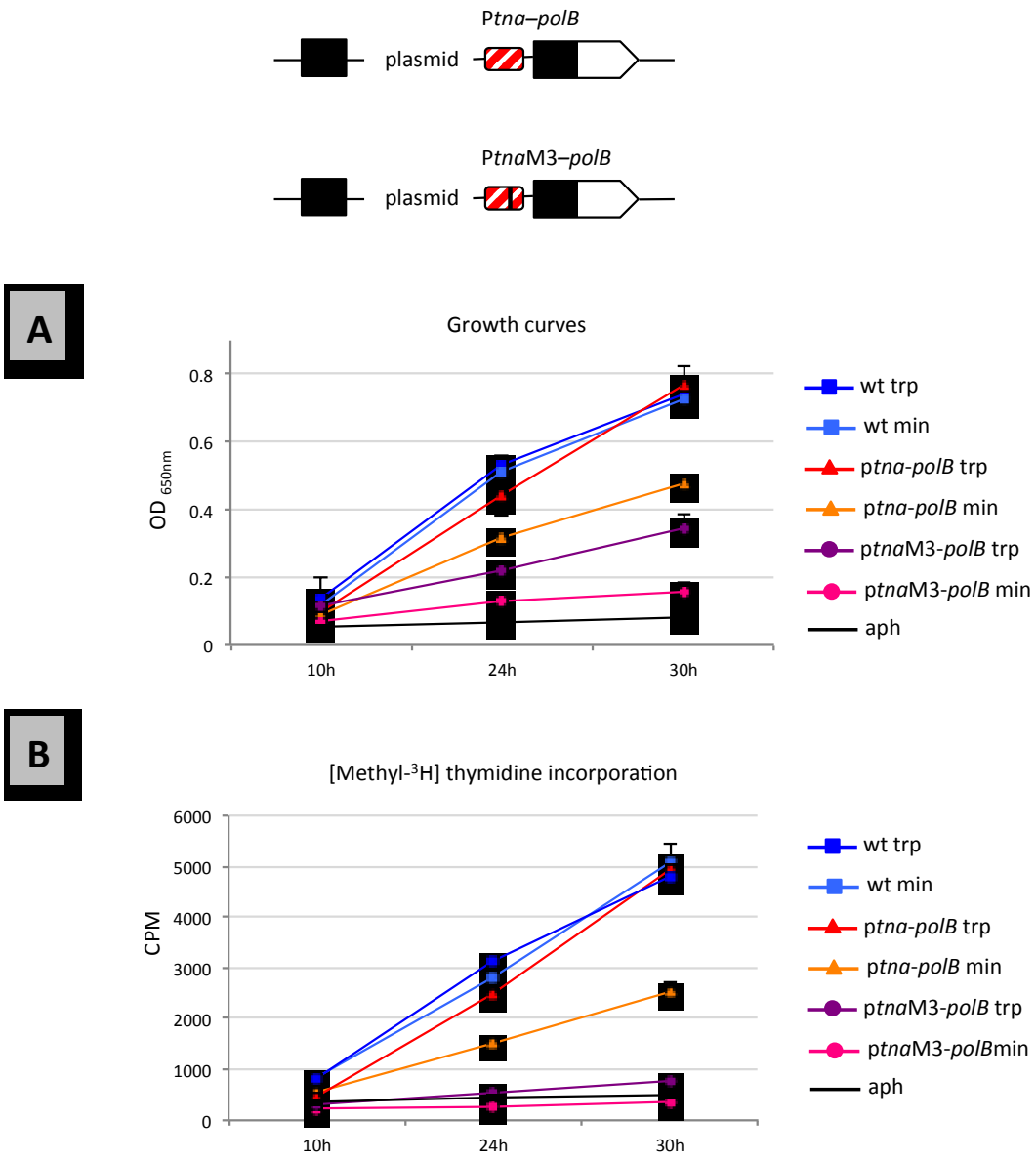
DNA replication was monitored by using scintillation counting to measure the incorporation of labelled thymidine into DNA after TCA precipitation of whole cells. As an outcome of two independent experiments we observed strong inhibition of [^{3}H]-thymidine incorporation in p*tna*M3-*PolB* cells grown in medium with and without tryptophan. The level of inhibition was comparable with those seen in the wild-type cells treated with aphidicolin. An integrant p*tna-polB* strain also showed reduction in [^{3}H]-thymidine incorporation on Hv-Min medium lacking tryptophan but not on medium containing tryptophan, in which the values of incorporation were as high as in the wild-type cells (Figure 5.3B).

The inhibition of DNA synthesis corresponds to reduction in growth rate: p*tna-polB* cells on medium lacking tryptophan, p*tna*M3-*PolB* on both types of Hv-Min medium and cells incubated with aphidicolin grew much weaker than the wild-type (Figure 5.3B).

Figure 5.3 Monitoring of DNA synthesis in p*tna-polB* and p*tna*M3-*polB* strains

A Growth of p*tna-polB* and p*tna*M3-*polB* strains in Hv-Min medium. Mean and standard deviation of three independent experiments are shown. *H.volcanii* DS70 is a wild-type control.

B [Methyl-3H] thymidine incorporation. DNA replication was monitored by using scintillation counting to measure the incorporation of labelled thymidine into DNA after TCA precipitation of whole cells. Mean and standard deviation of three independent experiments are shown. Aphidicolin was added to the wild-type cells grown on Hv-Min medium without tryptophan at the same time as [3H]-thymidine to the final concentration of 10 μg/ml.



5.2 Interplay between RPA proteins

The presence of multiple single-stranded DNA binding factors in *H.volcanii* raises the question of whether there is interplay between them. To find an answer to that question we examined whether overexpression of either *rpaA1-rpaA2-rpe* or *rpaB1-rpaB2* genes would rescue the lack of *rpaC* in the p*tna-rpaC* promoter replacement strain grown on medium lacking tryptophan. To facilitate constitutive high-level expression of *rpaA* and *rpaB* we constructed pNPM-derived plasmids carrying 5' regions of the genes inserted upstream of a 185-bp *fdx* promoter (plasmids PL27 and PL28, Table 2.3). The *fdx* promoter, derived from the ferredoxin gene from *Hbt.salinarum*, was shown to promote strong constitutive expression in *H.volcanii* (Gregor and Pfeifer, 2005). To this end, the *fdx* promoter was placed 5' to the *rpaA1-*

rpaA2-rpe and *rpaB1-rpaB2* loci in a p*tna-rpaC* background by plasmid integration (Figure 5.4 a). Serial dilutions of exponentially growing p*fdx-rpaA1-rpaA2-rpe* in p*tna-rpaC* and p*fdx-rpaB1-rpaB2* in p*tna-rpaC* strains were spotted onto HvMin plates with and without tryptophan. We found that elevated expression of *rpaB* but not *rpaA* could complement loss of RpaC: p*fdx-rpaB1-rpaB2* p*tna-rpaC* cells on medium lacking tryptophan grew as strong as wild-type, while growth of p*fdx-rpaA1-rpaA2-rpe* p*tna-rpaC* strain was strongly inhibited, comparable to that of the parental p*tna-rpaC* strain (Figure 5.4 b). The results indicate that products of the *rpaB* operon, when overexpressed, are able to perform the physiological functions of RpaC to the level required for normal cell growth.

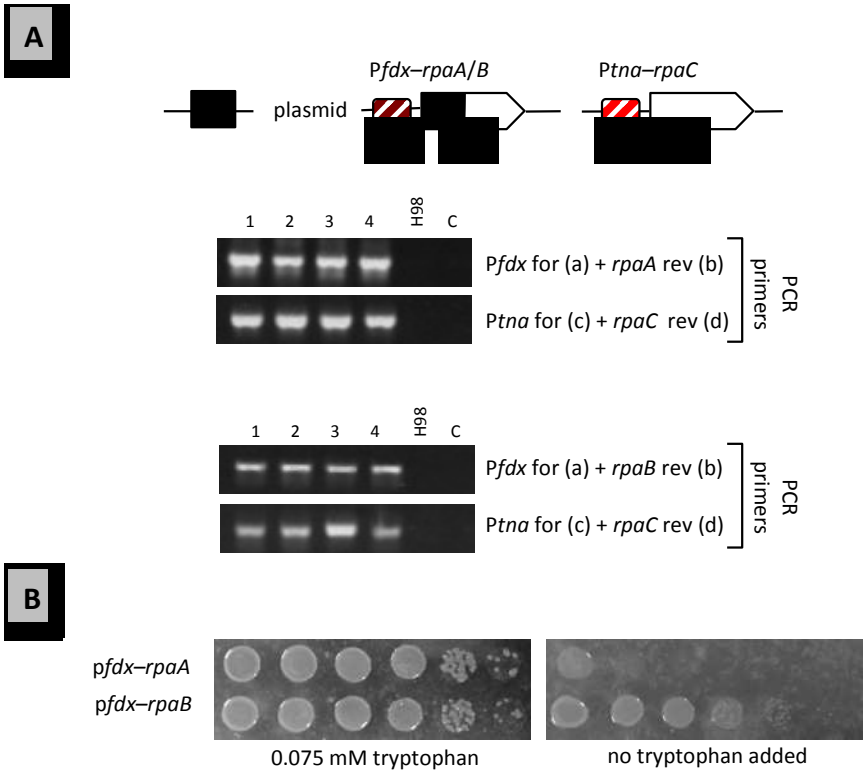


Figure 5.4 Rescue of p*tna-rpaC* strain by elevated expression of RpaB

A PCR analysis of chromosomal DNA prepared from four *rpaA1-rpaA2-rpe* p*tna-rpaC* (upper panel) and p*fdx-rpaB1-rpaB2* p*tna-rpaC* (bottom panel) colonies. Position of screening oligonucleotides is indicated by black arrows a-d.

B *H.volcanii* p*tna-rpaC* strain carrying integrated pNPM-fdx-RpaA and pNPM-fdx-RpaB plasmids (PL27 and PL28, Table 2.3) was grown up to mid-exponential phase at 45°C in HvMin medium containing 0.075 mM tryptophan. Cells were then washed twice to remove tryptophan and 10-fold serial dilutions were spotted onto HvMin plates with or without tryptophan. Plates were incubated at 45°C for 3 days.

5.3 DNA damage tests on cells overexpressing *H.volcanii* RpaC

Gene deletion analysis described in previous chapter indicated that loss of the *rpaB* locus causes increased sensitivity to DNA damage in *H.volcanii*. Consistent with that observation, previous work on *Hbt.salinarum* has shown a link between increased expression of the orthologous *rpaB* operon and cell response to UV treatment and enhanced resistance to ionizing radiation (DeVeaux et al., 2007, McCready et al., 2005). To test whether elevated level of *H.volcanii* RpaC might also lead to resistance to DNA damage, cells expressing the RpaC protein from the *tna* promoter (cells grown on medium supplemented with 0.075 mM tryptophan) or the *fdx* promoter were exposed to various DNA-damaging drugs. The p*fdx-rpaC* strain was constructed by transforming *H.volcanii* H98 with pNPM-derived plasmid carrying the 5'region of *rpaC* ORF inserted immediately downstream of the 185-bp *fdx* promoter (PL29, Table 2.3). Plasmid integration at the correct locus was confirmed by PCR using a forward oligonucleotide located in p*fdx* and reverse oligonucleotide located on *rpaC* in the chromosome (Figure 5.5A). It was already shown by RT-PCR that in p*tna-rpaC* strain the *rpaC* transcript level is elevated above the level seen in wild-type when cells are grown in medium with tryptophan (see Chapter 3). The *fdx* promoter is known to be a strong promoter in *H.volcanii*, driving constitutive high-level expression, presumably even higher than *tna* (Gregor and Pfeifer, 2005). Growth curves performed for both the p*tna-rpaC* and p*fdx-rpaC* strains showed no detectable defects in growth rate in comparison to the wild-type (Figure 5.2B). Cells expressing *rpaC* from *tna* and *fdx* promoter were exposed to a variety of DNA-damaging agents in liquid cultures and cell viability determined by colony counting. Four compounds that cause different type of DNA damage were used: UV irradiation and 4-NQO (4-nitroquinoline 1-oxide), MMS (methyl methanesulfonate) and phleomycin. Using a variety of DNA-damaging agents allow exploring all major DNA repair mechanism that *rpaC* might be involved in.

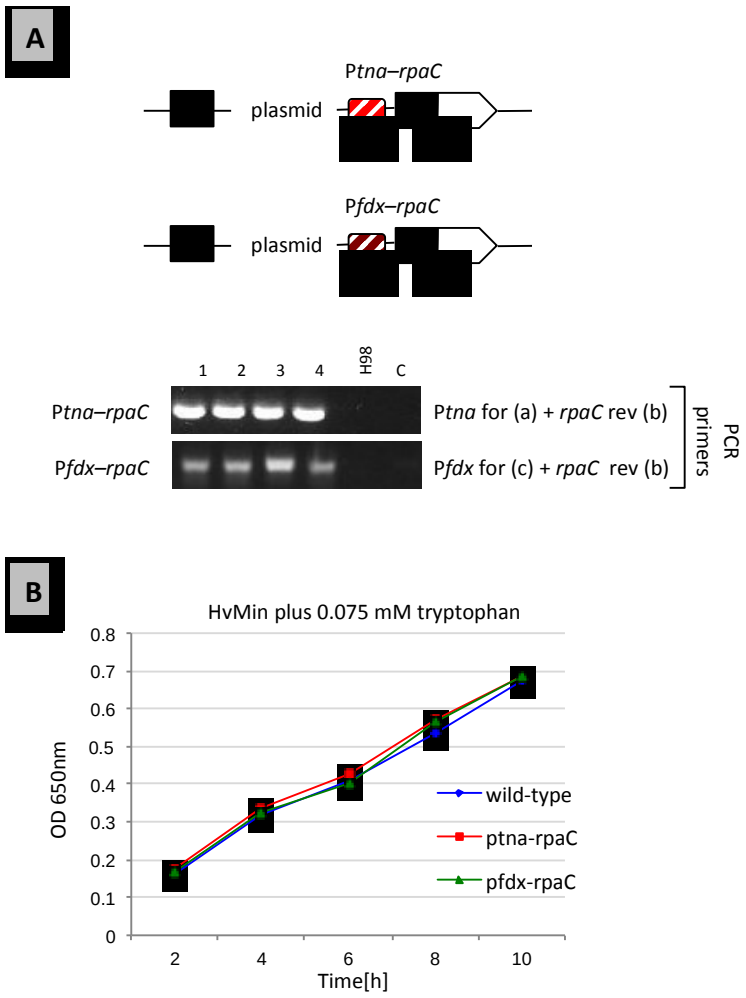


Figure 5.5 Strains overexpressing RpaC used in DNA sensitivity assays

A PCR analysis of chromosomal DNA prepared from four p*tna-rpaC* (upper panel) and p*fdx- rpaC* (bottom panel) colonies. Position of screening oligonucleotides is indicated by black arrows a-c.
B Growth curves of p*tna-rpaC* and p*fdx-rpaC* strains. Cells were grown in HvMin medium supplemented with 0.075 mM tryptophan and growth was monitored by measuring absorbance at 650 nm. Mean and standard deviations of three independent experiments are shown.

5.3.1 UV sensitivity assay

UV irradiation has a strong genotoxic effects to cause DNA damage. It introduces two of the most abundant DNA lesions: cyclobutane-pyrimidine dimers (CPDs) and 6-4 photoproducts (6-4PPs) where two pyrimidine bases are juxtaposed in tandem. These two types of lesions distort the DNA helix and if unrepaired can cause replication forks to arrest and subsequent double-strand breaks (reviewed by (Yang, 2011). Therefore, to survive on environment constantly exposed to UV irradiation, cells utilize a variety of DNA repair mechanisms. In the Archaeal domain, homologous of both bacterial and eukaryotic DNA repair genes have been identified. Species like *H.volcanii* and *Halobacterium* sp. NRC-1 display high resistance to DNA damage as being constantly challenged by intense solar radiation and the risk

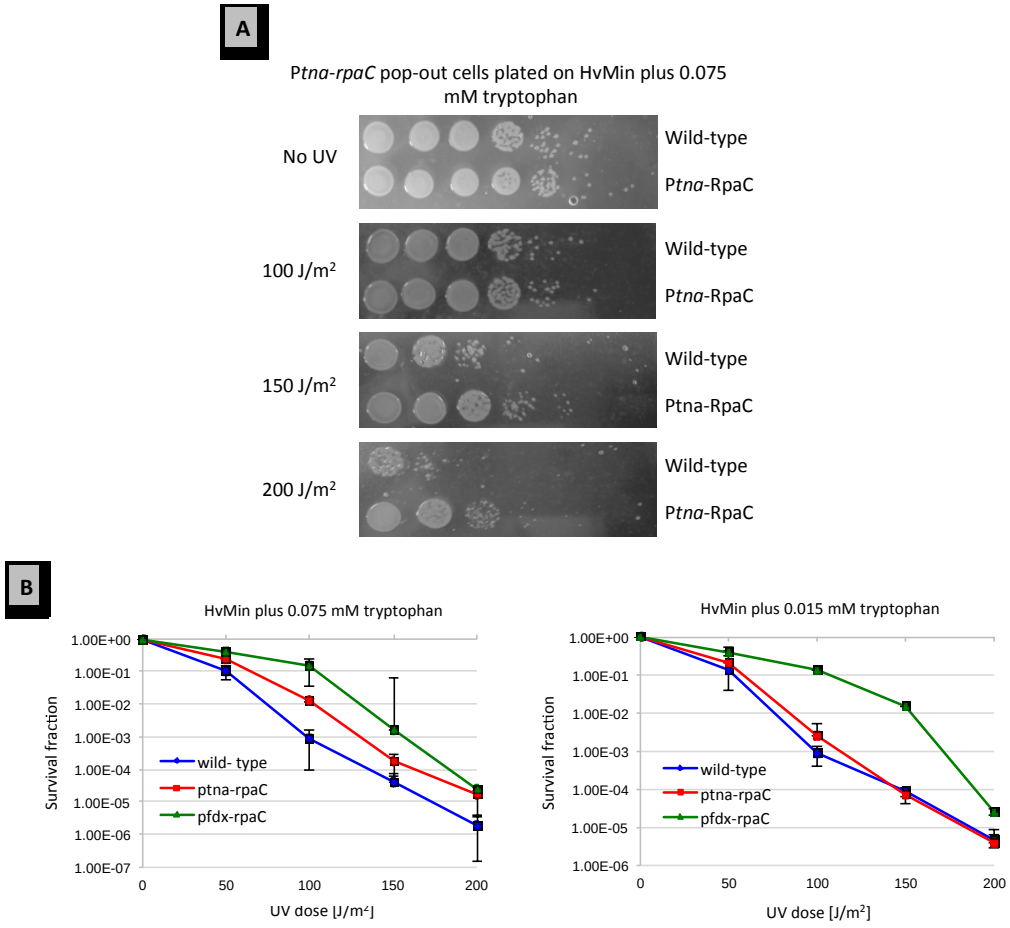
of desiccation in their natural habitat, which makes them an attractive model to study DNA repair in archaea. In haloarchaea, a key mechanism of repair of UV-induced damage is direct photoreactivation, a process dependent on the photolyase Phr2 that uses the energy of visible wavelengths of light to monomerize the pyrimidine dimers (McCready and Marcello, 2003). In addition, all haloarchaeal species encode homologous of bacterial nucleotide excision repair (UvrA, UvrB, UvrC, UvrD) and eukaryotic nucleotide excision repair (XPB, XPD, XPF) proteins (Hartman et al., 2010). Importantly, the co-occurrence of bacterial and eukaryotic NER proteins is seen only in haloarchaea and some methanogens indicating that these two groups may employ multiple DNA repair mechanisms not found in other archaea.

We exposed p*tna-rpaC* and p*fdx-rpaC* strains to various doses of UVC irradiation (from 50 to 200 J/m^2). Cells were kept in dark immediately after exposure to prevent DNA repair by photoreactivation. As shown in Figure 5.6, both strains showed enhanced resistance to UV, but the p*fdx-rpaC* strain was significantly more resistant in comparison to p*tna-rpaC*: at the dose 100 J/m^2, the wild-type survival fraction was 0.09%, while p*tna-rpaC* and p*fdx-rpaC* 1.4% and 15%, respectively. We also performed the experiment in Hv-Min medium containing a low level of tryptophan (0.015 mM), which is enough to activate p*tna* but not to cause protein overexpression. Under these conditions, UV sensitivity of p*tna-rpaC* strain was indistinguishable from the wild-type confirming that p*tna-rpaC* UV resistance seen on medium supplemented with 0.075 mM tryptophan was due to *rpaC* overproduction.

Figure 5.6 UV response of p*tna-rpaC* and p*fdx-rpaC* strains

A *H.volcanii* p*tna-rpaC* promoter replacement strain was grown up to mid-exponential phase at 45^{o}C in HvMin medium containing 0.075 mM tryptophan. After washing, ten-fold serial dilutions were spotted onto Hv-Min plates with tryptophan, irradiated with the indicated doses of UV and incubated at 45^{o}C for 3 days. *H.volcanii* DS70 is a wild-type control.

B Survival curves following UV irradiation. Cells were grown as described above and ten-fold serial dilutions were spotted onto Hv-Min plates with 0.075 mM and 0.015 mM tryptophan followed by UV irradiation. Survival was determined by colony counting after four days of incubation at 45^{o}C. Mean and standard deviation of three independent experiments are shown.



To confirm these results we repeated the experiment using 4-nitroquinoline-1-oxide (NQO) as a DNA-damaging agent that mimics the genotoxic properties of UV. P*tna-rpaC* and p*fdx-rpaC* cells were incubated for one hour in the presence of various concentrations of 4-NQO and plated on Hv-Min medium supplemented with 0.075 mM tryptophan. Consistent with the resistance pattern observed for UV irradiation, both strains expressing increased level of RpaC showed enhanced resistance to the drug; while growth of the wild-type was strongly inhibited when cells were incubated with 0.8 µg/ml 4-NQO (0.017% of cells survived), p*tna-rpaC* and p*fdx-rpaC* strains showed ten and nearly fifty times more resistance (Figure 5.7).

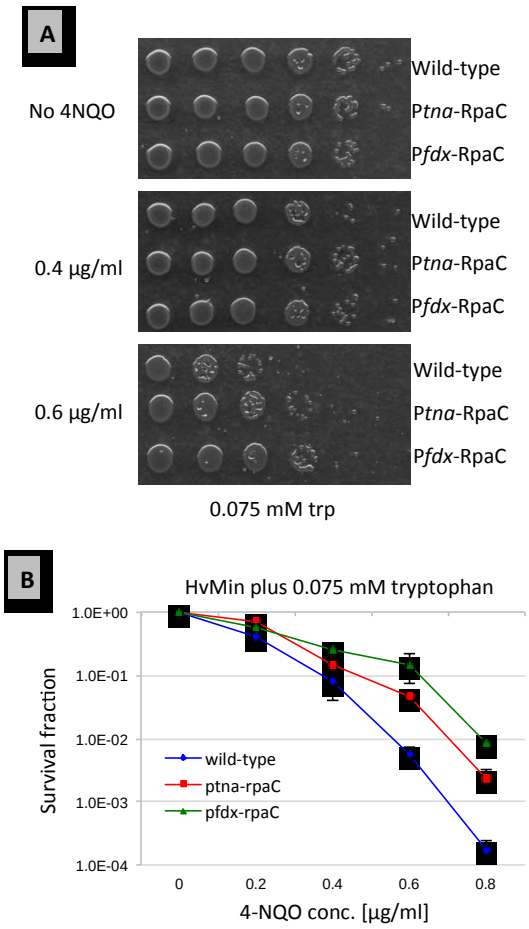


Figure 5.7 4-NQO response of p*tna-rpaC* and p*fdx-rpaC* strains

H.volcanii p*tna-rpaC* and p*fdx-rpaC* strains were grown up to mid-exponential phase at 45°C in HvMin medium containing 0.075 mM tryptophan. Cells were incubated for one hour with indicated concentration of 4-NQO and plated onto Hv-Min medium with tryptophan. *H.volcanii* H98 is a wild-type control. **A** Spotting assay, **B** Survival curves. Survival was determined by colony counting after four days of incubation at 45°C Mean and standard deviation of three independent experiments are shown.

5.3.2 MMS sensitivity test

Methyl methanesulfonate (MMS) is an alkylating drug that methylates nitrogen atoms in purines generating 7-methylguanine (7-meG) and 3-methyladenine. The latter is a strong cytotoxic lesion that blocks DNA replication and transcription. DNA damages caused by MMS are predominantly repaired by base excision repair (BER) in which damaged bases are removed by a lesion-specific DNA glycosylase (Sedgwick et al., 2007). Mid-exponentially growing p*tna-rpaC* and p*fdx-rpaC* cells was incubated with various concentrations of MMS (from 0.2 to 0.8%) for one hour followed by plating on Hv-Min medium supplemented with 0.075 mM tryptophan to maintain RpaC overproduction. A significant increase in resistance to MMS was observed in comparison to the wild-type, both in spotting assays and on survival curves, especially when *rpaC* was expressed from the *fdx* promoter. Significant differences in the surviving fraction were clearly seen when cells were treated with 0.06% and 0.08% MMS: MMS at concentration 0.06% caused lethality of 99.6% of the wild-type cells whereas p*fdx-rpaC* cells survival fraction was still as high as 7% (Figure 5.8).

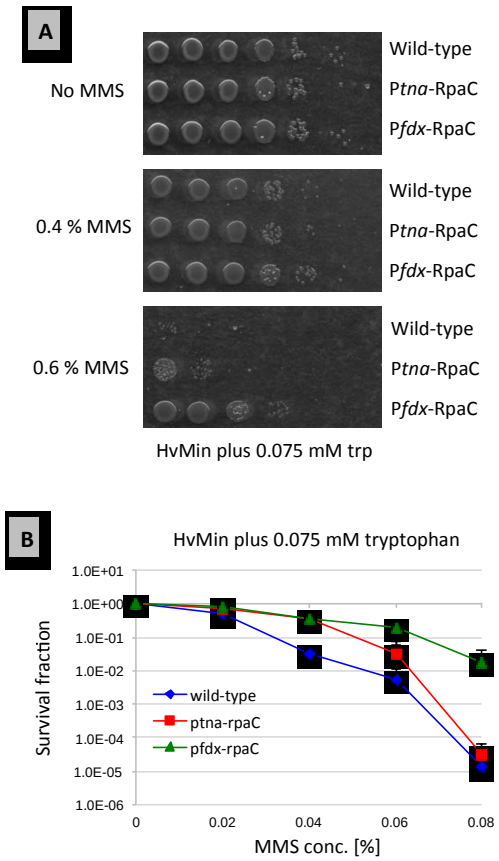


Figure 5.8 MMS response of p*tna-rpaC* and p*fdx-rpaC* strains

H.volcanii p*tna-rpaC* and p*fdx-rpaC* strains were grown up to mid-exponential phase at 45°C in HvMin medium containing 0.075 mM tryptophan. Cells were incubated for one hour with indicated concentration of MMS and plated onto Hv-Min medium with tryptophan. *H.volcanii* H98 is a wild-type control. **A** Spotting assay, **B** Survival curves. Survival was determined by colony counting after four days of incubation at 45°C Mean and standard deviation of three independent experiments are shown.

5.3.3 Phleomycin sensitivity test

Ionizing radiation is one of the most powerful sources of DNA damage that causes double strand breaks (DSBs) and DNA fragmentation. However in the laboratory, usage of γ-radiation is often limited and is replaced by radiomimetic drug phleomycin. Phleomycin is a mixture of copper-containing glycopeptides isolated from *Streptomyces verticillis* (Sleigh, 1976). The lethal dose of phleomycin for most aerobically growing cells ranges between 0.1 and 50 µg/ml but haloarchaea (which are in general also resistant to ionizing radiation) are able to thrive in higher concentrations of phleomycin. The major pathway to repair DSBs is homologous recombination (HR). Briefly, in Archaea, the resection of the 5'-ends that initiate HR is performed by the Rad50-Mre11-HerA-NurA complex, the strand invasion and exchange steps are promoted by RadA, and the junction-resolving enzyme is Hjc (reviewed by (White, 2011). While Rad50 and Mre11 are conserved proteins between archaea and eukaryotes, HerA helicase, NurA nuclease and Hjc resolvase are unique in the third domain of life. Interestingly, DSBs repair is more enigmatic in haloarchaea. In *H.volcanii*, a *Δmre11 Δrad51* double deletion mutant is more resistant to DNA damage than the wild-type suggesting that *H.volcanii* prefers other (still unspecified) pathway than HR to repair DSBs (see Discussion section).

To test whether elevated expression of *rpaC* leads to increased resistance to double strand breaks we incubated mid-exponential growing p*tna-rpaC* and p*fdx-rpaC* cells with two different concentrations of phleomycin, plated serial diluted cells on Hv-Min medium with 0.075 mM tryptophan and estimated survival fraction by counting colonies that have raised on plates after appropriate time of incubation. As an outcome from these experiments, both strains showed an increase in resistance to the drug in comparison to the wild-type (Figure 5.9).

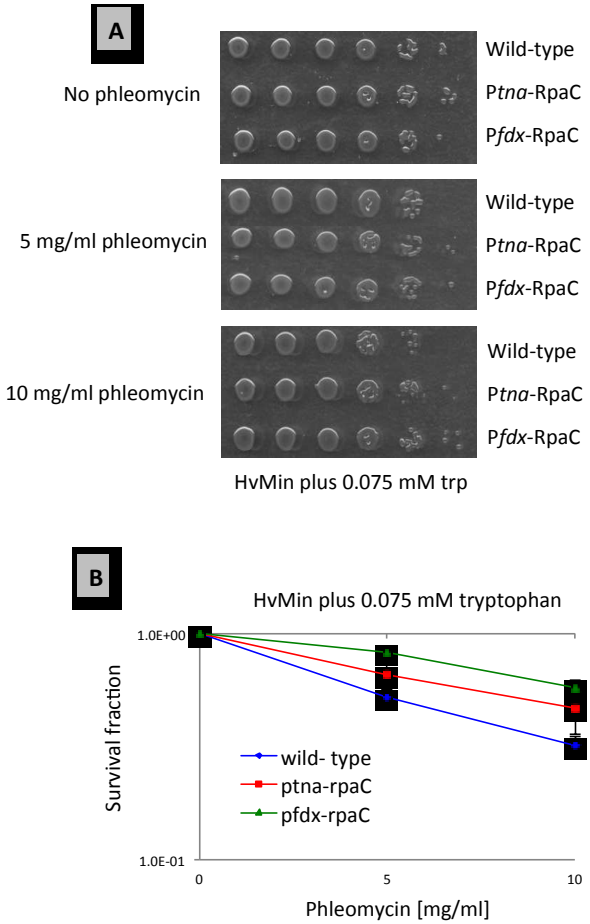


Figure 5.9 Phleomycin response of p*tna-rpaC* and p*fdx-rpaC* strains

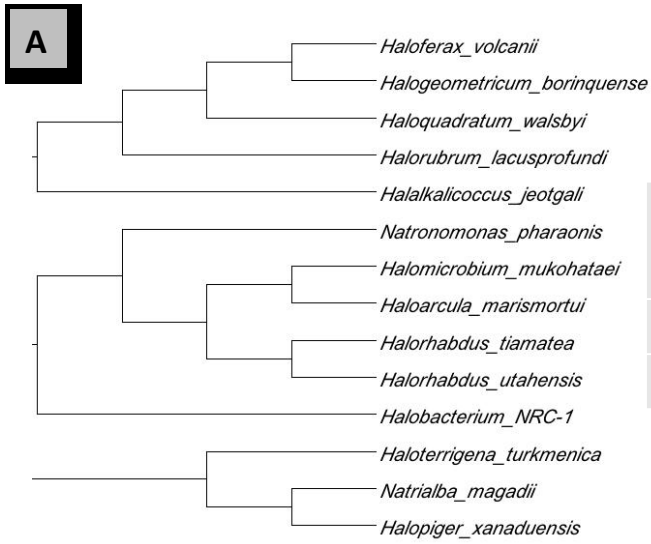
H.volcanii p*tna-rpaC* and p*fdx-rpaC* strains were grown up to mid-exponential phase at 45°C in HvMin medium containing 0.075 mM tryptophan. Cells were incubated for one hour with indicated concentrations of phleomycin and plated onto Hv-Min medium with tryptophan. *H.volcanii* H98 is a wild-type control. **A** Spotting assay, **B** Survival curves. Survival was determined by colony counting after four days of incubation at 45°C Mean and standard deviation of three independent experiments are shown.

5.4 Cross-species complementation

5.4.1 Complementation of *H.volcanii* RpaC by the full-length *H.borinquense* and *H.walsbyi* RpaC orthologues

In many model organisms, construction of a conditional-lethal mutant strain would ordinarily facilitate rapid structure-function analysis of the corresponding protein by providing *in trans* expression of mutated/truncated version of the gene of interest. However, *H.volcanii* possesses highly efficient homologous recombination that

causes gene conversion between the wild-type chromosomal copy of the gene and incoming mutated copies. This often results in loss of the mutant DNA sequence before any functional test can be applied (unpublished lab results). To bypass these difficulties, we identified RpaC orthologues in haloarchaeal species closely related to *H.volcanii* in order to test whether these proteins, displaying only limited sequence similarity at the nucleotide sequence level, would function in *H.volcanii* and rescue growth of the p*tna-rpaC* strain grown in the absence of tryptophan. The RpaC orthologues from *Halogeometricum borinquense* (Hbor_26830; GI 313127408) and *Haloquadratum walsbyi* (HQ_1435A; GI:110667397) showed the highest sequence identity to *H.volcanii* protein at amino acid sequence (75% and 59%, respectively, summarised in Figure 5.10A). We constructed replicating plasmids carrying these genes under the control of the *pfx* promoter (PL67 and PL68, Table 2.3). The full-length ORFs were amplified by PCR using the appropriate genomic DNA as a template and oligonucleotides P92-P95 (Appendix, Table). The PCR products were digested with *Nde*I and *Hind*III enzymes and ligated together with the *BamH*I-*Nde*I restriction fragment of *fdx* promoter into plasmid pTA230 (Allers et al., 2004) digested with *BamH*I and *Hind*III. Corresponding plasmid carrying the full-length *H.volcanii rpaC* (Hfx RpaC) was also constructed to serve as a control in the complementation experiment (PL66, Table 2.3). The resulting plasmids were then transformed into the p*tna-rpaC* promoter replacement strain and ability of transformant cells to grow was re-tested on medium containing and lacking tryptophan. As seen in Figure 5.10B, *in trans* expression of either *Halogeometricum borinquense* RpaC (Hbo RpaC) or *Haloquadratum walsbyi* RpaC (Hwa RpaC) rescues growth of the p*tna-rpaC* strain. As expected, no rescue was seen when the cells were transformed with the empty plasmid. The level of rescue observed with Hbo RpaC was indistinguishable from that seen when Hfx RpaC was expressed under equivalent conditions (data not shown).



	Hvo RpaC	
	sequence identity	sequence similarity
Hbo	75%	83%
Hwa	59%	74%

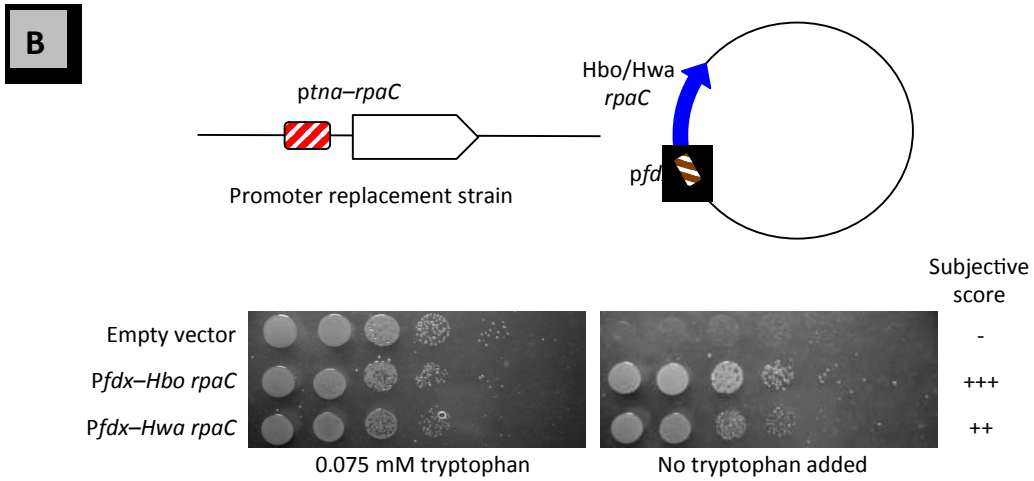


Figure 5.10 Rescue of p*tna-rpaC* strain by *H.borinquense* and *H.walsbyi* RpaC orthologues

A Phylogenetic tree based on the amino acid sequences of RpaC orthologues from fourteen haloarchaeal species showing close relevance between *H.volcanii* (Hvo), *H.borinquense* (Hbo) and *H.walsbyi* (Hwa) RpaC proteins. The tree was generated using Dendroscope. On the right, table summarizing sequence identity and similarity between Hvo RpaC and Hbo and Hwa RpaC orthologues.
B *H.volcanii* p*tna-rpaC* strain carrying pTA230-fdx-Hbo-rpaC and pTA230-fdx-Hwa-rpaC plasmids (PL67 and PL68, Table 2.3) was grown up to mid-exponential phase at 45°C in HvMin medium containing 0.075 mM tryptophan. Cells were then washed twice to remove tryptophan and 10-fold serial dilutions were spotted onto HvMin plates with or without tryptophan. Plates were incubated at 45°C for 3 days. *H.volcanii* p*tna-rpaC* strain carrying empty pTA230 plasmid is shown as a control.

5.4.2 Functional analysis of mutated RpaC proteins

Haloarchaeal RpaC orthologues display a highly conserved domain organization: three OB folds (OB folds A-C) and distinct N- and C-terminal domains (Figure 5.11). The N-terminal domain is found among haloarchaea and also some methanogens, including *Methanosarcinales*. The C-terminal domain is unique in the haloarchaea and might be a highly divergent form of the OB fold (see Discussion section); it is separated from the OB fold C by a non-conserved linker sequence of variable length.

The sequence alignment of RpaC orthologues from fourteen haloarchaeal species with an indication of all structural motifs is presented in Figure A1 (Appendix, Figure A1). Construction of OB fold-deletion proteins in *M.acetivorans* and *M.kandleri* have shown that in multiple OB fold-containing RPAs, one OB fold can be deleted without significant loss of ssDNA-binding abilities (Lin et al., 2008). However, it is not known how these changes to the structure of the protein influence *in vivo* function. To examine this, plasmids for expression seven deleted derivatives of *H.borinquense* RpaC were constructed. In addition, to test the functions of the N- and C-terminal domains we also deleted these parts of the protein. To summarize, the mutated ORFs expressed single OB fold deletions (ΔOB^{A}, ΔOB^{B} and ΔOB^{C}), double OB fold deletions (ΔOB^{AB}, ΔOB^{BC} and ΔOB^{AC}) and a triple OB deletion (ΔOB^{ABC}), N- and C-terminal deletions (ΔNTD and ΔCTD respectively) (Table 5.1 and Figure 5.9).

Table 5.1 Mutated derivatives of *H.borinquense* RpaC

Deleted/truncated RpaC	Sequence changes
Hbo RpaC-ΔNTD	deletion of amino acids 1–62 inclusive
Hbo RpaC-ΔOB^{A}	deletion of amino acids 58–167
Hbo RpaC-ΔOB^{B}	deletion of amino acids 168–267
Hbo RpaC-ΔOB^{C}	deletion of amino acids 268–367
Hbo RpaC-ΔOB^{AB}	deletion of amino acids 58–267
Hbo RpaC-ΔOB^{AC}	deletion of amino acids 58–167 and 268–367
Hbo RpaC-ΔOB^{BC}	deletion of amino acids 168–367
Hbo RpaC-ΔOB^{ABC}	deletion of amino acids 58–367
Hbo RpaC-ΔCTD	deletion of amino acids 405-483

The plasmid for expression of Hbo RpaC-ΔNTD was generated by excising the full-length *H.volcanii rpaC* gene from pTA230-HvoRpaC plasmid, and replacing it with RpaC ORF, lacking 186 nucleotides from N terminus, amplified by PCR as *Nde*I-*Hind*III fragment using plasmid pTA230-HboRpaC as a template and oligonucleotides P93 and P108 (Appendix, TableA1). Plasmids expressing OB fold-mutant forms of the RpaC protein were constructed by overlap extension PCR mutagenesis using plasmid pTA230-HboRpaC as a template and oligonucleotides P92 and P93 in combination with oligonucleotides P96–P107 (Appendix, TableA1). The mutated ORFs were re-cloned, as *Nde*I-*Hind*III fragments, into pTA230-

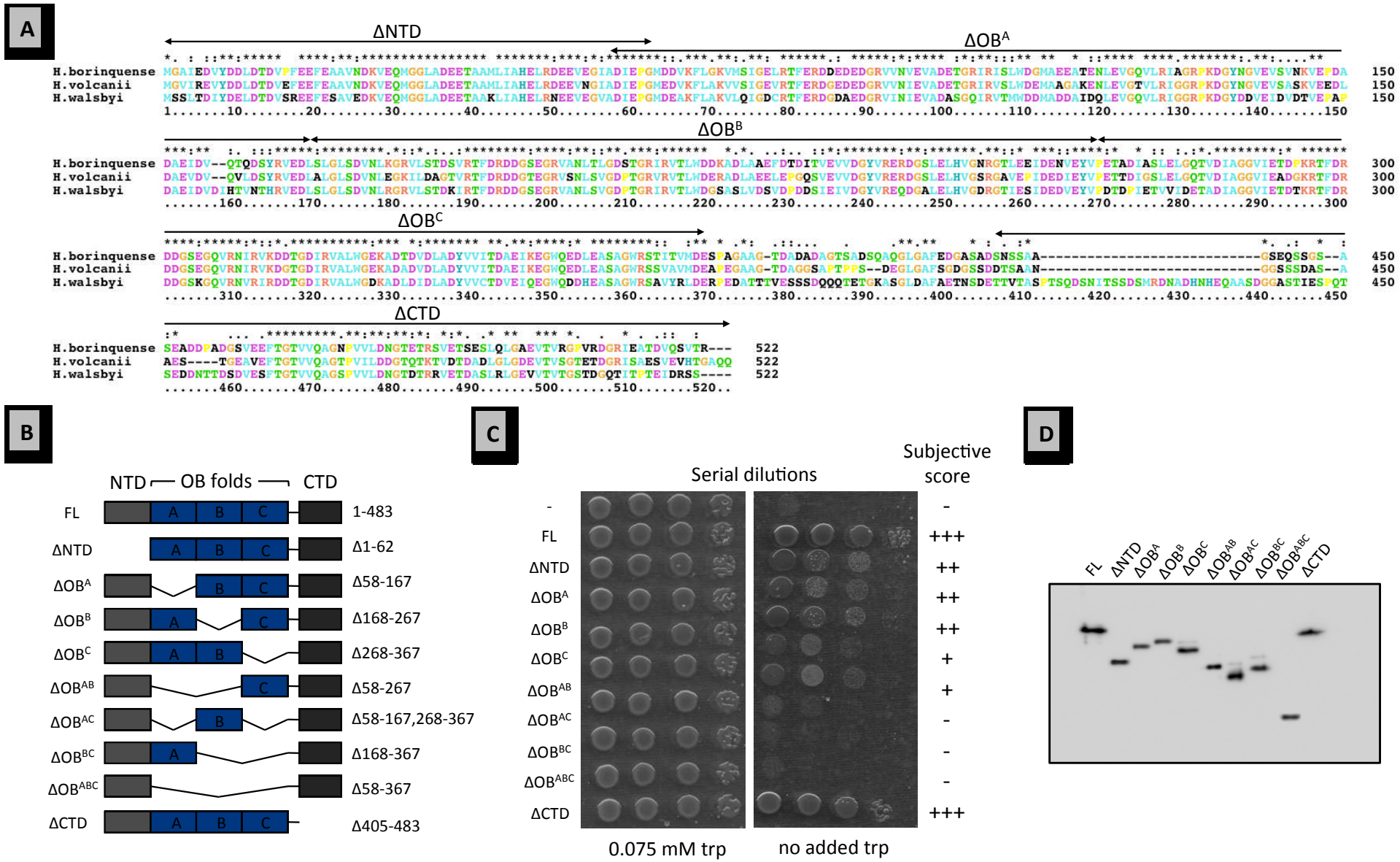
HvoRpaC plasmid, replacing the full-length *H.volcanii rpaC* gene carried by the plasmid. The resulting plasmids (pTA230-Hgm RpaC-ΔOB^{A}, etc., PL70-PL76, Table 2.3) were sequenced to ensure the absence of unwanted sequence changes. The ORF to express Hbo RpaC-ΔCTD was constructed using internal *EcoR*I restriction site located in the gene at a position that corresponds to the poorly conserved linker between the third OB fold and the CTD. Plasmid expressing Hbo RpaC-ΔCTD was constructed by digesting plasmid pTA230-HboRpaC with *EcoR*I and ligating in a short dsDNA with *EcoR*I cohesive ends assembled from oligonucleotides P109 and P110 (Appendix, TableA1), containing a stop codon. Correct insertion of stop codon-containing DNA was confirmed by sequencing.

Plasmids expressing each of the mutated or truncated *rpaC* ORFs from the *fdx* promoter were transformed into the *H.volcanii* RpaC *tna*-promoter replacement strain and transformants tested for their ability to restore growth of the cells in medium lacking tryptophan. The obtained results were as follows. None of the three OB folds is individually essential for RpaC function- all three single domain deletions (ΔOB^{A}, ΔOB^{B} and ΔOB^{C}) rescued growth of the p*tna-rpaC* strain, although in each deletion the degree of rescue was reduced in comparison to the full-length *H.borinquense* RpaC protein indicating that the presence of all three OB folds is required for optimal RpaC function. Interestingly, OB fold C appears to have the most important role as the growth of the cells expressing single ΔOB^{C} was weaker than the two other single deletions and no growth was observed when ΔOB^{AC} and ΔOB^{BC} were expressed, while ΔOB^{AB} retained some rescue activity. No rescue was observed in cells expressing the triple OB fold deletion mutant ΔOB^{ABC}. The N- and C-terminal domains are individually non-essential for RpaC function, however cells expressing Hbo RpaC-ΔNTD grew more poorly than cells expressing the full-length protein. The growth of cells expressing ΔCTD was indistinguishable from the wild-type (Figure 5.11B).

To ensure that all nine *H.borinquense* RpaC mutant proteins are expressed in *H.volcanii* we constructed the N-terminally Flag-tagged versions of these proteins (see Materials and Methods 2.2.8 for details of plasmid construction). The resulting proteins were expressed with the N-terminal sequence M_1DYKDDDDKHM_2, where M_2 corresponds to the native N-terminal methionine (Met63 in the case of the RpaC-ΔNTD protein). Subsequent western blotting confirmed that all nine proteins were expressed to similar levels in *H.volcanii* (Figure 5.11C).

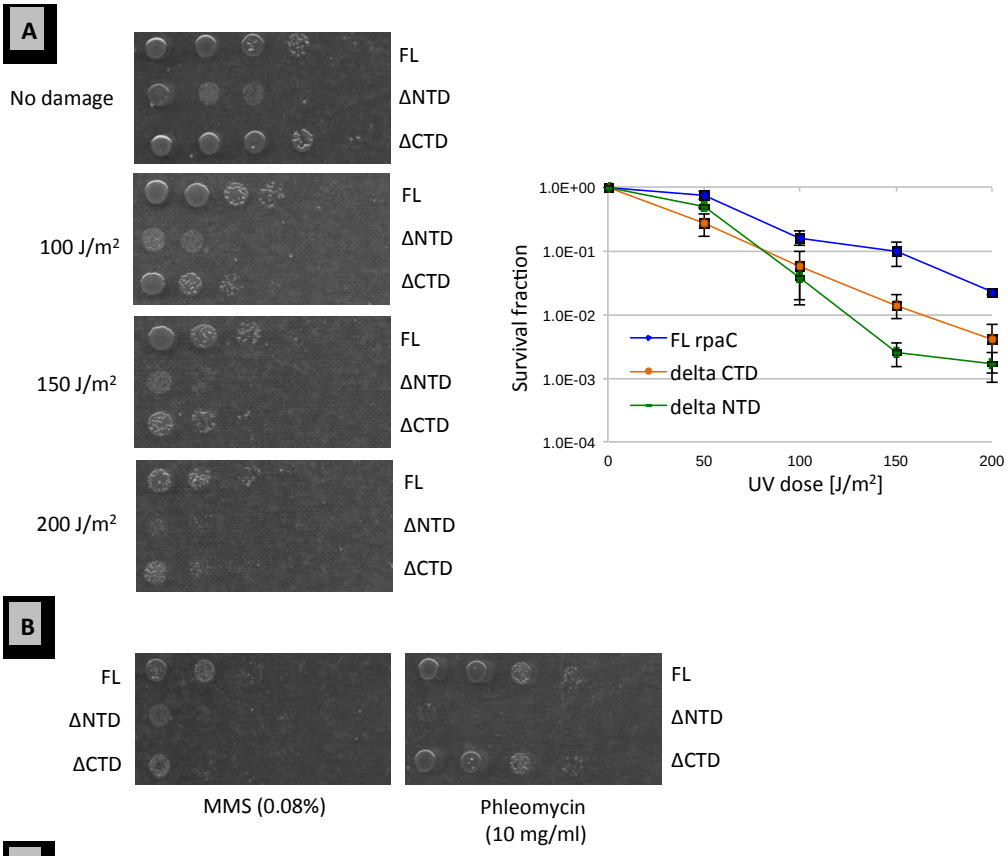
Figure 5.11 Complementation of *H.volcanii* RpaC by mutated *H.borinquense* orthologues

A Multiple sequence alignment of *H.borinquense*, *H.volcanii* and *H.walsbyi* RpaC proteins. Arrows above the sequence indicate residues deleted in **B**. Sequence alignment was generated using Clustal X2.0.12 (Larkin et al., 2007) with default parameters.
B Schematic representation of nine mutant *H.borinquense* RpaC proteins indicating the location of the N-terminal domain (NTD), OB folds A-C and C-terminal domain (CTD) and the extent of the deletion regions. The short black line indicates a linker sequence between OB fold C and CTD.
C *H.volcanii* p*tna-rpaC* strain carrying various pTA230-fdx-HboRpaC plasmids (PL78-PL87, Table 2.3) was grown up to mid-exponential phase at 45^{o}C in HvMin medium containing 0.075 mM tryptophan. Cells were then washed twice to remove tryptophan and 10-fold serial dilutions were spotted onto HvMin plates with or without tryptophan. Plates were incubated at 45^{o}C for 3 days.
D Western blot analysis of N-terminally Flag-tagged full-length *H.borinquense* RpaC and its nine mutant versions expressed from *fdx* promoter in *H.volcanii*.


A
ΔNTD
ΔOB^A
ΔOB^B
ΔOB^C
ΔCTD
H.borinquense
H.volcanii
H.walsbyi
B
NTD
OB folds
CTD
FL
ΔNTD
ΔOB^A
ΔOB^B
ΔOB^C
ΔOB^AB
ΔOB^AC
ΔOB^BC
ΔOB^ABC
ΔCTD
1-483
Δ1-62
Δ58-167
Δ168-267
Δ268-367
Δ58-267
Δ58-167,268-367
Δ168-367
Δ58-367
Δ405-483
C
Serial dilutions
Subjective score
0.075 mM trp
no added trp
D
FL
ΔNTD
ΔOB^A
ΔOB^B
ΔOB^C
ΔOB^AB
ΔOB^AC
ΔOB^BC
ΔOB^ABC
ΔCTD

5.4.3 DNA damage sensitivity of ΔNTD and ΔCTD RpaC

Although the RpaC N- and C-terminal domains are non-essential for cell viability, both domains are well conserved across haloarchaeal species, suggesting that they play important role *in vivo*. To elucidate the possible role of this motifes in DNA repair, it was examined, whether p*tna-rpaC* strains expressing truncated *H.borinquense* RpaC-ΔNTD and RpaC-ΔCTD proteins are more sensitive to the DNA damage caused by various agents than cells expressing full-length RpaC. The results of these experiments were as follows. Cells expressing RpaC-ΔNTD showed increased sensitivity to UV irradiation, MMS and phleomycin, whereas cells expressing RpaC-ΔCTD protein were more sensitive to UV and MMS but not to phleomycin (Figure 5.12).



C

RpaC ΔNTD		RpaC ΔCTD	
Growth rate	Weak growth	Growth rate	as the wild-type
UV	Sensitive	UV	Sensitive
MMS	Sensitive	MMS	Sensitive
Phleomycin	Sensitive	Phleomycin	as the wild-type

Figure 5.12 Sensitivity of cells expressing RpaC-ΔNTD and RpaC-ΔCTD to DNA damage.

A Response to UV treatment. *H.volcanii* p*tna–rpaC* cells carrying various pTA230-fdx-HboRpaC plasmids were grown up to mid-exponential phase in Hv-Min medium containing 0.075mM tryptophan. The cells were then washed and ten-fold serial dilutions spotted onto Hv-Min plates, irradiated with the indicated dose of UV and incubated at 45^{o}C for 3 days. On the left: spotting assay, on the right survival curves. Survival was determined by colony counting after four days of incubation at 45^{o}C. Mean and standard deviation of three independent experiments are shown.
B Response to MMS and phleomycin treatment. The same strains were grown up and exposed to MMS (0.08%) or phleomycin (10 mg/ml) for 1 h, before being serially diluted, spotted and incubated at 45^{o}C for 3 days.
C Tables summarizing growth properties and sensitivity to DNA damage of the cells expressing RpaC-ΔNTD (left panel) and RpaC-ΔCTD (right panel) proteins.

Our results suggest that the N-terminal domain may have a general role in RpaC function in DNA repair while the role of the C-terminal domain may be limited to certain types of DNA repair only. However, the exact role of RpaC protein and its N- and C-terminal domain has to be confirmed and further explored.

5.5 Expression of *H.volcanii* full-length RpaC and RpaC NTD

A highly conserved feature of the haloarchaeal and also methanogenic RpaC proteins is the presence of the N-terminal domain (NTD). The NTD consists of sixty residues with centres on a fifteen amino acid sequence (residues 27-41 in the *H.volcanii* RpaC protein) that is identical in fourteen haloarchaeal RpaC proteins in current databases (Figure 5.13 left panel). Although the crystal structure of RpaC is not available, structure modelling using Phyre2 (Kelley and Sternberg, 2009, Bennett-Lovsey et al., 2008) revealed, albeit with low confidence, that the NTD may form a three-helix bundle structure, with the perfectly conserved fifteen amino acid sequence spanning the C-terminal end of the second helix, an unstructured loop and the N-terminal end of the third helix (Figure 5.13 right panel). Our structure-function analysis of truncated RpaC derivatives showed that deletion of the NTD impairs protein function; cells expressing RpaC ΔNTD are more sensitive to various types of DNA damage. Thus, we hypothesised that this domain may play a role in DNA repair pathways as a platform for protein-protein interaction. To elucidate the exact pathways that RpaC participates in via NTD, NTD was expressed in order to screen for proteins that interact directly with the NTD.

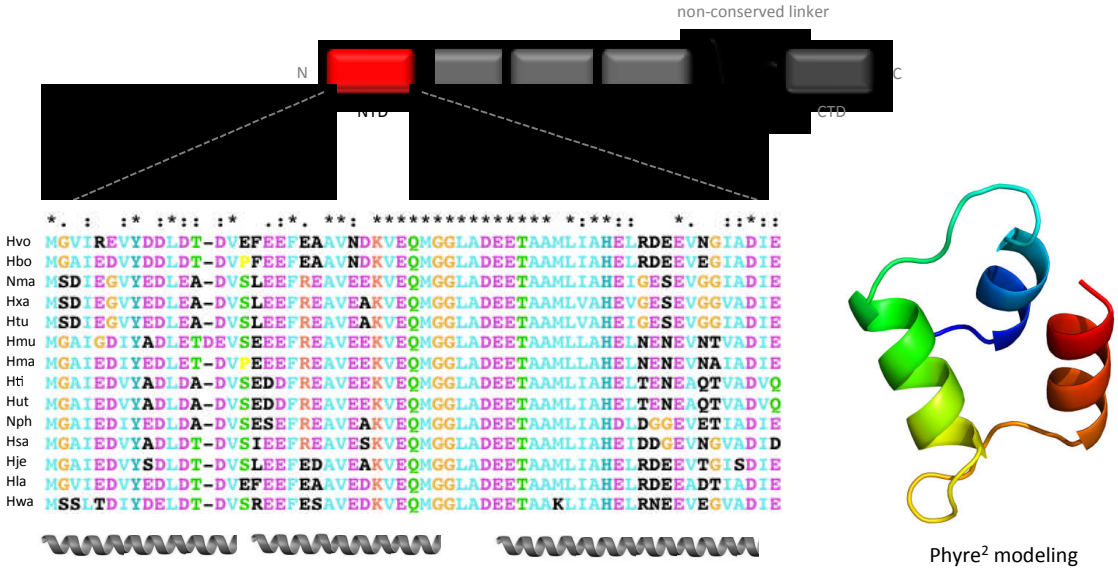


Figure 5.13 Sequence and predicted structure of RpaC N-terminal domain (NTD)

Left panel: The multiple sequence alignment of NTD (residues 1-60) of RpaC proteins from fourteen haloarchaeal species (see Table A3 for accession numbers of all proteins). Sequence alignment was generated using Clustal X 2.0 (Larkin et al., 2007) with default parameters.
Right panel: Predicted model of *H.volcanii* RpaC NTD structure obtained using Phyre2 (Bennett-Lovsey et al., 2008, Kelley and Sternberg, 2009). Image coloured by rainbow from N to C terminus. 69% of the sequence has been modeled with 26.7% confidence by the single highest scoring template. The position of the three helices predicted in Phyre2 model is indicated by the grey helix symbol below the sequence alignment.

The initial attempt was to express the *H.volcanii* RpaC NTD with an C-terminal Strep-tag (pASK-IBA vector) in *E.coli* Rosetta 2 but protein was not expressed under any conditions tested, despite using a synthetic gene with the sequence optimized to improve the efficiency of gene expression in *E.coli* (data not shown). Instead, the RpaC NTD was expressed in the native host using the cellulose-binding domain (CBD) of the *Clostridium thermocellum* cellulosome as a tag. The CBD has ability to interact with cellulose in a salt-insensitive manner, which is particularly desirable for expressing halophilic proteins (Irihimovitch and Eichler, 2003, Morag et al., 1995). To facilitate future expression of CBD-fused proteins, plasmid pTA233-CCBD (PL88, Table 2.3), carrying the *cbd* gene with two stop codons introduced and *fdx* promoter inserted 5' to the *cbd*, was constructed. The *cbd* gene was PCR amplified from plasmid pWL-CBD-Sec11b, which was a kind gift from Professor Jerry Eichler (University of the Negev) (see Materials and Methods 2.2.10.1 section for details of plasmid construction). The pTA233 plasmid contains a *H.volcanii* replication origin and the *hdrB* auxotrophic marker for selection of transformed cells in Hv-YPC rich

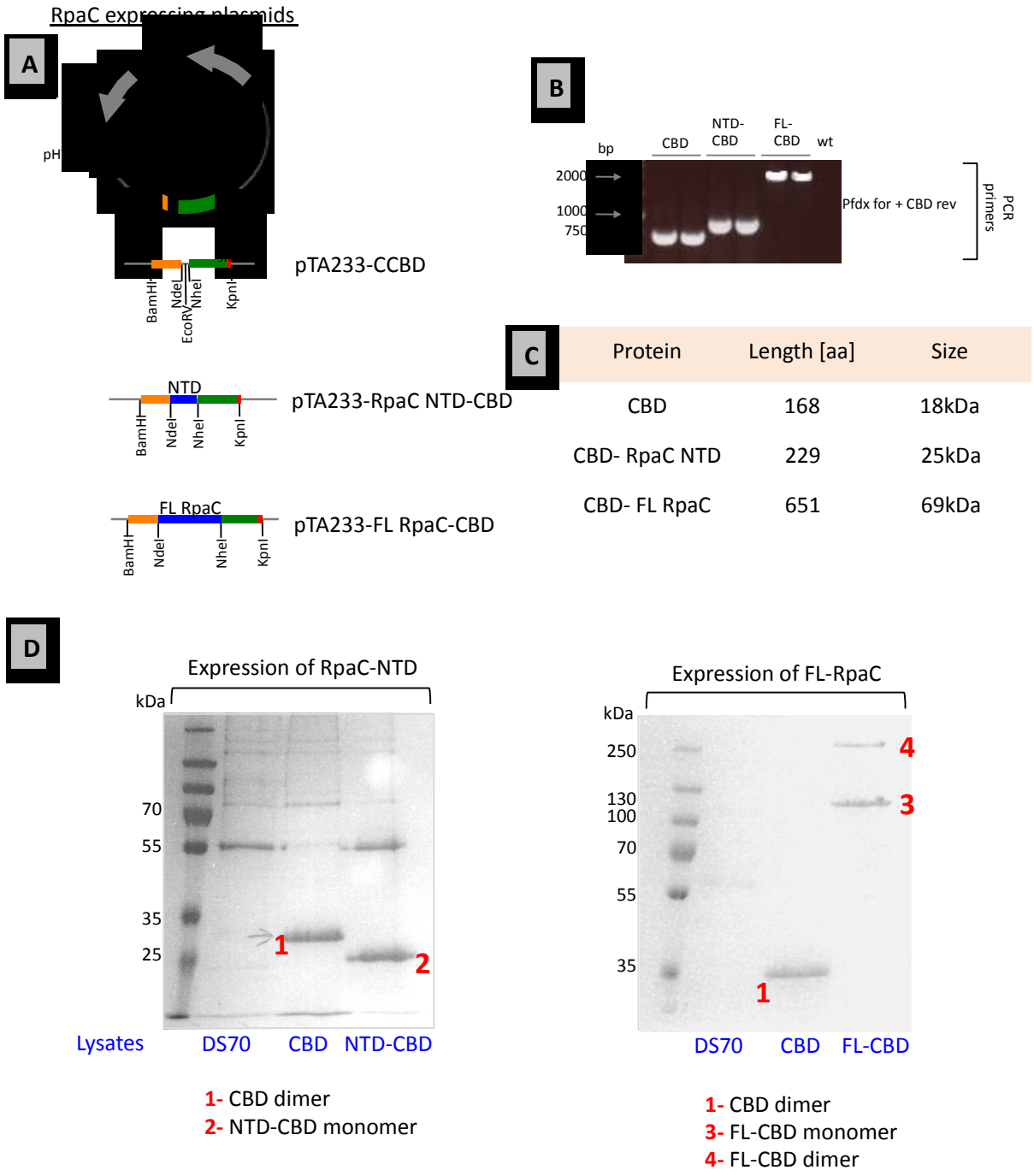
medium. Next, the full-length *rpaC* ORF (lacking the native stop codon) and *rpaC* NTD were PCR-amplified from *H.volcanii* DS70 genomic DNA and cloned in the centre of p*fdx-cbd* cassette so the *cbd* was fused at the C-terminus of *rpaC* (PL89 and PL90, Table 2.3, Figure 5.14A). All three plasmids were then transformed into *H.volcanii* H26 and transformants were obtained on HvCa plates before being purified by streaking to single colonies on Hv-YPC medium.

To assess expression of CBD, RpaC NTD-CBD and FL-CBD fusion proteins, the cell lysates from 100 ml of exponentially growing *H.volcanii* cells transformed with the appropriate plasmid were incubated with 200 µl of cellulose beads for 1h and after several washing steps, beads were resuspended in 2xSB and analysed by SDS-PAGE.

Coomassie staining revealed the presence of a bands corresponding in size to CBD (15 kDa), NTD-CBD (25 kDa) and FL-CBD (69 kDa) (Figure 5.14). In some experiments, protein bands corresponding in size to CBD, NTD-CBD, FL-CBD dimers could be also detected. Their appearance, however, was not consistent. To validate the identity of these proteins, the polypeptides were excised from the gel, subjected to trypsin digestion and analysed by MALDI-MS and MALDI-MS/MS mass spectrometry. The results confirmed the identification of CBD, NTD-CBD and FL-CBD proteins (data not shown).

Figure 5.14 A small-scale expression of CBD, RpaC NTD-CBD and FL RpaC-CBD proteins

A Schematic representation of plasmids pTA233-CCBD, pTA233-RpaC NTD-CCBD and pTA233-FL-CCBD for protein expression. For expression of RpaC proteins, the coding sequences are inserted between the *Nde*I-*Nhe*I sites 5' to the *fdx* promoter and 3' to the *cbd* tag. See Materials and Methods chapter for details of the plasmid construction.
B PCR analysis of DNA prepared from two randomly picked *H.volcanii* colonies transformed with pTA233-CCBD, pTA233-RpaC NTD-CCBD and pTA233-FL-CCBD plasmids confirming the presence of expressing cassettes.
C Table summarizing the length and predicted size of expressed proteins.
D SDS-PAGE gels of the CBD-tagged proteins expressed in 100 ml cultures of *H.volcanii* transformant strains and purified on cellulose beads as described in Materials and Methods chapter. Samples were separated on 10% acrylamide gel and bands were visualized by Coomasie Blue staining. Lysate from *H.volcanii* DS70 wild-type strain was used as a control. Bands 1-4 were identified based on the size and confirmed by mass spectrometry.



Strong expression of CBD-tagged RpaC proteins seen in small-scale experiment encouraged us to perform large-scale purification using 1L cultures of appropriate strains (100 ml culture expressing CBD-fusion combined with 900 ml of the wild-type culture). Unfortunately, several attempts failed to identify any bands unique for NTD-CBD or FL-CBD that might indicate protein-protein interaction (Figure 5.15).

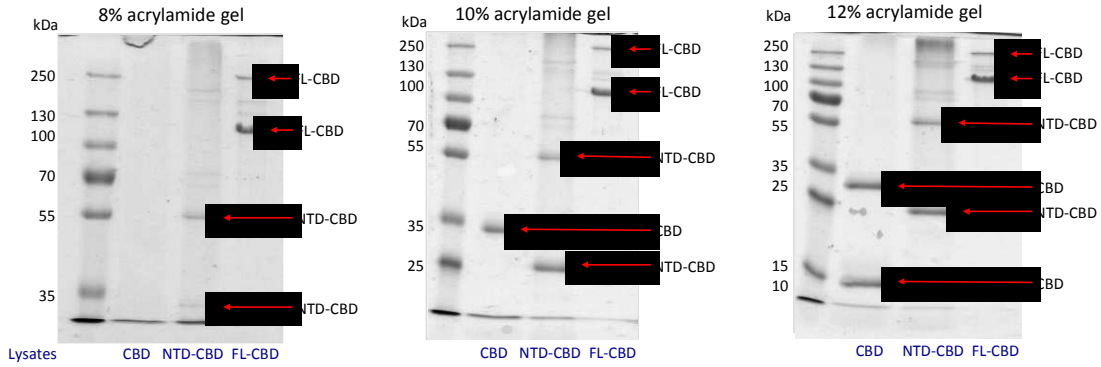


Figure 5.15 Large-scale purification of CBD, RpaC NTD-CBD and FL RpaC-CBD proteins

SDS-PAGE gels of the CBD-tagged proteins expressed in 1L cultures of *H.volcanii* cells and purified on cellulose beads as described in Materials and Methods chapter. Samples were separated on acrylamide gels (percentage indicated above the images) and bands were visualized by Coomasie Blue staining. Bands indicated by red arrows were identified based on the size and confirmed by mass spectrometry.

5.6 Discussion

In this chapter the structure-function analysis of RpaC protein has been presented. RpaC is the only SSB that is individually essential for cell viability in *H.volcanii*. We were unable to delete the *rpaC* gene from its native chromosomal locus (Figure 4.4) and down-regulation of *rpaC* expression using the *tna* promoter caused conditional growth arrest (Figure 5.1).

As an essential protein, RpaC is a good candidate to provide SSB function required during DNA replication. Although the indispensable role of SSBs in DNA replication is beyond doubt, to date no analysis determining which of the multiple euryarchaeal RPA is a replicative protein is available. It was shown using p*tna-rpaC* and p*tna*M3-*rpaC* integrant strains that lack of RpaC expression has a dramatic effect on on-going DNA synthesis *in vivo*; in cells labelled with [methyl-^{3}H] thymidine and grown on medium lacking tryptophan (*tna* promoter shut-off) the incorporation of radio-nucleotides was about 3-5 times lower than in cells grown on medium supplemented with tryptophan or in the wild-type (Figure 5.2). It is proof that RpaC is required for DNA synthesis *in vivo*.

RpaC might not be the only SSB involved in DNA replication. Indeed, elevated expression of the *rpaB1-rpaB2* operon, driven by the constitutive *fdx* promoter, is sufficient to partially suppress the growth defect of a p*tna-rpaC* strain (Figure 5.4), indicating that the products of *rpaB* operon are able to perform some of the essential cellular functions of RpaC. To further explore the role of RpaC in DNA replication

we attempted to construct p*tna-rpaC* Δ*rpaA* and p*tna-rpaC* Δ*rpaB* strains but this was unsuccessful. In addition, it was not possible to express *rpaC* from *tna* in Δ*rpaA* and Δ*rpaB* background by plasmid integration, or to delete *rpaA* and *rpaB* in p*tna-rpaC* (data not shown).

Haloarchaeal RpaC proteins have a conserved domain organisation: they possess three OB fold motifs (OB^{A-C}) and also N- and C- terminal domains. A cross-species complementation assay was used to investigate what parts of the protein are essential for it function. We constructed nine mutated versions (single, double and triple OB fold deletions and ΔNTD and ΔCTD) of *H.borinquense* RpaC, a protein closely related to *H.volcanii* RpaC, and tested their ability to rescue the p*tna-rpaC* strain on medium lacking tryptophan (Figure 5.11). Deletion of a single OB fold reduced but did not abolish growth, implying that the presence of all three OB folds provides optimal RpaC function but is not essential. Interestingly, some growth was seen in a double ΔOB^{AB} deletion, but not ΔOB^{AC} nor ΔOB^{BC}, indicating that RpaC protein containg an OB fold C only retains residual activity. Our results are in agreement with similar biochemical work done in *Methanosarcina acetivorans* and *Methanopyrus kandleri*, where construction of an single OB fold-deletion proteins as well as artificial RPAs consisting of chimeric OB fold modules (for example MacRPA3/Rpa1-OB chimera made by replacing the two OB folds of MacRPA3 with the two highly similar OB folds from MacRPA1) resulted in functional proteins. Deletion of a single OB fold in the four-OB fold-containing RpaC homologous did not influence the size of the ssDNA- binding site but binding affinity and ability to discriminate between ssDNA and dsDNA were slightly reduced (Lin and Ha, 2006, Lin et al., 2008, Robbins et al., 2005). It remains to be tested how deletions in *H.borinquense* RpaC affect DNA binding and whether there is a direct correlation between *in vivo* function and affinity for ssDNA.

As the further outcome of the cross-species complementation experiment it was found that deleting RpaC C-terminal domain does not have deleterious effect on cell growth, whereas cells expressing Hbo RpaC-ΔNTD grow more poorly than cells expressing the full-length protein. The RpaC CTD displayes a low sequence similarity to MacRPA1 OB fold D. MacRPA1, the *M.acetivorans* homologue of *H.volcanii* RpaC, possess four OB folds (OB folds A-D); in haloarchaea RpaC OB folds A-C are equivalent to folds A-C in the MacRPA1. The sequence similarity between haloarchaeal RpaC CTD and MacRPA1 OB fold D suggests that haloarchaea protein contained four OB folds at some point of evolution but OB fold

D has diverged and is no longer recognizable as an OB fold. Although cells expressing Hbo RpaC-ΔCTD in p*tna-rpaC* grew as well as the *H.volcanii* wild-type, this strain was significantly sensitive to DNA damage caused by UV irradiation and MMS but not phleomycin (Figure 5.12). This observation suggests that RpaC CTD is involved in DNA repair but not in the repair of dsDNA breaks.

In contrast to CTD, lack of the RpaC N-terminus impaired cell growth: cells expressing Hbo RpaC-ΔNTD did rescue growth of p*tna-rpaC* strain on medium lacking tryptophan but the level of rescue was much lower in comparison to cells expressing the full-length protein. The N-terminal domain is well conserved in all haloarchaeal species whose genomes are publically available (15 residues are identical in all species) and also in some methanogens. The RpaC NTD does not display sequence similarity to the OB fold and therefore is not likely to have any role in ssDNA binding. Deletion of the NTD impairs RpaC function: cells expressing Hbo RpaC-ΔNTD are significantly more sensitive to different kinds of DNA damage (Figure 5.12). It was concluded that this part of RpaC might provide a protein-protein interaction in DNA repair pathways. To specify the individual interaction partners, as well as the repair pathways that RpaC participates in, we attempted to express NTD (alongside the full-length RpaC) as an N-terminal fusion with cellulose binding domain (CBD) in *H.volcanii*. The CBD of the *Clostridium thermocellum* cellulosome has ability to interact with cellulose in a salt-insensitive manner and allow affinity-based purification in high salt concentration (Irihimovitch and Eichler, 2003, Morag et al., 1995). Although the full-length RpaC and RpaC NTD were strongly expressed in *H.volcanii*, we were unable to identify any interacting proteins (Figure 5.15).

In contrast to the DNA damage sensitivity observed when N- and C-terminal domains were deleted, overexpressing the full-length RpaC (either from the constitutive *H.salinarum fdx* promoter or from *tna* promoter in cells grown in the presence of 0.075 mM tryptophan) resulted in an increased resistance to several DNA-damaging factors causing different types of DNA damage (UV irradiation and 4NQO cause bulky adducts such as cyclobutane-pyrimidine dimers (CPDs) and 6-4 photoproducts (6-4PPs), MMS methylates DNA which can stall replication forks and phleomycin causes double strand breaks). One explanation for increased resistance to DNA damage would be reduction in growth rate, as has been shown for *D.radiodurans* (Cox and Battista, 2005). To rule out that possibility for the p*tna-rpaC* and p*fdx-rpaC* strains, we verified the generation times of these strains in

comparison to each other and the *H.volcanii* DS70 wild-type strain. Cells grown under standard conditions showed no measurable difference in growth rate (Figure 5.5).

To summarise, the results presented in this chapter offer the first insights into the functional and structural properties of the haloarchaeal RpaC protein. We showed that RpaC is the only individually essential SSB protein in *H.volcanii* and its presence is required for DNA replication *in vivo*. Also, we demonstrated that RpaC participates, most likely via the N-terminal domain, in several DNA repair pathways. Future work is needed to understand the exact roles of multiple RPAs in euryarchaeal DNA metabolism and to identify the proteins RPAs interact with in order to perform their cellular functions.

Chapter 6

Characterisation of the PriS-GinS operon in Archaea

6.1 Introduction

Primase and GINS are components of chromosomal DNA replication in Eukaryotes and Archaea. Primases are DNA-dependent RNA polymerases that synthesise short ribonucleotide primers for use by the replicative DNA polymerases (Arezi and Kuchta, 2000, Frick and Richardson, 2001). Eukaryotic and archaeal cells encode a heterodimeric primase composed of the catalytic and non-catalytic subunit (in Archaea, PriS and PriL, respectively). The crystal structures of the catalytic subunit of *P.furiosus* (Augustin et al., 2001), *P.horikoshii* (Ito et al., 2003) and *S.solfataricus* primases revealed presence of two distinct domains: the Prim domain composed of mixed α/β domain which includes the catalytic site of the enzyme and a smaller α-helical domain of unknown function. The exact role of the non-catalytic primase subunit, PriL, is less understood; data suggests that this protein might have a regulatory function, determining the length of the RNA primers synthesised by PriS (Le Breton et al., 2007).

The GINS protein is the archaeo-eukaryotic factor essential for the initiation and elongation stages of chromosomal replication that in eukaryotic cells is assembled together with the Cdc45 and MCM proteins into the CMG complex, unwinding DNA duplex ahead of the moving replication fork (reviewed by (MacNeill, 2010)).

Eukaryotic GINS is a heterotetramer composed of four distinct subunits designated Sld5, Psf1, Psf2 and Psf3. Each of the subunit comprises two structural domains: an α-helical A domain and a smaller B domain composed mainly of β-strands. In the archaeal domain of life, all species with sequenced genomes possess a single GINS protein (termed the GINS51), which resembles eukaryotic Sld5 and Psf1 subunits with the A-domain at the N-terminus and the B-domain at the C-terminus. In addition to GINS51, some archaeal species encode also GINS23, a protein with domain organisation described for eukaryotic Psf2 and Psf3 subunits: the B-domain is N-terminal and the A-domain is C-terminal (Marinsek et al., 2006, Makarova et al., 2005). In *S.solfataricus* and *P.furiosus*, the GINS51 and GINS23 form a tetrameric complex of two GINS51 subunits and two GINS23 subunits that is similar in structure to the eukaryotic counterpart (Marinsek et al., 2006, Yoshimochi et al.,

2008a). In many archaeal species belonging to the major phylogenetic groups genes encoding PriS and GINS51 are organised in one transcription unit.

This chapter will describe characterisation of *H.volcanii* PriS-GINS operon using the tryptophan- regulated *tna* promoter system in combination with cross-species complementation assay and bioinformatics study that lead us to identify a conserved structural domain shared by the archaeal PriS and GINS.

6.2 PriS-GINS operon in Archaea

Analysing the genome context of genes encoding small subunit of primase, GINS 51 and GINS 23 in species representative of the individual orders of three major archaeal species belonging to different phylogenetic groups revealed that that the ORFs encoding PriS and GINS51 are often adjacent to each other on the chromosome: in 27 of 43 analysed species the gene encoding PriS is located immediately upstream of that encoding GINS51 and in six species gene encoding GINS51 is adjacent to that encoding PCNA. In the remaining ten species, the genes are not co-localised. In addition, in 18 archaeal species encoding GINS23 protein, in 11 species the corresponding ORF is located adjacent to that encoding MCM helicase.

In *H.volcanii*, the PriS (HVO_2697) and GINS (HVO_2698) ORFs overlap by four nucleotides. Analysing the nucleotide sequence of PriS ORF immediately upstream of GINS start codon did not identify putative promoter region indicating that both genes are likely to be transcribed as a bicistronic mRNA. Whereas PriS was shown to be essential protein in *Halobacterium* sp. NRC-1 (Berquist et al., 2007), to our knowledge, no attempt has been made to elucidate essentiality of GINS protein in Haloarchaea.

To facilitate analysis of the essentiality of particular components of PriS-GINS operon in *H.volcanii* and future functional analysis of corresponding proteins we constructed *priS-ginS* conditional-lethal mutant using tryptophan-regulated promoter system. Initially, we generated plasmid-integrant strain in with the *tna* promoter was displacing the native promoter of *priS-ginS* (SMH731, Table 2.1). Little growth was seen when p*tna-priS-ginS* strain was grown on medium lacking tryptophan (promoter turned off), both in spotting assays and on growth curves, suggesting that at least one gene encodes a protein essential for *H.volcanii* cell viability (Figure 6.1A).

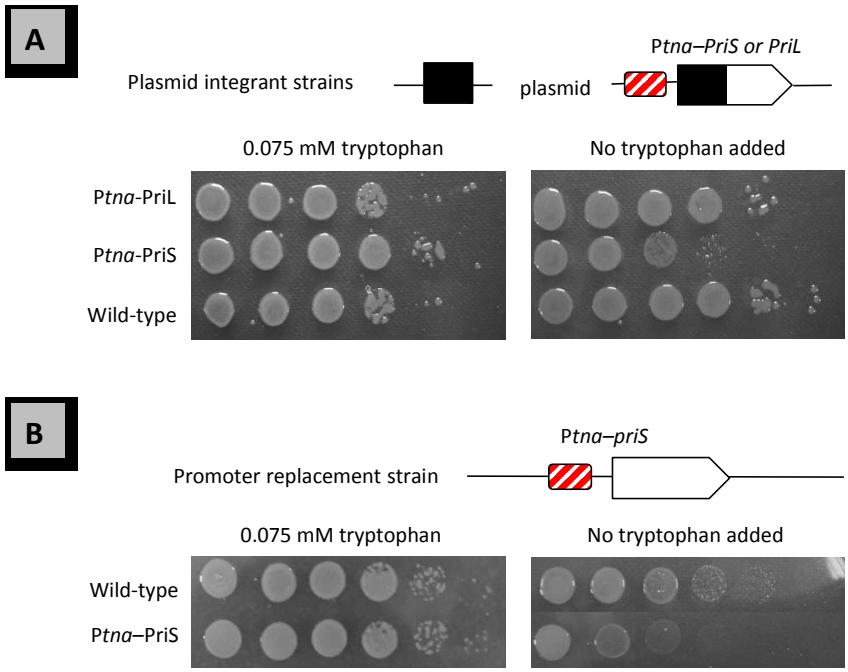


Figure 6.1 Growth of the p*tna-priS-ginS* conditional lethal mutants

A *H.volcanii* H98 strain carrying integrated pNPM-tna-PriS plasmids (PL19-PL21 Table 2.3). Strain DS70 is a wild-type control.
B The *H.volcanii* p*tna-pris-ginS* promoter replacement strain. Strain H98 is a wild-type control.
In **A** and **B** cells were grown up to mid-exponential phase at 45^{o}C in HvMin medium containing 0.075 mM tryptophan. Cells were then washed twice to remove tryptophan and 10-fold serial dilutions were spotted onto HvMin plates with or without tryptophan. Plates were incubated at 45^{o}C for 3 days.

Next, the genetically stable p*tna-pris-ginS* strain was engineered (SMH739, Table 2.1) with the p*tna* replacing the native promoter. Growth arrest of that strain in the absence of tryptophan was confirmed by spotting assay (Figure 6.1B).

6.3 Separation of the PriS-GINS operon in *H.volcanii*

The pop-in/pop-out method commonly used for gene knockdown in *H.volcanii* is based on construction deletion plasmid carrying 5' and 3'- flanking regions of the gene to be deleted allowing plasmid integration at correct locus on chromosome. In the case of genes located in the close proximity to the adjacent ORF on the chromosome, this deletion strategy is complicated by the possibility of disturbing expression of the adjacent genes, which hampers analysis of the results. In order to test whether *priS* and *ginS* encode essential proteins in *H.volcanii*, a complementation assay was chosen instead of the pop-in/pop-out method taking advantage of already constructed conditional lethal promoter replacement mutant *tna-priS-ginS* (SMH739, Table 2.1).

The *tna-priS-ginS* strain, which showed significant growth arrest on medium lacking tryptophan, was transformed with replicating plasmids expressing either both *H.volcanii* PriS and GINS proteins (plasmid PL93, Table 2.3) or only GINS (plasmid PL94, Table 2.3) and ability to grow of transformant cells was re-tested on medium containing and lacking tryptophan. The PL93 and PL94 are pTA230 derivatives carrying appropriate ORFs under the control of the *fdx* promoter. To construct plasmid PL93 carrying 2154-nucleotides fragment encoding PriS and GINS (with additional twenty four nucleotides downstream of the GINS locus to maintain the putative transcriptional terminator) the internal *Dde*I restriction site located at position 1174-1178 with the respect to the *priS* start codon was used. Appropriate ORFs were amplified by PCR using *H.volcanii* DS70 genomic DNA as a template and oligonucleotides P134-P140 (Appendix, Table A1). As shown on Figure 6.2, *in trans* expression of PriS and GINS proteins rescued growth of the *tna-priS-ginS* strain, whereas no rescue was seen when cells were transformed either with empty plasmid or with plasmid expressing *Haloferax* GINS protein only. These results indicate that the small primase subunit is indispensable for *H.volcanii* cell viability and is agreement with deletion study performed in *Halobacterium* sp. NRC-1 (Berquist et al., 2007).

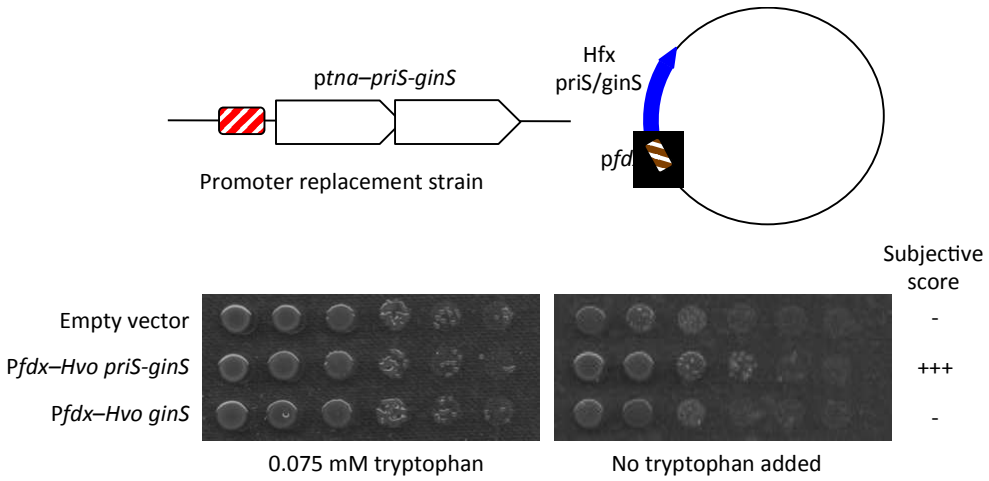


Figure 6.2 Rescue of p*tna-priS-ginS* strain by *in trans* expression *H.volcanii* PriS-GINS and GINS proteins.

H.volcanii p*tna-priS-ginS* strain carrying pTA230-fdx-Hvo-priS-ginS and pTA230-fdx-Hvo-ginS plasmids (PL93 and PL94, Table 2.3) was grown up to mid-exponential phase at 45^{o}C in HvMin medium containing 0.075 mM tryptophan. Cells were then washed twice to remove tryptophan and 10-fold serial dilutions were spotted onto HvMin plates with or without tryptophan. Plates were incubated at 45^{o}C for 3 days. *H.volcanii* p*tna-priS-ginS* strain carrying empty pTA230 plasmid is shown as a control.

In order to examine whether GINS itself is an essential protein in *H.volcanii* we used complementation assay where PriS only was expressed *in trans* from the *fdx* promoter in the p*tna-priS-ginS* and ability of that strain to grow on medium lacking tryptophan was re-tested. This was done with GINS orthologue from *Halogeometricum borinquense* (Hbo GINS) to avoid homologous recombination between the p*tna-priS-ginS* sequence present in the SMH739 strain and p*fdx*-Hfx-*ginS* construct on the plasmid. The genome context of *H.borinquense priS* and *ginS* ORFs corresponds to that seen in *H.volcanii*, with *ginS* (Hbor_11520) overlapping *priS* (Hbor_11510) by four nucleotides. *H.borinquense* proteins show the highest sequence identity to *H.volcanii* proteins at amino acid level (99% sequence identity for both PriS and GINS) with only limited similarity at the nucleotide sequence level. *H.borinquense ginS* ORF was amplified by PCR using the appropriate genomic DNA as a template and oligonucleotides P143-P144 (Appendix, Table A1), PCR product was digested with *Nde*I and *Hind*III enzymes and ligated into plasmid PL93, leaving *H.borinquense* GINS under the *fdx* promoter (plasmid PL97, Table 2.3). As a control, corresponding plasmids carrying the *H.borinquense priS* ORF and *priS-ginS* sequences were also constructed (plasmids PL96 and PL95, respectively). Resulting plasmids were then transformed into the p*tna-priS-ginS* strain and transformants were tested for their ability to restore growth of the cells on medium lacking tryptophan. As a result of this experiment we found that expression of *H.borinquense* PriS and GINS rescued growth of the p*tna-priS-ginS* strain and the level of rescue was indistinguishable from that seen when *H.volcanii* PriS and GINS was expressed under equivalent conditions. Interestingly, a similar observation was made for cells expressing *H.borinquense* PriS on its own: p*tna-priS-ginS* cells expressing *H.borinquense* PriS grew equally well on medium lacking tryptophan as cells expressing PriS-GINS proteins. Expression of *H.borinquense* GINS did not restore growth of cells on medium with no tryptophan supplementation (Figure 6.3).

A

	Hvo PriS	
	sequence similarity	sequence identity
Hbo	99%	72%

	Hvo GINS	
	sequence similarity	sequence identity
Hbo	99%	52%

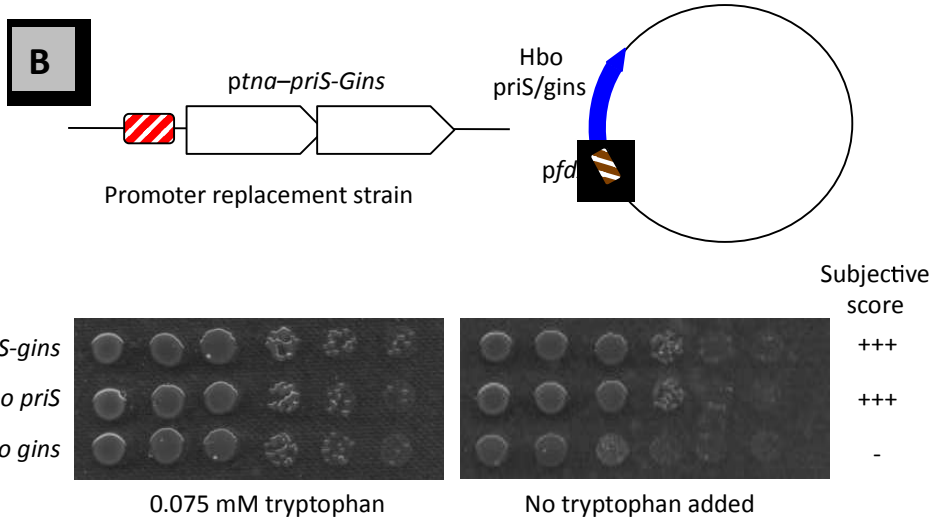


Figure 6.3 Rescue of p*tna-priS-ginS* strain by *H.borinquense* PriS and GINS orthologues

A Table summarizing sequence identity and similarity between *H.volcanii* and H.borinquense PriS (upper table) and *H.volcanii* and *H.borinquense* GINS (lower table) orthologues.
B *H.volcanii* p*tna-priS-ginS* strain carrying pTA230-fdx-Hbo-priS-ginS, pTA230-fdx-Hbo-priS and pTA230-fdx-Hbo-ginS plasmids (PL95-97, Table 2.3) was grown up to mid-exponential phase at 45°C in HvMin medium containing 0.075 mM tryptophan. Cells were then washed twice to remove tryptophan and 10-fold serial dilutions were spotted onto HvMin plates with or without tryptophan. Plates were incubated at 45°C for 3 days.

6.4 Bioinformatic analysis of PriS CTD

All archaeal species with sequenced genomes encode GINS51 proteins, whereas many species appear to lack of GINS23. In order to identify putative GINS23 proteins in diverse archaeal species, we performed BLAST searching using default parameters (Altschul et al., 1990) against archaeal proteins in the NCBI Reference Sequence database using the sequence of the *Cenarchaeum symbiosum* (strain A) GINS23 protein (CENSYa_1724; GI: 118576897) as the query. We observed that sequences corresponding to the C-terminal domain of the catalytic subunit of archaeal primase, termed PriS-CTD, were often found as a result of that search. For

example, the PriS protein (PAE3036; GI: 1463797) from *Pyrobaculum aerophilum* was identified with an E-value 0.003 (amino acids 14-63 of the *C.symbiosum* B-domain of the GinS23 are 42% identical to residues 261-310 of *P.aerophilum* PriS). In additional *Pyrobaculum* PriS proteins, from *P.calidifontis* (Pcal_0991, GI: 4909914), *P.arsenaticum* (Pars_1787, GI:5055591) and *P. islandicum* [Pisl_0437, GI: 4617745] were found with E-values of 0.058, 0.063 and 0.22, while PriS from *Thermoproteus neutrophilus* (Tneu 1683; GI:6165219) was found with an E-value of 5.6.

The PriS-CTD is a fifty amino acids domain at the C-terminus of archaeal PriS, which is comprised of a three-stranded antiparallel β-sheet adjacent to an α-helix and a two-stranded antiparallel β-sheet. Multiple sequence alignment performed for PriS sequences from fifty one archaeal species representative to all major archaeal lineages indicated that PriS-CTD is conserved in all phylogenetic groups with the exceptions of the *Thermococcales* (including *Pyrococcus* and *Thermococcus* species) and the *Methanobacteriales* **(***Methanosphaera* and *Methanothermobacter* species). Also, this domain is not present in the eukaryotic primase small subunit. The exact role of PriS-CTD is unknown but it has been suggested that it may play a structural role supporting and positioning the zinc-binding motif located at the end of extended β hairpin structure (Lao-Sirieix et al., 2005). Consistently with that hypothesis, in the *P.horikoshii* PriS, which does not possess the β hairpin structure, C-terminal domain is replaced by a single α-helix (Ito et al., 2003).

As the structural relationship between PriS and GINS proteins had not been reported previously this initial observation prompted us to explore it further. A multiple sequence alignment of PriS CTD, B-domain of GINS51 and B-domain of GINS23 from a representative set of archaeal species was performed. As a result, low-level sequence conservation across the entire CTD and B-domain regions was revealed (Figure 6.4).

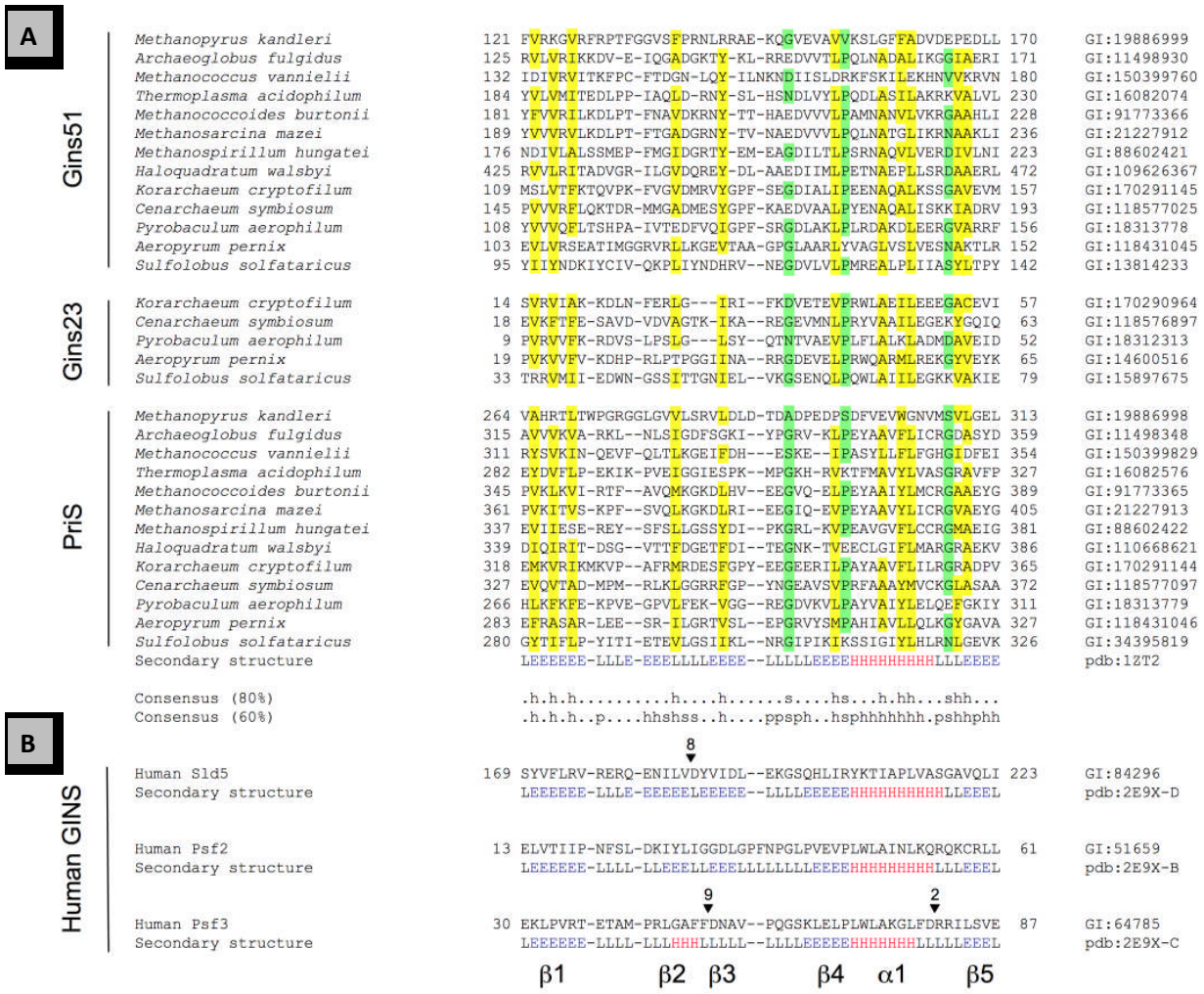


Figure 6.4 Multiple sequence alignment of archaeal primase CTD, archaeal GINS B-domain and eukaryotic GINS B-domain

A The multiple sequence alignment of PriS-CTD and GINS51 and GINS23 B-domains of selected archaeal species representative to all major lineages generated using Clustal X 2.0 with default paramiters (Jeanmougin et al., 1998, Larkin et al., 2007). Sequences are marked with the species name on the left and Genbank Identifiers GI on the right. The colouring is based on the sequence consensus shown underneath the alignment. Hydrophobic positions (ACFILMVWYH) are indicated by the letter h and shaded yellow when present in 80% of the sequences shown; small residues (ACDGNPSTV) are indicated by the letters and shaded green. The secondary structure of the CTD of the *S. solfataricus* PriS protein (PDB code 1TZ2) is shown underneath the alignment (with H, E and L indicating α-helix, β-strand and loop regions respectively, with α-helices shown in red and β-strands in blue), as are the primary sequences and secondary structures of the B-domains of three of the four human GINS proteins: Sld5, Psf2 and Psf3 (derived from PDB file 2E9X).

B The alignment of the human GINS and *S. solfataricus* PriS CTD sequences was generated by pairwise structure comparison (1ZT2 versus 2E9X with default parameters) using DaliLite (Holm and Park, 2000). The inverted triangles above the Sld5 and Psf3 sequences indicate that amino acids have been omitted at these positions; the number of amino acids omitted is shown.

Next, we compared the three-dimensional structures of *S.solfataricus* primase (PDB 1ZT2) (Lao-Sirieix et al., 2005) with the human GINS structure (PDB 2E9X) (Kamada et al., 2007, Chang et al., 2007a, Choi et al., 2007) using DaliLite (Holm and Park, 2000) (Figure 6.5). Structural similarities between the PriS CTD and B-domains of Sld5, Psf2 and Psf3 were identified, with Z-scores of 5.6, 5.5 and 3.4 and rmsd values of 2.6, 2.6 and 1.9 Å for 50, 48 and 38 Cα atoms, respectively. The structural similarity is most apparent with the Psf2 B-domain (Figure 6.5A and B).

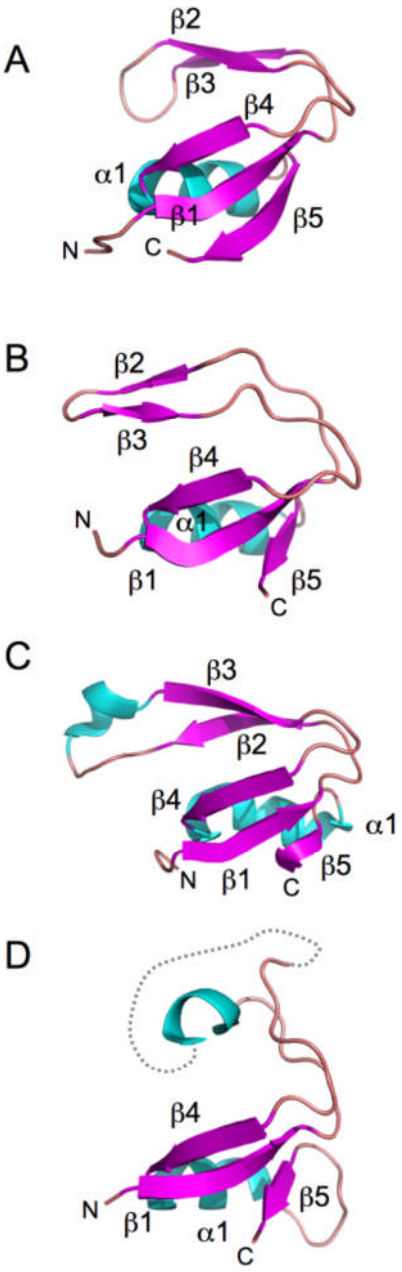


Figure 6.5 Comparison of the three-dimensional structures of *S.solfataricus* PriS-CTD with the B-domain of human GINS subunits

A Structure of the C-terminal domain (amino acids 274-329) of the *S.solfataricus* PriS protein (PDB code 1ZT2, chain A) showing five conserved β-strands β1-β5 and helix α1
B Structure of the N-terminal B-domain (amino acids 12-61) of the human GINS subunit Psf2 (PDB code 2E9X, chain B)
C Structure of the C-terminal B-domain (amino acids 167-223) of human Sld5 (PDB code
2E9X, chain D) indicating the presence of an additional α-helix between β-strands β2 and β3.
D Structure of the N-terminal B-domain (amino acids 30-87) of the human Psf3 (PDB code 2E9X, chain C) Amino acids 48-56 are missing from the structure (indicated by broken line).

The finding that PriS and GINS share a common domain together with the fact that the ORFs encoding PriS and GINS51 as well as GINS23 and MCM are often adjacent to each other on the chromosome, suggests a simple mechanism for the evolution of these proteins (Figure 6.6). In this model the last common archaeao-eukaryotic ancestor encoded both GINS51 and GINS23, in such chromosomal arrangement that GINS51 was adjacent to Prim and GINS23 was adjacent to MCM. A tandem duplication of GINS51 and following deletion of GINS51 A-domain sequence resulted in formation of PriS protein with acquired B-domain as a Prim-B-domain fusion. This model is supported by the fact that the members of deep-branching, ancient archaeal phylum,

Thaumarchaeota, possess genes encoding both GINS51 and GINS23 proteins in the genomic context presented above. The gene encoding GINS23 has been lost from many archaeal species at a latter stage of evolution and the CTD lost from the *Thermococcales*, including *Pyrococcus* and *Thermococcus* species and *Methanobacteriales* PriS. Because the eukaryotic small subunit of primase does not possess CTD sequence, proposed duplication and fusion events are most likely to take place after the archaeal and eukaryotic lineages had diverged.

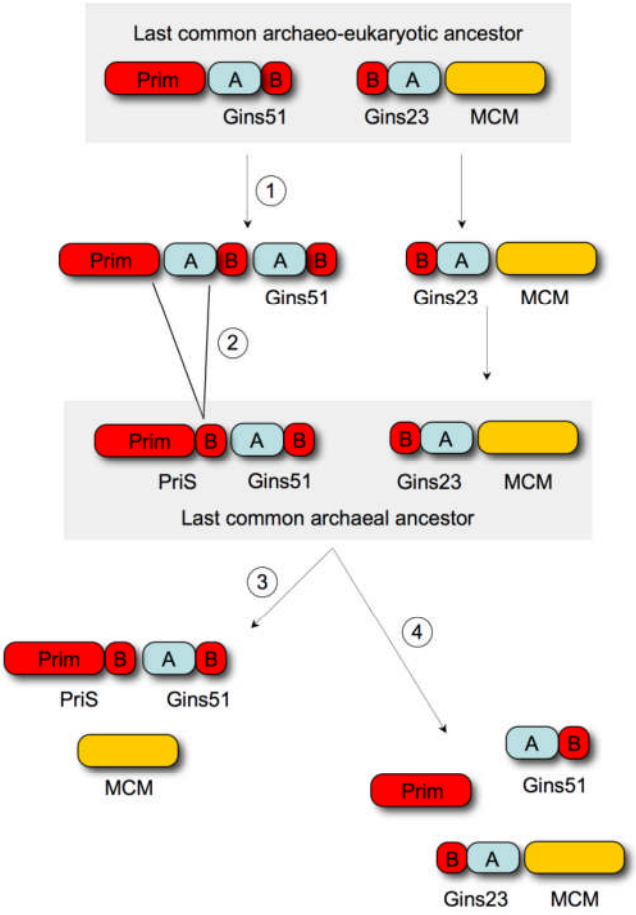


Figure 6.6 Model for acquisition of the CTD by PriS

Tandem duplication (labelled 1) of a GinS51 ORF found adjacent to a Prim domain ORF in the last common archaeo-eukaryotic ancestor is followed by deletion (labelled 2) of GinS51 A-domain sequences resulting in fusion of Prim domain and B-domain sequences. Subsequent archaeal evolution has seen loss of GinS23 (labelled 3) in many species and loss of the CTD (labelled 4) from PriS in the *Thermococcales* and *Methanobacteriales*.

6.5 Expression and purification of PriS CTD

The crystallographic and biochemical analysis of the primase small subunit from *Sulfolobus solfataricus* showed that the C-terminal domain might have a structural role supporting and positioning the zinc-binding motif located at the end of extended β hairpin structure (Lao-Sirieix et al., 2005).

The initial aim was to elucidate function of PriS-CTD in *H.volcanii* primase by expressing this domain in the native host and looking for proteins that CTD may interact with. The C-terminal domain in *H.volcanii* small primase subunit (protein

ID:292656816) was identified as a 48 amino acids fragment (residues 339-385), based on the multiple alignment of PriS sequences from forty one archaeal species (data not shown). For expression purposes, a plasmid carrying PriS CTD fragment as an N-terminal fusion with the CBD was constructed (PL92, Table 2.3). Expression was carried out according to the protocol described in Materials and Methods, which was successfully used for expression of *H.volcanii* RpaC protein (see Chapter 5.5). Although several attempts have been made, we were unable to detect expression of CBD-PriS CTD fusion (data not shown).

Our next attempt was to express *S.solfataricus* PriS-CTD (residues 227-330) in *E.coli* using the Glutathione S-transferase (GST) Gene Fusion System (GE Healthcare). For protein expression, plasmid pGEX-6P-1-CTD was constructed carrying sequence encoding *S,solfataricus* PriS-CTD fused to N-terminal GST tag. This plasmid, alongside with the empty pGEX-6P-1 were transformed into *E.coli* Rosetta 2 (DEX3) [pLysS] (Table 2.1.2). A small-scale protein expression, using 250 ml cultures, was carried out as described in Material and Methods (section 2.2.11). To determine the optimal regime for protein expression, different conditions were tested as listed in Table 6.1.

Table 6.1 Condition tested for expression of PriS CTD

Lysis buffer	150mM NaCl, 20mM sodium phosphate buffer 300mM NaCl, 20mM sodium phosphate buffer 500mM NaCl, 20mM sodium phosphate buffer
Detergent	Triton 1% Tween 0.1%
Temperature	20^{o}C, 25^{o}C, 37^{o}C
Time of induction	3h, overnight

Aliquots derived from the non-induced, induced, soluble and purification steps were examined on a 10% and 15% SDS-polyacrylamide gel followed by Coomassie blue staining. As showed on Figure 6.7, the PriS-CTD recombinant protein was highly expressed in *E.coli* after 3 hours induction at 25^{o}C with 0.1 mM of IPTG and could be easily purified using Glutatione Sepharose 4B beads. The GST-PriS CTD appeared on gel as a band of approximately size of 30 kDa, which corresponds to the predicted molecular mass for that protein fusion.

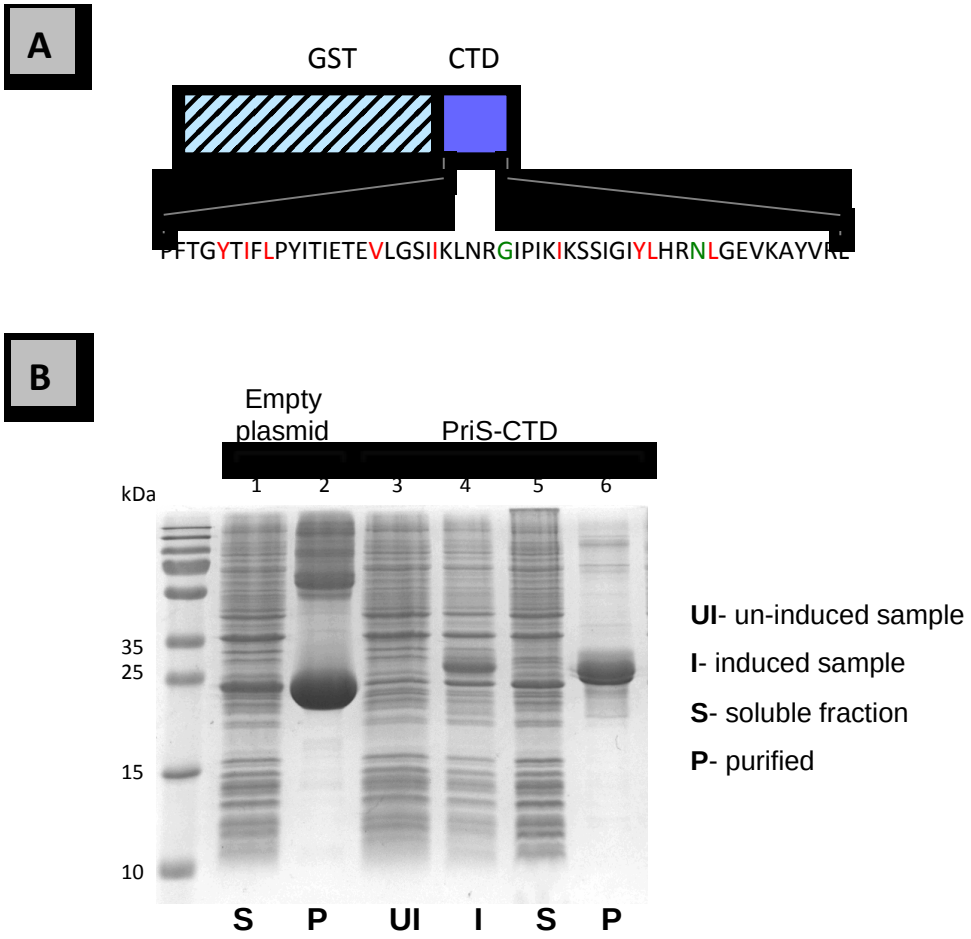


Figure 6.7 Expression of *S.solfataricus* PriS CTD in *E.coli*

A Schematic diagram of CBD-GST fusion protein expressed from pGEX-GP-1 plasmid (GE Healthcare). The striped area represents the GST tag and blue area represents PriS CTD. The amino acid sequence of CTD is shown. In accordance to the multiple sequence alignment of the archaeal PriS CTD and GINS B-domain on Figure 6.1, hydrophobic residues present in 80% of the sequences are marked on red and small conserved residues are marked on green.
B SDS-PAGE analysis of recombinant GST-PriS CTD. The extent of GST-PriS CTD fusion expression and purification was examined by loading 10 µl samples of un-induced cells (UI), induced cells (I), soluble fraction of induced whole cell lysate (S) and protein purified from the Glutatione Sepharose 4B beads. Lines 1 and 2 show samples derived from cells harboring empty plasmid pGEX-GP-1 and lines 3-6 from cells harboring plasmid pGEX-GP-1-PriS-CTD.

The small-scale expression presented above showed that *S.solfataricus* PriS-CTD could be expressed and purified in *E.coli* as a GST-tagged recombinant protein. That construct can be used to perform a pull-down experiment with *S.solfataricus* cell extract in order to identify proteins that interact with PriS via its C-terminal domain.

6.6 Discussion

Primase and GINS are essential components of eukaryotic DNA replication apparatus. Whereas primase has long history of structural and functional research, GINS is the most recently identified replication factor (Kubota et al., 2003, Kanemaki et al., 2003, Takayama et al., 2003) and therefore, its exact role is not fully established. Eukaryotic GINS is a tetramer composed of four subunit: Sld5, Psf1, Psf2 and Psf3 that are related to one other and arranged in a vertical pseudo two-fold axis in the middle of the complex (Chang et al., 2007a, Choi et al., 2007, Kamada et al., 2007). The GINS tetramer is a part of bigger protein assembly known as the CMG (Cdc45-Mcm-GINS) complex that functions as the replicative helicase (Gambus et al., 2006). The eukaryotic type of primase and GINS are also present in the third domain of life, Archaea. Bioinformatics searches revealed the presence of two types of GINS proteins in Archaea: all species appear to encode GINS51, which resembles eukaryotic Sld5 and Psf1 in terms of domain organisation. In addition, some, but not all, archaeal species encode also GINS23, with domain organisation similar to that seen in Psf2 and Psf3 subunits. It is not clear if GINS23 is truly absent in some species or rather its primary sequence is so divergent that is not longer recognised as GINS by database searching. In organisms encoding both types of proteins, two GINS51 and GINS23 molecules assemble in a tetrameric complex (Marinsek et al., 2006, Yoshimochi et al., 2008b).

The exact role that GINS plays in DNA synthesis in archaeal cells is also not fully resolved yet. The biochemical study performed for *S.solfataricus* and *P.furiosus* proteins place the archaeal GINS at the heart of replication apparatus, making contact with Orc1/Cdc6 at origins and with primase and MCM at the moving replication fork (Marinsek et al., 2006, Yoshimochi et al., 2008b), which suggest analogy with eukaryotic system. However, the presence of Cdc45 homologue was not confirmed in archaea. The RecJdbd, homologue of DNA-binding domain of bacterial RecJ nuclease, is a candidate protein to play a Cdc45 role in archaeal cells (Makarova et al., 2012). It is tempting to speculate that Archaea form a RecJ-MCM-GINS complex that is the functional equivalent of eukaryotic CDC45-MCM-GINS complex but this hypothesis has to be verified.

As the function of GINS is so poorly understood in Archaea, we decided to look closer at the gene encoding this factor in Haloarchaea. *H.volcanii* possesses GINS51 only and this ORF (HVO_2698) is adjacent to the gene encoding PriS (HVO_2697) (in fact, ORFs overlap by four nucleotides). Organizing this two genes in a putative

operon is commonly seen in Archaea: in more than the half species (27 out of 43) analysed in terms of the *priS* and *ginS* genome context, the two genes were adjacent to each other on the chromosome. To examine whether GINS is an essential protein in *H.volcanii*, we employed the cross-species complementation assay to take advantage of the genetically-stable p*tna-priS-ginS* promoter replacement strain (SMH739), whose construction was described in Chapter 3. The p*tna-priS-ginS* strain showed poor growth on medium lacking tryptophan (promoter turned-off), whereas growth of medium supplemented with 0.075 mM tryptophan was as good as the wild-type indicating that the presence of at least one protein in PriS-GINS cluster is indispensable for cell viability. When the SMH739 strain was transformed with replicating plasmid expressing either *H.volcanii* PriS and GINS proteins or only GINS we found that the presence of the GINS on its own was not sufficient to restore the wild-type phenotype. Next, we transformed the SMH739 strain with plasmids expressing PriS and GINS, only PriS or only GINS from the closely related species *H.borinquense*. Surprisingly, expression of PriS only rescued growth of the p*tna-priS-ginS* strain on medium without tryptophan and the growth properties of those cells were indistinguishable from the cells expressing both PriS and GINS proteins or from the wild-type. These results suggest that the lack of GINS51 has no effect on cell viability and define GINS as a non-essential protein in *H.volcanii*. If this is true, can it be extended for other Archaea? Is the RecJ-MCM-GINS complex formed in Haloarchaea? What is the function of GINS if it's not required for chromosomal DNA replication? The p*tna-priS-ginS* promoter replacement strain might be valuable tool in further studies aiming to answer these questions, particularly when combined with down-regulation or overexpression of other replicative proteins.

The second project described in this chapter was identification of similarity between the C-terminal domain of PriS (PriS-CTD) and the B-domain of GINS. A multiple sequence alignment performed for archaeal PriS-CTD, B-domain of GINS51 and B-domain of GINS23 from a representative set of organisms revealed sequence conservation (albeit low-level) across these regions. Also, comparison of the three-dimensional structures of *S.solfataricus* PriS-CTD with the B-domain of human Sld5, Psf2 and Psf3 GINS subunits identified similarities at the structure level. As the small subunit of eukaryotic primase lacks this putative domain, we proposed a model of co-evolution of *priS* and *ginS*, in which the C-terminal domain was acquired by PriS from GINS51, after the archaeal and eukaryotic lineages had diverged. To elucidate the possible role of PriS CTD as a protein-protein interaction region,

S.solfataricus recombinant CTD was expressed in *E.coli*. However, by the end of this PhD project, we were unable to identify any interacting partners.

Chapter 7
General discussion

The overarching aim of the project described in this thesis was to establish a system for conditional inactivation of genes involved in chromosomal DNA replication in the halophilic euryarchaeon *H.volcanii*, as a framework for further study of the structure and function of the corresponding proteins. *H.volcanii* was chosen as a model organism because this species is particularly suitable for genetic study (reviewed in (Leigh et al., 2011)).

The generation of conditional lethal mutants of replication genes was achieved by using the tryptophan-regulated promoter, p*tna*, as described in Chapter 3. The *tna* promoter, derived from the tryptophanase gene *trpA* is strongly induced in the presence of tryptophan, whereas in the absence of tryptophan its activity is very low (Large et al., 2007). The feasibility of using this system to study DNA replication was initially tested by construction of the p*tna* plasmid-integrant strains of fourteen genes that were known or predicted to participate in DNA synthesis; this showed that the wild-type p*tna* can be used for down-regulation of some, but not all, replication genes (three out of twelve tested genes). Growth of the cells expressing *polB*, *rpaC* and *priS* from p*tna* in medium lacking tryptophan was significantly inhibited due to a reduction of the transcript levels of the genes to below the levels seen in the wild-type, as confirmed by RT-PCR. Similar analysis performed for the *tna* plasmid-integrant strains that did not show reduced growth on medium without tryptophan (specifically, strains expressing *polD1*and *polD2* from p*tna*) indicated that the *tna* promoter is not completely inactivated in such conditions and provides a low level of expression that for some genes is sufficient to maintain the wild-type growth properties. The low mRNA levels seen in RT-PCR analysis of *tna-polD1* and *tna-polD2* constructs were indeed provided by the p*tna* activity as placing the transcriptional terminator L11e upstream of *tna* to prevent the read-through transcription did not change the phenotype of the analysed strains.

To extend usefulness of p*tna* to more replication genes, altered versions of the promoter were generated, to obtain promoters that were still regulated by tryptophan but displayed reduced minimal/maximal expression levels. This was successfully achieved by construction of a mutant form of *tna*, designated *tna*M3, carrying a

single substitution at position -26 (T to G), within a region identified as the *tna* TATA-box. The TATA-box, found in Eukaryotes and Archaea, is a region located approximately 25 bp upstream of the transcriptional start site that is important for initiation of transcription (Jun et al., 2011, Peng et al., 2011). It was previously reported that alterations generated in this region in Archaea decrease transcription efficiency from the affected promoter (Palmer and Daniels, 1995, Reiter et al., 1990). The *tna*M3 promoter was used to drive expression of *polB*, *mcm*, *rpaC* and *polD1* and, for all of these genes, overall expression levels were reduced by 20-40% in comparison to the wild-type *tna*. Construction of *tna*M3 allows conditional down-regulation of genes displaying the low native transcript levels, which do not show the desired phenotype when expressed from the wild-type *tna*. It should be noted, however, that use of *tna*M3 is still limited by the native transcript level of particular gene, so it probably cannot be used for every essential gene.

As mentioned above, our intention was to generate collection of conditional lethal mutants of replication genes that could be used to explore unknown aspects of DNA replication in *H.volcanii* and Archaea in general. We have mainly focused on two issues. The first one is a cellular function of the putative single-stranded DNA binding protein, RpaC: this was studied using the wild-type *tna* and *tna*M3- plasmid integrant and promoter replacement strains together with the strains overexpressing RpaC and the two other putative SSB factors, RpaA and RpaB. The second aspect that we focused on is the biological relevance of the PriS-GINS operon present in Haloarchaea and other archaeal lineages. This work aimed to address questions about the essentiality of PriS and, primarily, GINS, as a framework to establish the cellular functions of GINS as a putative component of the Cdc45/RecJ-MCM-GINS complex. This analysis was facilitated by construction of the *tna-priS-ginS* promoter replacement strain. In addition, we also used the wild-type *tna* and *tna*M3-*polB* plasmid integrant and promoter replacement strains to show that the DNA polymerase B is required for DNA synthesis *in vivo*.

Characterization of RpaC in *H.volcanii*

Chapters 4 and 5 describe identification and structure-function analysis of the RpaC protein. Euryarchaea display a great variety in the number and type of encoded RPA proteins and it is not clear how these multiple proteins function together in binding and stabilizing ssDNA. In fact, apart from biochemical study performed for RPA

homologous in Methanogens (Lin et al., 2008, Robbins et al., 2005, Robbins et al., 2004a), very little is known about euryarchaeal RPAs; no genetic study was performed for any RPA protein and no crystal structures are available.

Three putative RPA proteins in *H.volcanii* (RpaA, RpaB and RpaC) were identified and it was shown that RpaC, which contains three OB fold, is the only individually essential single-stranded DNA-binding protein in this archaeon.

Down-regulation of RpaC causes growth retardation and a significant reduction of DNA synthesis *in vivo*, as was shown by labelling the p*tna-rpaC* and p*tna*M3-*rpaC* cells with [methyl-^{3}H] thymidine in medium lacking tryptophan, which provides direct evidence that the presence of RpaC is required for DNA replication. In addition, RpaC seems to have general role in DNA repair as its overexpression makes cells resistant to different types of DNA damage. This function is probably mediated by the N-terminus as deletion of this region makes cells sensitive to DNA damaging agents. The N-terminal domain of RpaC (RpaC NTD) is well conserved across haloarchaea and methanogens. As the NTD does not display sequence similarity to OB fold, and therefore is less likely to have role in ssDNA binding, we concluded it might mediate protein-protein interactions. The RpaC NTD was expressed in the native host but we were unable to identify any interacting partners.

Deletion analysis was performed to investigate the roles of the individual OB folds in RpaC function. It was found that loss of one OB fold reduced but did not abolish cell growth, indicating that the presence of all three OB folds is not essential. When more than one OB fold was deleted, growth was strongly inhibited, especially when OB fold C was missing. That might imply a sequential mode of ssDNA binding: DNA might bind to OB fold C first, allowing subsequent interaction OB folds A and B and/or stabilizing DNA-protein binding. In recently published paper, recombinant *H.volcanii* RpaB1, which contains a single OB fold, was shown to be capable of binding 18 nucleotide ssDNA under physiological salt conditions *in vitro* (Winter et al., 2012). Similar studies performed for the full-length and OB fold-deleted RpaC proteins could be performed to test how alterations made in protein structure affect ssDNA binding.

The fact that RpaC is involved in DNA replication and repair raises the question of the cellular functions of the two other SSBs in Haloarchaea, RpaA and RpaB. We have shown that the *rpaA1* and *rpaB1* genes (together with downstream genes, *rpaA2* and *rpaB2*, respectively) are individually non-essential in *H.volcanii*, but that the Δ*rpaA*Δ*rpaB* double deletion is not viable. The *rpaB* deletion strain was

sensitive to UV treatment whereas the *rpaA* deletion did not cause any obvious phenotype in terms of growth properties or sensitivity to DNA damage. Stroud et al., 2012 confirmed our findings and extended the deletion analysis of the *rpaA* and *rpaB* operons, showing that the downstream genes cannot complement each other. They also demonstrated, using hexahistidine-tagged proteins expressed in *H.volcanii,* that RpaA1 and RpaB1 interact with their respective downstream proteins, RpaA2 and RpaB2 (Stroud et al., 2012). It will be very interesting to reveal whether RPAs form a complex with their respective downstream proteins in *vivo* and if they do, what is the role of these downstream genes as, based on the biochemical studies mentioned above (Winter et al., 2012), at least RpaB on its own is capable of ssDNA binding.

Another aspect of RPA's function that remains to be explored is the interplay between RpaA/RpaB and RpaC. We have seen that elevated level of RpaB operon, but not RpaA, can complement lack of RpaC.

Characterisation of PriS-GINS operon in Archaea

In many archaeal lineages, genes encoding the small subunit of primase (PriS) and the replication factor GINS are located adjacent to on another (sometimes overlapping) on the chromosome. This is the case in *H.volcanii* where *priS* and *ginS* overlap by four nucleotides, and both genes are likely to be transcribed from the same promoter. We used the p*tna-priS-ginS* promoter replacement strain to investigate whether both genes are essential for cell viability in *H.volcanii*. While PriS was shown to be indispensable protein in Halophiles (Berquist et al., 2007), no attempts were made to try to delete or down regulate the most recently identified replication factor, GINS, in Archaea.

Surprisingly, given that GINS is widely believed to be an essential factor, results from cross-species complementation experiments indicated that GINS could be down-regulated without severe consequences for the cells; *in trans* expression of PriS only from the episomal plasmid was sufficient to suppress the growth defect of p*tna-priS-ginS* strain. These findings are particular exciting considering a growing body of evidence for the existence of the Cdc45/RecJ-MCM-GINS complex in Archaea, which would be an equivalent of eukaryotic CMG (CDC45-MCM-GINS) complex (Krastanova et al., 2012, Makarova et al., 2012). It is easy to imagine that such a complex, if it exists, would have important role in DNA replication in Archaea and thus all its components should be indispensable for cell survival. It is

therefore crucial to confirm that GINS is indeed non-essential gene in Archaea and investigate the synergetic effects of down-regulation of RecJ, MCM and GINS.

PriS and GINS share a common domain

In addition to genetic analysis of PriS-GINS operon, Chapter 6 also described a bioinformatics study that lead us to identify low-level sequence similarity between the C-terminal domain of the catalytic subunit of archaeal primase (PriS CTD) and B-domain of GINS51 and B-domain of GINS23. This common domain is comprised of a three-stranded antiparallel β-sheets adjacent to an α-helix and a two-stranded antiparallel β-sheet. Comparison of the three-dimensional structures of *S.solfataricus* primase with the structure of subunits of human GINS tetramer revealed structural similarities between regions of interest with Z-scores of 5.6, 5.5 and 3.4 (when Sso PriS was compared with Psf2, Sld5 and Psf3, respectively). We hypothesized an evolutionary model explaining the acquisition of the CTD by PriS from GINS51 as a result of a tandem duplication of GINS51. This model assumes that the last common archaeao-eukaryotic ancestor encoded both GINS51 and GINS23, in such a chromosomal arrangement that GINS51 was adjacent to PriS and GINS23 was adjacent to MCM. This assumption is based on the corresponding genome context in a significant number of species from different archaeal lineages and the fact that the members of deep-branching, ancient archaeal phylum, Thaumarchaeota, possess genes encoding both GINS51 and GINS23. These findings offer important insights into evolution of archaeal GINS proteins.

References

ALBERS, S. V., BIRKELAND, N. K., DRIESSEN, A. J., GERTIG, S., HAFERKAMP, P., KLENK, H. P., KOURIL, T., MANICA, A., PHAM, T. K., RUOFF, P., SCHLEPER, C., SCHOMBURG, D., SHARKEY, K. J., SIEBERS, B., SIEROCINSKI, P., STEUER, R., VAN DER OOST, J., WESTERHOFF, H. V., WIELOCH, P., WRIGHT, P. C. & ZAPARTY, M. 2009. SulfoSYS (*Sulfolobus* Systems Biology): towards a silicon cell model for the central carbohydrate metabolism of the archaeon *Sulfolobus solfataricus* under temperature variation. *Biochem Soc Trans* 37**,** 58-64.

ALLERS, T., BARAK, S., LIDDELL, S., WARDELL, K. & MEVARECH, M. 2010. Improved Strains and Plasmid Vectors for Conditional Overexpression of His-Tagged Proteins in *Haloferax volcanii*. *Appl Environ Microbiol* 76**,** 1759-1769.

ALLERS, T. & MEVARECH, M. 2005. Archaeal genetics - the third way. *Nat Rev Genet* 6**,** 58-73.

ALLERS, T., NGO, H. P., MEVARECH, M. & LLOYD, R. G. 2004. Development of additional selectable markers for the halophilic Archaeon *Haloferax volcanii* based on the *leuB* and *trpA* genes. *Appl Environ Microbiol* 70**,** 943-953.

ALTSCHUL, S. F., GISH, W., MILLER, W., MYERS, E. W. & LIPMAN, D. J. 1990. Basic local alignment search tool. *J Mol Biol* 215**,** 403-10.

APARICIO, T., GUILLOU, E., COLOMA, J., MONTOYA, G. & MENDEZ, J. 2009. The human GINS complex associates with Cdc45 and MCM and is essential for DNA replication. *Nucleic Acids Res* 37**,** 2087-95.

ARAVIND, L. & KOONIN, E. V. 1998. Phosphoesterase domains associated with DNA polymerases of diverse origins. *Nucleic Acids Res* 26**,** 3746-52.

AREZI, B. & KUCHTA, R. D. 2000. Eukaryotic DNA primase. *Trends Biochem Sci* 25**,** 572-6.

ARMACHE, K. J., MITTERWEGER, S., MEINHART, A. & CRAMER, P. 2005. Structures of complete RNA polymerase II and its subcomplex, Rpb4/7. *J Biol Chem* 280**,** 7131-4.

AUGUSTIN, M. A., HUBER, R. & KAISER, J. T. 2001. Crystal structure of a DNA-dependent RNA polymerase (DNA primase). *Nat Struct Biol* 8**,** 57-61.

BAE, B., CHEN, Y. H., COSTA, A., ONESTI, S., BRUNZELLE, J. S., LIN, Y., CANN, I. K. & NAIR, S. K. 2009. Insights into the architecture of the replicative helicase from the structure of an archaeal MCM homolog. *Structure* 17**,** 211-22.

BAE, S. H., BAE, K. H., KIM, J. A. & SEO, Y. S. 2001. RPA governs endonuclease switching during processing of Okazaki fragments in eukaryotes. *Nature* 412**,** 456-461.

BAN, N., NISSEN, P., HANSEN, J., MOORE, P. B. & STEITZ, T. A. 2000. The complete atomic structure of the large ribosomal subunit at 2.4 A resolution. *Science* 289**,** 905-20.

BARRY, E. R. & BELL, S. D. 2006. DNA replication in the archaea. *Microbiol Mol Biol R* 70**,** 876.

BARRY, E. R., MCGEOCH, A. T., KELMAN, Z. & BELL, S. D. 2007. Archaeal MCM has separable processivity, substrate choice and helicase domains. *Nucleic Acids Res* 35**,** 988-98.

BELL, S. D. 2011. DNA replication: archaeal oriGINS. *BMC biology*, 9**,** 36.

BELL, S. D., CAIRNS, S. S., ROBSON, R. L. & JACKSON, S. P. 1999. Transcriptional regulation of an archaeal operon in vivo and in vitro. *Mol Cell* 4**,** 971-982.

BELL, S. D. & JACKSON, S. P. 1998. Transcription and translation in Archaea: A mosaic of eukaryal and bacterial features. *Trends Microbiol* 6**,** 222-228.
BELL, S. D. & JACKSON, S. P. 2001. Mechanism and regulation of transcription in archaea. *Curr Opin Microbiol* 4**,** 208-213.
BELL, S. D., MAGILL, C. P. & JACKSON, S. P. 2001. Basal and regulated transcription in Archaea. *Biochem Soc Trans* 29**,** 392-395.
BELL, S. P. & DUTTA, A. 2002. DNA replication in eukaryotic cells. *Annu Rev Biochem* 71**,** 333-74.
BENNETT-LOVSEY, R. M., HERBERT, A. D., STERNBERG, M. J. & KELLEY, L. A. 2008. Exploring the extremes of sequence/structure space with ensemble fold recognition in the program Phyre. *Proteins* 70**,** 611-25.
BERQUIST, B. R., DASSARMA, P. & DASSARMA, S. 2007. Essential and non-essential DNA replication genes in the model halophilic Archaeon, Halobacterium sp NRC-I. *Bmc Genetics* 8.
BITAN-BANIN, G., ORTENBERG, R. & MEVARECH, M. 2003. Development of a gene knockout system for the halophilic archaeon *Haloferax volcanii* by use of the pyrE gene. *J Bacteriol* 185**,** 772-778.
BLABY, I. K., PHILLIPS, G., BLABY-HAAS, C. E., GULIG, K. S., EL YACOUBI, B. & DE CRECY-LAGARD, V. 2010. Towards a systems approach in the genetic analysis of archaea: Accelerating mutant construction and phenotypic analysis in Haloferax volcanii. *Archaea* 2010**,** 426239.
BLASEIO, U. & PFEIFER, F. 1990. Transformation of *Halobacterium halobium* - Development of Vectors and Investigation of Gas Vesicle Synthesis. *Proc Natl Acad Sci USA* 87**,** 6772-6776.
BOCHKAREV, A. & BOCHKAREVA, E. 2004. From RPA to BRCA2: lessons from single-stranded DNA binding by the OB-fold. *Curr Opin Struct Biol* 14**,** 36-42.
BOCQUIER, A. A., LIU, L., CANN, I. K., KOMORI, K., KOHDA, D. & ISHINO, Y. 2001. Archaeal primase: bridging the gap between RNA and DNA polymerases. *Curr Biol* 11**,** 452-6.
BOLHUIS, H., PALM, P., WENDE, A., FALB, M., RAMPP, M., RODRIGUEZ-VALERA, F., PFEIFFER, F. & OESTERHELT, D. 2006. The genome of the square archaeon *Haloquadratum walsbyi* : life at the limits of water activity. *BMC genomics,* 7**,** 169.
BOSKOVIC, J., COLOMA, J., APARICIO, T., ZHOU, M., ROBINSON, C. V., MENDEZ, J. & MONTOYA, G. 2007. Molecular architecture of the human GINS complex. *Embo Rep* 8**,** 678-684.
BOUCHER, Y., DOUADY, C. J., PAPKE, R. T., WALSH, D. A., BOUDREAU, M. E., NESBO, C. L., CASE, R. J. & DOOLITTLE, W. F. 2003. Lateral gene transfer and the origins of prokaryotic groups. *Annual Rev Genet* 37**,** 283-328.
BREUERT, S., ALLERS, T., SPOHN, G. & SOPPA, J. 2006. Regulated polyploidy in halophilic archaea. *PLoS One* 1**,** e92.
BREWSTER, A. S. & CHEN, X. S. 2010. Insights into the MCM functional mechanism: lessons learned from the archaeal MCM complex. *Crit Rev Biochem Mol Biol* 45**,** 243-56.
BREWSTER, A. S., WANG, G., YU, X., GREENLEAF, W. B., CARAZO, J. M., TJAJADI, M., KLEIN, M. G. & CHEN, X. S. 2008. Crystal structure of a near-full-length archaeal MCM: functional insights for an AAA+ hexameric helicase. *Proc Natl Acad Sci USA* 105**,** 20191-6.
BROCHIER-ARMANET, C., FORTERRE, P. & GRIBALDO, S. 2011. Phylogeny and evolution of the Archaea: one hundred genomes later. *Curr Opin Microbiol* 14**,** 274-81.

BRODERICK, S., REHMET, K., CONCANNON, C. & NASHEUER, H. P. 2010. Eukaryotic single-stranded DNA binding proteins: central factors in genome stability. *Subcell Biochem* 50**,** 143-63.
BRUGGER, K., REDDER, P., SHE, Q., CONFALONIERI, F., ZIVANOVIC, Y. & GARRETT, R. A. 2002. Mobile elements in archaeal genomes. *FEMS Microbiol Lett* 206**,** 131-41.
BUBECK, D., REIJNS, M. A., GRAHAM, S. C., ASTELL, K. R., JONES, E. Y. & JACKSON, A. P. 2011. PCNA directs type 2 RNase H activity on DNA replication and repair substrates. *Nucleic Acids Res* 39**,** 3652-66.
CANN, I. K., ISHINO, S., YUASA, M., DAIYASU, H., TOH, H. & ISHINO, Y. 2001. Biochemical analysis of replication factor C from the hyperthermophilic archaeon Pyrococcus furiosus. *J Bacteriol* 183**,** 2614-23.
CAPES, M. D., COKER, J. A., GESSLER, R., GRINBLAT-HUSE, V., DASSARMA, S. L., JACOB, C. G., KIM, J. M. & DASSARRNA, P. 2011. The information transfer system of halophilic archaea. *Plasmid* 65**,** 77-101.
CASTREC, B., LAURENT, S., HENNEKE, G., FLAMENT, D. & RAFFIN, J. P. 2010. The Glycine-Rich Motif of *Pyrococcus abyssi* DNA Polymerase D Is Critical for Protein Stability. *J Mol Biol* 396**,** 840-848.
CASTREC, B., ROUILLON, C., HENNEKE, G., FLAMENT, D., QUERELLOU, J. & RAFFIN, J. P. 2009. Binding to PCNA in Euryarchaeal DNA Replication requires two PIP motifs for DNA polymerase D and one PIP motif for DNA polymerase B. *J Mol Biol* 394**,** 209-18.
CHAI, Q., QIU, J., CHAPADOS, B. R. & SHEN, B. 2001. Archaeoglobus fulgidus RNase HII in DNA replication: enzymological functions and activity regulation via metal cofactors. *Biochem Biophys Res Commun* 286**,** 1073-81.
CHANG, Y. P., WANG, G., BERMUDEZ, V., HURWITZ, J. & CHEN, X. S. 2007a. Crystal structure of the GINS complex and functional insights into its role in DNA replication. *Proc Natl Acad Sci USA* 104**,** 12685-90.
CHANG, Y. P., WANG, G. G., BERMUDEZ, V., HURWITZ, J. & CHEN, X. J. S. 2007b. Crystal structure of the GINS complex and functional insights into its role in DNA replication. *Proc Natl Acad Sci USA* 104**,** 12685-12690.
CHAPADOS, B. R., CHAI, Q., HOSFIELD, D. J., QIU, J., SHEN, B. & TAINER, J. A. 2001. Structural biochemistry of a type 2 RNase H: RNA primer recognition and removal during DNA replication. *J Mol Biol* 307**,** 541-56.
CHAPADOS, B. R., HOSFIELD, D. J., HAN, S., QIU, J., YELENT, B., SHEN, B. & TAINER, J. A. 2004. Structural basis for FEN-1 substrate specificity and PCNA-mediated activation in DNA replication and repair. *Cell* 116**,** 39-50.
CHARLEBOIS, R. L., LAM, W. L., CLINE, S. W. & DOOLITTLE, W. F. 1987. Characterization of Phv2 from *Halobacterium volcanii* and Its Use in Demonstrating Transformation of an Archaebacterium. *Proc Natl Acad Sci USA* 84**,** 8530-8534.
CHARLEBOIS, R. L., SCHALKWYK, L. C., HOFMAN, J. D. & DOOLITTLE, W. F. 1991. Detailed Physical Map and Set of Overlapping Clones Covering the Genome of the Archaebacterium *Haloferax volcanii* Ds2. *J Mol Bio*, 222**,** 509-524.
CHOI, J. M., LIM, H. S., KIM, J. J., SONG, O. K. & CHO, Y. 2007. Crystal structure of the human GINS complex. *Genes Dev* 21**,** 1316-21.
CONNOLLY, B. A., FOGG, M. J., SHUTTLEWORTH, G. & WILSON, B. T. 2003. Uracil recognition by archaeal family B DNA polymerases. *Biochem Soc Trans* 31**,** 699-702.
COSTA, A., PAPE, T., VAN HEEL, M., BRICK, P., PATWARDHAN, A. & ONESTI, S. 2006. Structural studies of the archaeal MCM complex in different functional states. *J Struct Biol* 156**,** 210-9.

COX, M. M. & BATTISTA, J. R. 2005. *Deinococcus radiodurans* - the consummate survivor. *Nat Rev Microbiol* 3**,** 882-92.
CRAMER, P., BUSHNELL, D. A. & KORNBERG, R. D. 2001. Structural basis of transcription: RNA polymerase II at 2.8 angstrom resolution. *Science,* 292**,** 1863-76.
CROWLEY, D. J., BOUBRIAK, I., BERQUIST, B. R., CLARK, M., RICHARD, E., SULLIVAN, L., DASSARMA, S. & MCCREADY, S. 2006. The *uvrA*, *uvrB* and *uvrC* genes are required for repair of ultraviolet light induced DNA photoproducts in *Halobacterium* sp. NRC-1. *Saline Syst* 2**,** 11.
CUBEDDU, L. & WHITE, M. F. 2005. DNA damage detection by an archaeal single-stranded DNA-binding protein. *J Mol Biol* 353**,** 507-16.
CULLMANN, G., FIEN, K., KOBAYASHI, R. & STILLMAN, B. 1995. Characterization of the five replication factor C genes of *Saccharomyces cerevisiae*. *Mol Cell Biol* 15**,** 4661-71.
DANIEL, R. M., PETERSON, M. E., DANSON, M. J., PRICE, N. C., KELLY, S. M., MONK, C. R., WEINBERG, C. S., OUDSHOORN, M. L. & LEE, C. K. 2010. The molecular basis of the effect of temperature on enzyme activity. *Biochem J* 425**,** 353-60.
DASSARMA, S., BERQUIST, B. R., COKER, J. A., DASSARMA, P. & MULLER, J. A. 2006. Post-genomics of the model haloarchaeon *Halobacterium* sp. NRC-1. *Saline Syst* 2**,** 3.
DASSARMA, S. L., CAPES, M. D., DASSARMA, P. & DASSARMA, S. 2010. HaloWeb: the haloarchaeal genomes database. *Saline Syst* 6**,** 12.
DE FALCO, M., FUSCO, A., DE FELICE, M., ROSSI, M. & PISANI, F. M. 2004. The DNA primase of Sulfolobus solfataricus is activated by substrates containing a thymine-rich bubble and has a 3'-terminal nucleotidyl-transferase activity. *Nucleic Acids Res* 32**,** 5223-30.
DE VRIEZE, J., HENNEBEL, T., BOON, N. & VERSTRAETE, W. 2012. Methanosarcina: The rediscovered methanogen for heavy duty biomethanation. *Bioresour Technol* 112**,** 1-9.
DEMOTT, M. S., ZIGMAN, S. & BAMBARA, R. A. 1998. Replication protein A stimulates long patch DNA base excision repair. *J Biol Chem* 273**,** 27492-8.
DESOGUS, G., ONESTI, S., BRICK, P., ROSSI, M. & PISANI, F. M. 1999. Identification and characterization of a DNA primase from the hyperthermophilic archaeon *Methanococcus jannaschii*. *Nucleic Acids Res* 27**,** 4444-50.
DEVEAUX, L. C., MULLER, J. A., SMITH, J., PETRISKO, J., WELLS, D. P. & DASSARMA, S. 2007. Extremely radiation-resistant mutants of a halophilic archaeon with increased single-stranded DNA-binding protein (RPA) gene expression. *Radiat Res* 168**,** 507-14.
DIONNE, I., NOOKALA, R. K., JACKSON, S. P., DOHERTY, A. J. & BELL, S. D. 2003. A heterotrimeric PCNA in the hyperthermophilic archaeon *Sulfolobus solfataricus*. *Mol Cell* 11**,** 275-82.
DOVE, S. L., DARST, S. A. & HOCHSCHILD, A. 2003. Region 4 of sigma as a target for transcription regulation. *Mol Microbiol* 48**,** 863-74.
DYALL-SMITH, M. 2009. The Halohandbook.
EGGLER, A. L., INMAN, R. B. & COX, M. M. 2002. The Rad51-dependent pairing of long DNA substrates is stabilized by replication protein A. *J Biol Chem* 277**,** 39280-8.
EGOROVA, K. & ANTRANIKIAN, G. 2005. Industrial relevance of thermophilic Archaea. *Curr Opin Microbiol* 8**,** 649-55.
EVGUENIEVA-HACKENBERG, E., WALTER, P., HOCHLEITNER, E., LOTTSPEICH, F. & KLUG, G. 2003. An exosome-like complex in *Sulfolobus solfataricus*. *Embo Rep* 4**,** 889-93.

FAIRMAN, M. P. & STILLMAN, B. 1988. Cellular factors required for multiple stages of SV40 DNA replication in vitro. *The EMBO J* 7**,** 1211-8.
FAN, J. & PAVLETICH, N. P. 2012. Structure and conformational change of a replication protein A heterotrimer bound to ssDNA. *Genes Dev* 26**,** 2337-47.
FANNING, E., KLIMOVICH, V. & NAGER, A. R. 2006. A dynamic model for replication protein A (RPA) function in DNA processing pathways. *Nucleic Acids Res* 34**,** 4126-37.
FERRY, J. G. 2010. How to make a living by exhaling methane. *Annu Rev Microbiol* 64**,** 453-73.
FINE, A., IRIHIMOVITCH, V., DAHAN, I., KONRAD, Z. & EICHLER, J. 2006. Cloning, expression, and purification of functional Sec11a and Sec11b, type I signal peptidases of the Archaeon *Haloferax volcanii*. *J Bacteriol* 188**,** 1911-1919.
FLETCHER, R. J., BISHOP, B. E., LEON, R. P., SCLAFANI, R. A., OGATA, C. M. & CHEN, X. S. 2003. The structure and function of MCM from archaeal *M. Thermoautotrophicum*. *Nat Struct Biol* 10**,** 160-7.
FLYNN, R. L. & ZOU, L. 2010. Oligonucleotide/oligosaccharide-binding fold proteins: a growing family of genome guardians. *Crit Rev Biochem Mol Biol* 45**,** 266-75.
FOGG, M. J., PEARL, L. H. & CONNOLLY, B. A. 2002. Structural basis for uracil recognition by archaeal family B DNA polymerases. *Nat Struct Biol* 9**,** 922-7.
FORSBURG, S. L. 2004. Eukaryotic MCM proteins: beyond replication initiation. *Microbiol Mol Biol R: MMBR* 68**,** 109-31.
FORTERRE, P., ELIE, C. & KOHIYAMA, M. 1984. Aphidicolin inhibits growth and DNA synthesis in halophilic arachaebacteria. *J Bacteriol* 159**,** 800-2.
FRICK, D. N. & RICHARDSON, C. C. 2001. DNA primases. *Annu Rev Biochem* 70**,** 39-80.
GAI, D., LI, D., FINKIELSTEIN, C. V., OTT, R. D., TANEJA, P., FANNING, E. & CHEN, X. S. 2004. Insights into the oligomeric states, conformational changes, and helicase activities of SV40 large tumor antigen. *J Biol Chem* 279**,** 38952-9.
GAMBUS, A., JONES, R. C., SANCHEZ-DIAZ, A., KANEMAKI, M., VAN DEURSEN, F., EDMONDSON, R. D. & LABIB, K. 2006. GINS maintains association of Cdc45 with MCM in replisome progression complexes at eukaryotic DNA replication forks. *Nat Cell Biol* 8**,** 358-66.
GEORGE, N. P., NGO, K. V., CHITTENI-PATTU, S., NORAIS, C. A., BATTISTA, J. R., COX, M. M. & KECK, J. L. 2012. Structure and Cellular Dynamics of Deinococcus radiodurans Single-stranded DNA (ssDNA)-binding Protein (SSB)-DNA Complexes. *J Biol Chem* 287**,** 22123-32.
GOMEZ-LLORENTE, Y., FLETCHER, R. J., CHEN, X. S., CARAZO, J. M. & SAN MARTIN, C. 2005. Polymorphism and double hexamer structure in the archaeal minichromosome maintenance (MCM) helicase from *Methanobacterium thermoautotrophicum*. *J Biol Chem* 280**,** 40909-15.
GRABOWSKI, B. & KELMAN, Z. 2003. Archaeal DNA replication: Eukaryal proteins in a bacterial context. *Annu Rev Microbiol* 57**,** 487-516.
GRAHAM, B. W., SCHAUER, G. D., LEUBA, S. H. & TRAKSELIS, M. A. 2011. Steric exclusion and wrapping of the excluded DNA strand occurs along discrete external binding paths during MCM helicase unwinding. *Nucleic Acids Res* 39**,** 6585-95.
GREGOR, D. & PFEIFER, F. 2005. In vivo analyses of constitutive and regulated promoters in halophilic archaea. *Microbiology* 151**,** 25-33.
GRISTWOOD, T., DUGGIN, I. G., WAGNER, M., ALBERS, S. V. & BELL, S. D. 2012. The sub-cellular localization of Sulfolobus DNA replication. *Nucleic Acids Res*.

GUEGUEN, Y., ROLLAND, J. L., LECOMPTE, O., AZAM, P., LE ROMANCER, G., FLAMENT, D., RAFFIN, J. P. & DIETRICH, J. 2001. Characterization of two DNA polymerases from the hyperthermophilic euryarchaeon *Pyrococcus abyssi*. *Eur J Biochem* 268**,** 5961-5969.
HAMPSEY, M. 1998. Molecular genetics of the RNA polymerase II general transcriptional machinery. *Microbiol Mol Biol R: MMBR* 62**,** 465-503.
HARING, S. J., HUMPHREYS, T. D. & WOLD, M. S. 2010. A naturally occurring human RPA subunit homolog does not support DNA replication or cell-cycle progression. *Nucleic Acids Res* 38**,** 846-58.
HARTMAN, A. L., NORAIS, C., BADGER, J. H., DELMAS, S., HALDENBY, S., MADUPU, R., ROBINSON, J., KHOURI, H., REN, Q. H., LOWE, T. M., MAUPIN-FURLOW, J., POHLSCHRODER, M., DANIELS, C., PFEIFFER, F., ALLERS, T. & EISEN, J. A. 2010. The Complete Genome Sequence of *Haloferax volcanii* DS2, a Model Archaeon. *PLoS One,* 5**,** -.
HARUKI, M., HAYASHI, K., KOCHI, T., MUROYA, A., KOGA, Y., MORIKAWA, M., IMANAKA, T. & KANAYA, S. 1998. Gene cloning and characterization of recombinant RNase HII from a hyperthermophilic archaeon. *J Bacteriol* 180**,** 6207-14.
HENNEKE, G., FLAMENT, D., HUBSCHER, U., QUERELLOU, J. & RAFFIN, J. P. 2005. The hyperthermophilic euryarchaeota *Pyrococcus abyss*i likely requires the two DNA polymerases D and B for DNA replication. *J Mol Biol* 350**,** 53-64.
HINGORANI, M. M. & O'DONNELL, M. 2000. A tale of toroids in DNA metabolism. *Nat Rev Mol Cell Biol* 1**,** 22-30.
HIRATA, A., KLEIN, B. J. & MURAKAMI, K. S. 2008. The X-ray crystal structure of RNA polymerase from Archaea. *Nature* 451**,** 851-4.
HLINKOVA, V., XING, G., BAUER, J., SHIN, Y. J., DIONNE, I., RAJASHANKAR, K. R., BELL, S. D. & LING, H. 2008. Structures of monomeric, dimeric and trimeric PCNA: PCNA-ring assembly and opening. *Acta Crystallogr Sect D Biol Crystallogr* 64**,** 941-9.
HOLM, L. & PARK, J. 2000. DaliLite workbench for protein structure comparison. *Bioinformatics* 16**,** 566-7.
HOLMES, M. L. & DYALL-SMITH, M. L. 2000. Sequence and expression of a halobacterial beta-galactosidase gene. *Mol Microbiol* 36**,** 114-22.
HOLZLE, A., FISCHER, S., HEYER, R., SCHUTZ, S., ZACHARIAS, M., WALTHER, P., ALLERS, T. & MARCHFELDER, A. 2008. Maturation of the 5S rRNA 5' end is catalyzed in vitro by the endonuclease tRNase Z in the archaeon *H. volcanii*. *RNA,* 14**,** 928-37.
HOPFNER, K. P., EICHINGER, A., ENGH, R. A., LAUE, F., ANKENBAUER, W., HUBER, R. & ANGERER, B. 1999. Crystal structure of a thermostable type B DNA polymerase from *Thermococcus gorgonarius*. *Proc Natl Acad Sci USA* 96**,** 3600-5.
HOPKINS, B. B. & PAULL, T. T. 2008. The P. furiosus mre11/rad50 complex promotes 5' strand resection at a DNA double-strand break. *Cell* 135**,** 250-60.
HOSFIELD, D. J., FRANK, G., WENG, Y., TAINER, J. A. & SHEN, B. 1998a. Newly discovered archaebacterial flap endonucleases show a structure-specific mechanism for DNA substrate binding and catalysis resembling human flap endonuclease-1. *J Biol Chem* 273**,** 27154-61.
HOSFIELD, D. J., MOL, C. D., SHEN, B. & TAINER, J. A. 1998b. Structure of the DNA repair and replication endonuclease and exonuclease FEN-1: coupling DNA and PCNA binding to FEN-1 activity. *Cell* 95**,** 135-46.
HUANG, J., GONG, Z., GHOSAL, G. & CHEN, J. 2009. SOSS complexes participate in the maintenance of genomic stability. *Mol Cell* 35**,** 384-93.

HWANG, K. Y., BAEK, K., KIM, H. Y. & CHO, Y. 1998. The crystal structure of flap endonuclease-1 from *Methanococcus jannaschii*. *Nat Struct Biol* 5**,** 707-13.
IFTODE, C., DANIELY, Y. & BOROWIEC, J. A. 1999. Replication protein A (RPA): The eukaryotic SSB. *Crit Rev Biochem Mol Biol* 34**,** 141-180.
IRIHIMOVITCH, V. & EICHLER, J. 2003. Post-translational secretion of fusion proteins in the halophilic archaea *Haloferax volcanii*. *J Biol Chem* 278**,** 12881-7.
ISHINO, Y., KOMORI, K., CANN, I. K. O. & KOGA, Y. 1998. A novel DNA polymerase family found in Archaea. *J Bacteriol* 180**,** 2232-2236.
ITO, J. & BRAITHWAITE, D. K. 1991. Compilation and alignment of DNA polymerase sequences. *Nucleic Acids Res* 19**,** 4045-57.
ITO, N., NUREKI, O., SHIROUZU, M., YOKOYAMA, S. & HANAOKA, F. 2003. Crystal structure of the Pyrococcus horikoshii DNA primase-UTP complex: implications for the mechanism of primer synthesis. *Genes Cells* 8**,** 913-923.
JEANMOUGIN, F., THOMPSON, J. D., GOUY, M., HIGGINS, D. G. & GIBSON, T. J. 1998. Multiple sequence alignment with Clustal X. *Trends Biochem Sci* 23**,** 403-405.
JUN, S. H., REICHLEN, M. J., TAJIRI, M. & MURAKAMI, K. S. 2011. Archaeal RNA polymerase and transcription regulation. *Crit Rev Biochem Mol Biol* 46**,** 27-40.
KAMADA, K., KUBOTA, Y., ARATA, T., SHINDO, Y. & HANAOKA, F. 2007. Structure of the human GINS complex and its assembly and functional interface in replication initiation. *Nat Struct Mol Biol* 14**,** 388-396.
KANEMAKI, M., SANCHEZ-DIAZ, A., GAMBUS, A. & LABIB, K. 2003. Functional proteomic identification of DNA replication proteins by induced proteolysis *in vivo*. *Nature* 423**,** 720-4.
KELLEY, L. A. & STERNBERG, M. J. 2009. Protein structure prediction on the Web: a case study using the Phyre server. *Nat Protoc* 4**,** 363-71.
KELLY, T. J., SIMANCEK, P. & BRUSH, G. S. 1998. Identification and characterization of a single-stranded DNA-binding protein from the archaeon *Methanococcus jannaschii*. *Proc Natl Acad Sci USA*, 95 14634-14639.
KELMAN, Z. & HURWITZ, J. 2000. A unique organization of the protein subunits of the DNA polymerase clamp loader in the archaeon *Methanobacterium thermoautotrophicum* deltaH. *J Biol Chem* 275**,** 7327-36.
KELMAN, Z., PIETROKOVSKI, S. & HURWITZ, J. 1999. Isolation and characterization of a split B-type DNA polymerase from the archaeon *Methanobacterium thermoautotrophicum* deltaH. *J Biol Chem* 274**,** 28751-61.
KELMAN, Z. & WHITE, M. F. 2005. Archaeal DNA replication and repair. *Curr Opin Microbiol* 8**,** 669-76.
KEMP, M. G., MASON, A. C., CARREIRA, A., REARDON, J. T., HARING, S. J., BORGSTAHL, G. E., KOWALCZYKOWSKI, S. C., SANCAR, A. & WOLD, M. S. 2010. An alternative form of replication protein a expressed in normal human tissues supports DNA repair. *J Biol Chem* 285**,** 4788-97.
KERR, I. D., WADSWORTH, R. I. M., CUBEDDU, L., BLANKENFELDT, W., NAISMITH, J. H. & WHITE, M. F. 2003. Insights into ssDNA recognition by the OB fold from a structural and thermodynamic study of *Sulfolobus* SSB protein. *EMBO J* 22**,** 2561-2570.
KESHAV, K. F., CHEN, C. & DUTTA, A. 1995. Rpa4, a homolog of the 34-kilodalton subunit of the replication protein A complex. *Mol Cell Biol* 15**,** 3119-28.
KIM, J., KIM, O., KIM, H. W., KIM, H. S., LEE, S. J. & SUH, S. W. 2009. ATP-dependent DNA ligase from Archaeoglobus fulgidus displays a tightly closed conformation. *Acta Crystallogr Sect F Struct Biol Crystalliz Commun* 65**,** 544-50.
KOMORI, K. & ISHINO, Y. 2001. Replication protein A in *Pyrococcus furiosus* is involved in homologous DNA recombination. *J Biol Chem* 276**,** 25654-25660.

KOSA, P. F., GHOSH, G., DEDECKER, B. S. & SIGLER, P. B. 1997. The 2.1-A crystal structure of an archaeal preinitiation complex: TATA-box-binding protein/transcription factor (II)B core/TATA-box. *Proc Natl Acad Sci USA* 94**,** 6042-7.
KOSTER, D. A., CRUT, A., SHUMAN, S., BJORNSTI, M. A. & DEKKER, N. H. 2010. Cellular strategies for regulating DNA supercoiling: a single-molecule perspective. *Cell* 142**,** 519-30.
KOSTREWA, D., ZELLER, M. E., ARMACHE, K. J., SEIZL, M., LEIKE, K., THOMM, M. & CRAMER, P. 2009. RNA polymerase II-TFIIB structure and mechanism of transcription initiation. *Nature* 462**,** 323-30.
KRASTANOVA, I., SANNINO, V., AMENITSCH, H., GILEADI, O., PISANI, F. M. & ONESTI, S. 2012. Structural and functional insights into the DNA replication factor Cdc45 reveal an evolutionary relationship to the DHH family of phosphoesterases. *J Biol Chem* 287**,** 4121-8.
KREJCI, L., ALTMANNOVA, V., SPIREK, M. & ZHAO, X. 2012. Homologous recombination and its regulation. *Nucleic Acids Res* 40**,** 5795-818.
KROKAN, H., WIST, E. & KROKAN, R. H. 1981. Aphidicolin inhibits DNA synthesis by DNA polymerase alpha and isolated nuclei by a similar mechanism. *Nucleic Acids Res* 9**,** 4709-19.
KRUPOVIC, M., GRIBALDO, S., BAMFORD, D. H. & FORTERRE, P. 2010. The evolutionary history of archaeal MCM helicases: a case study of vertical evolution combined with hitchhiking of mobile genetic elements. *Mol Biol Evol* 27**,** 2716-32.
KUBOTA, Y., TAKASE, Y., KOMORI, Y., HASHIMOTO, Y., ARATA, T., KAMIMURA, Y., ARAKI, H. & TAKISAWA, H. 2003. A novel ring-like complex of Xenopus proteins essential for the initiation of DNA replication. *Genes Dev* 17**,** 1141-52.
KUCHTA, R. D. & STENGEL, G. 2009. Mechanism and evolution of DNA primases. *Biochim Biophys Acta*.
LAI, L., YOKOTA, H., HUNG, L. W., KIM, R. & KIM, S. H. 2000. Crystal structure of archaeal RNase HII: a homologue of human major RNase H. *Structure* 8**,** 897-904.
LAI, X., SHAO, H., HAO, F. & HUANG, L. 2002. Biochemical characterization of an ATP-dependent DNA ligase from the hyperthermophilic crenarchaeon *Sulfolobus shibatae*. *Extremophiles: life under extreme conditions* 6**,** 469-77.
LAKSANALAMAI, P. & ROBB, F. T. 2004. Small heat shock proteins from extremophiles: a review. *Extremophiles: life under extreme conditions* 8**,** 1-11.
LAM, W. L., COHEN, A., TSOULUHAS, D. & DOOLITTLE, W. F. 1990. Genes for tryptophan biosynthesis in the archaebacterium *Haloferax volcanii*. *Proc Natl Acad Sci USA* 87**,** 6614-8.
LAM, W. L., LOGAN, S. M. & DOOLITTLE, W. F. 1992. Genes for tryptophan biosynthesis in the halophilic archaebacterium *Haloferax volcanii*: the trpDFEG cluster. *J Bacteriol* 174**,** 1694-7.
LAO-SIRIEIX, S. H. & BELL, S. D. 2004. The heterodimeric primase of the hyperthermophilic archaeon *Sulfolobus solfataricus* possesses DNA and RNA primase, polymerase and 3'-terminal nucleotidyl transferase activities. *J Mol Biol* 344**,** 1251-63.
LAO-SIRIEIX, S. H., NOOKALA, R. K., ROVERSI, P., BELL, S. D. & PELLEGRINI, L. 2005. Structure of the heterodimeric core primase. *Nat Struct Mol Biol* 12**,** 1137-44.
LARGE, A., STAMME, C., LANGE, C., DUAN, Z. H., ALLERS, T., SOPPA, J. & LUND, P. A. 2007. Characterization of a tightly controlled promoter of the

halophilic archaeon *Haloferax volcanii* and its use in the analysis of the essential cct1 gene. *Mol Microbiol* 66**,** 1092-1106.
LARKIN, M. A., BLACKSHIELDS, G., BROWN, N. P., CHENNA, R., MCGETTIGAN, P. A., MCWILLIAM, H., VALENTIN, F., WALLACE, I. M., WILM, A., LOPEZ, R., THOMPSON, J. D., GIBSON, T. J. & HIGGINS, D. G. 2007. Clustal W and Clustal X version 2.0. *Bioinformatics* 23**,** 2947-8.
LAWRENCE, C. M. & WHITE, M. F. 2011. Recognition of archaeal CRISPR RNA: No P in the alindromic repeat? *Structure* 19**,** 142-4.
LE BRETON, M., HENNEKE, G., NORAIS, C., FLAMENT, D., MYLLYKALLIO, H., QUERELLOU, J. & RAFFIN, J. P. 2007. The heterodimeric primase from the euryarchaeon *Pyrococcus abyssi*: a multifunctional enzyme for initiation and repair? *J Mol Biol* 374**,** 1172-85.
LEE, J. Y., CHANG, C., SONG, H. K., MOON, J., YANG, J. K., KIM, H. K., KWON, S. T. & SUH, S. W. 2000. Crystal structure of NAD(+)-dependent DNA ligase: modular architecture and functional implications. *EMBO J* 19**,** 1119-29.
LEHMAN, I. R. 1974. DNA ligase: structure, mechanism, and function. *Science* 186**,** 790-7.
LEIGH, J. A., ALBERS, S. V., ATOMI, H. & ALLERS, T. 2011. Model organisms for genetics in the domain Archaea: methanogens, halophiles, Thermococcales and Sulfolobales. *FEMS Microbiol Rev* 35**,** 577-608.
LESTINI, R., DUAN, Z. H. & 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* 9**,** 994-1002.
LI, D., ZHAO, R., LILYESTROM, W., GAI, D., ZHANG, R., DECAPRIO, J. A., FANNING, E., JOCHIMIAK, A., SZAKONYI, G. & CHEN, X. S. 2003. Structure of the replicative helicase of the oncoprotein SV40 large tumour antigen. *Nature* 423**,** 512-8.
LI, Y., BOLDERSON, E., KUMAR, R., MUNIANDY, P. A., XUE, Y., RICHARD, D. J., SEIDMAN, M., PANDITA, T. K., KHANNA, K. K. & WANG, W. 2009. HSSB1 and hSSB2 form similar multiprotein complexes that participate in DNA damage response. *J Biol Chem* 284**,** 23525-31.
LI, Z., SANTANGELO, T. J., CUBONOVA, L., REEVE, J. N. & KELMAN, Z. 2010. Affinity purification of an archaeal DNA replication protein network. *mBio* 1.
LIN, Y., GUZMAN, C. E., MCKINNEY, M. C., NAIR, S. K., & HA, T., CANN, I. K. O. 2006. *Methanosarcina acetivorans* Flap Endonuclease 1 Activity Is Inhibited by a Cognate Single-Stranded-DNA-Binding Protein. *J Bacteriol* 9**,** 6153–6167.
LIN, Y., LIN, L. J., SRIRATANA, P., COLEMAN, K., HA, T., SPIES, M. & CANN, I. K. 2008. Engineering of functional replication protein a homologs based on insights into the evolution of oligonucleotide/oligosaccharide-binding folds. *J Bacteriol* 190**,** 5766-80.
LIN, Y., ROBBINS, J. B., NYANNOR, E. K., CHEN, Y. H. & CANN, I. K. 2005. A CCCH zinc finger conserved in a replication protein a homolog found in diverse Euryarchaeotes. *J Bacteriol* 187**,** 7881-9.
LITTLEFIELD, O., KORKHIN, Y. & SIGLER, P. B. 1999. The structural basis for the oriented assembly of a TBP/TFB/promoter complex. *Proc Natl Acad Sci USA* 96**,** 13668-73.
LIU, H., RUDOLF, J., JOHNSON, K. A., MCMAHON, S. A., OKE, M., CARTER, L., MCROBBIE, A. M., BROWN, S. E., NAISMITH, J. H. & WHITE, M. F. 2008a. Structure of the DNA repair helicase XPD. *Cell* 133**,** 801-12.
LIU, J., SMITH, C. L., DERYCKERE, D., DEANGELIS, K., MARTIN, G. S. & BERGER, J. M. 2000. Structure and function of Cdc6/Cdc18: implications for origin recognition and checkpoint control. *Mol Cell* 6**,** 637-48.

LIU, L., KOMORI, K., ISHINO, S., BOCQUIER, A. A., CANN, I. K., KOHDA, D. & ISHINO, Y. 2001. The archaeal DNA primase: biochemical characterization of the p41-p46 complex from *Pyrococcus furiosus*. *J Biol Chem* 276**,** 45484-90.
LIU, W., PUCCI, B., ROSSI, M., PISANI, F. M. & LADENSTEIN, R. 2008b. Structural analysis of the *Sulfolobus solfataricus* MCM protein N-terminal domain. *Nucleic Acids Res* 36**,** 3235-43.
LUNDGREN, M., ANDERSSON, A., CHEN, L., NILSSON, P. & BERNANDER, R. 2004. Three replication origins in *Sulfolobus* species: synchronous initiation of chromosome replication and asynchronous termination. *Proc Natl Acad Sci USA* 101**,** 7046-51.
MACNEILL, S. A. 2001. Understanding the enzymology of archaeal DNA replication: progress in form and function. *Mol Microbiol* 40**,** 520-9.
MACNEILL, S. A. 2009. The haloarchaeal chromosome replication machinery. *Biochem Soc Trans* 37**,** 108-13.
MACNEILL, S. A. 2010. Structure and function of the GINS complex, a key component of the eukaryotic replisome. *Biochem J* 425**,** 489-500.
MADERN, D., EBEL, C. & ZACCAI, G. 2000. Halophilic adaptation of enzymes. *Extremophiles: life under extreme conditions* 4**,** 91-8.
MAGA, G., FROUIN, I., SPADARI, S. & HUBSCHER, U. 2001. Replication protein A as a "fidelity clamp" for DNA polymerase alpha. *J Biol Chem* 276**,** 18235-42.
MAGA, G. & HUBSCHER, U. 1996. DNA replication machinery: functional characterization of a complex containing DNA polymerase alpha, DNA polymerase delta, and replication factor C suggests an asymmetric DNA polymerase dimer. *Biochemistry* 35**,** 5764-77.
MAGA, G. & HUBSCHER, U. 2003. Proliferating cell nuclear antigen (PCNA): a dancer with many partners. *J Cell Sci* 116**,** 3051-60.
MAINE, G. T., SINHA, P. & TYE, B. K. 1984. Mutants of *S. cerevisiae* defective in the maintenance of minichromosomes. *Genetics* 106**,** 365-85.
MAJKA, J. & BURGERS, P. M. 2004. The PCNA-RFC families of DNA clamps and clamp loaders. *Prog Nucleic Acid Res Mol Biol* 78**,** 227-60.
MAKAROVA, K. S., KOONIN, E. V. & KELMAN, Z. 2012. The CMG (CDC45/RecJ, MCM, GINS) complex is a conserved component of the DNA replication system in all archaea and eukaryotes. *Biol Direct* 7**,** 7.
MAKAROVA, K. S., WOLF, Y. I., MEKHEDOV, S. L., MIRKIN, B. G. & KOONIN, E. V. 2005. Ancestral paralogs and pseudoparalogs and their role in the emergence of the eukaryotic cell. *Nucleic Acids Res* 33**,** 4626-38.
MARG, B. L., SCHWEIMER, K., STICHT, H. & OESTERHELT, D. 2005. A two-alpha-helix extra domain mediates the halophilic character of a plant-type ferredoxin from halophilic archaea. *Biochemistry* 44**,** 29-39.
MARINSEK, N., BARRY, E. R., MAKAROVA, K. S., DIONNE, I., KOONIN, E. V. & BELL, S. D. 2006. GINS, a central nexus in the archaeal DNA replication fork. *EMBO Rep* 7**,** 539-545.
MARTIN, I. V. & MACNEILL, S. A. 2002. ATP-dependent DNA ligases. *Genome Biol* 3**,** REVIEWS3005.
MASE, T., KUBOTA, K., MIYAZONO, K., KAWARABAYASI, Y. & TANOKURA, M. 2011. Structure of flap endonuclease 1 from the hyperthermophilic archaeon *Desulfurococcus amylolyticus*. *Acta Crystallogr Sect F Struct Biol Crystalliz Commun* 67**,** 209-13.
MATOBA, K., MAYANAGI, K., NAKASU, S., KIKUCHI, A. & MORIKAWA, K. 2002. Three-dimensional electron microscopy of the reverse gyrase from *Sulfolobus tokodaii*. *Biochem Biophys Res Commun* 297**,** 749-55.

MATSUI, E., KAWASAKI, S., ISHIDA, H., ISHIKAWA, K., KOSUGI, Y., KIKUCHI, H., KAWARABAYASHI, Y. & MATSUI, I. 1999. Thermostable flap endonuclease from the archaeon, *Pyrococcus horikoshii*, cleaves the replication fork-like structure endo/exonucleolytically. *J Biol Chem* 274**,** 18297-309.
MATSUI, E., NISHIO, M., YOKOYAMA, H., HARATA, K., DARNIS, S. & MATSUI, I. 2003. Distinct domain functions regulating de novo DNA synthesis of thermostable DNA primase from hyperthermophile *Pyrococcus horikoshii*. *Biochemistry* 42**,** 14968-76.
MATSUMIYA, S., ISHINO, Y. & MORIKAWA, K. 2001. Crystal structure of an archaeal DNA sliding clamp: proliferating cell nuclear antigen from *Pyrococcus furiosus*. *Protein Sci,* 10**,** 17-23.
MATTHEWS, J. M. & SUNDE, M. 2002. Zinc fingers--folds for many occasions. *IUBMB Life* 54**,** 351-5.
MCCREADY, S., JOCHEN A MÜLLER, J. A., BOUBRIAK, I., BERQUIST, B. R., LOON, W. & DASSARMA, S. 2005. UV irradiation induces homologous recombination genes in the model archaeon, *Halobacterium* sp. NRC-1. *Saline Syst* 1.
MCCREADY, S. & MARCELLO, L. 2003. Repair of UV damage in *Halobacterium salinarum*. *Biochem Soc Trans* 31**,** 694-8.
MCNALLY, R., BOWMAN, G. D., GOEDKEN, E. R., O'DONNELL, M. & KURIYAN, J. 2010. Analysis of the role of PCNA-DNA contacts during clamp loading. *BMC Struct Biol* 10**,** 3.
MESLET-CLADIERE, L., NORAIS, C., KUHN, J., BRIFFOTAUX, J., SLOOSTRA, J. W., FERRARI, E., HUBSCHER, U., FLAMENT, D. & MYLLYKALLIO, H. 2007. A novel proteomic approach identifies new interaction partners for proliferating cell nuclear antigen. *J Mol Biol* 372**,** 1137-48.
MEVARECH, M., FROLOW, F. & GLOSS, L. M. 2000. Halophilic enzymes: proteins with a grain of salt. *Biophys Chem* 86**,** 155-64.
MOLDOVAN, G. L., PFANDER, B. & JENTSCH, S. 2007. PCNA, the maestro of the replication fork. *Cell* 129**,** 665-679.
MORAG, E., LAPIDOT, A., GOVORKO, D., LAMED, R., WILCHEK, M., BAYER, E. A. & SHOHAM, Y. 1995. Expression, Purification, and Characterization of the Cellulose-Binding Domain of the Scaffoldin Subunit from the Cellulosome of *Clostridium thermocellum*. *Appl Environ Microbiol* 61**,** 1980-1986.
MORGUNOVA, E., GRAY, F. C., MACNEILL, S. A. & LADENSTEIN, R. 2009. Structural insights into the adaptation of proliferating cell nuclear antigen (PCNA) from *Haloferax volcanii* to a high-salt environment. *Acta Crystallogr Sect D Biol Crystallog* 65**,** 1081-8.
MOYER, S. E., LEWIS, P. W. & BOTCHAN, M. R. 2006. Isolation of the Cdc45/Mcm2-7/GINS (CMG) complex, a candidate for the eukaryotic DNA replication fork helicase. *Proc Natl Acad Sci USA* 103**,** 10236-41.
MULLAKHANBHAI, M. F. & LARSEN, H. 1975. *Halobacterium volcanii* Spec Nov a Dead Sea *Halobacterium* with a Moderate Salt Requirement. *Arch Microbiol* 104**,** 207-214.
MURZIN, A. G. 1993. OB(oligonucleotide/oligosaccharide binding)-fold:
common structural and functional solution for non-homologous sequences. *EMBO J* 12**,** 861-867.
MURZIN, A. G., BRENNER, S. E., HUBBARD, T. & CHOTHIA, C. 1995. SCOP: a structural classification of proteins database for the investigation of sequences and structures. *J Mol Biol* 247**,** 536-40.
MYLLYKALLIO, H., LOPEZ, P., LOPEZ-GARCIA, P., HEILIG, R., SAURIN, W., ZIVANOVIC, Y., PHILIPPE, H. & FORTERRE, P. 2000. Bacterial mode of

replication with eukaryotic-like machinery in a hyperthermophilic archaeon. *Science* 288**,** 2212-5.
NAJI, S., GRUNBERG, S. & THOMM, M. 2007. The RPB7 orthologue E' is required for transcriptional activity of a reconstituted archaeal core enzyme at low temperatures and stimulates open complex formation. *J Biol Chem* 282**,** 11047-57.
NAKATANI, M., EZAKI, S., ATOMI, H. & IMANAKA, T. 2002. Substrate recognition and fidelity of strand joining by an archaeal DNA ligase. *Eur J Biochem/FEBS* 269**,** 650-6.
NAKAYA, R., TAKAYA, J., ONUKI, T., MORITANI, M., NOZAKI, N. & ISHIMI, Y. 2010. Identification of proteins that may directly interact with human RPA. *J Biochem* 148**,** 539-47.
NEVES, C., DA COSTA, M. S. & SANTOS, H. 2005. Compatible solutes of the hyperthermophile *Palaeococcus ferrophilus*: osmoadaptation and thermoadaptation in the order thermococcales. *Appl Environ Microbiol* 71**,** 8091-8.
NG, W. V., KENNEDY, S. P., MAHAIRAS, G. G., BERQUIST, B., PAN, M., SHUKLA, H. D., LASKY, S. R., BALIGA, N. S., THORSSON, V., SBROGNA, J., SWARTZELL, S., WEIR, D., HALL, J., DAHL, T. A., WELTI, R., GOO, Y. A., LEITHAUSER, B., KELLER, K., CRUZ, R., DANSON, M. J., HOUGH, D. W., MADDOCKS, D. G., JABLONSKI, P. E., KREBS, M. P., ANGEVINE, C. M., DALE, H., ISENBARGER, T. A., PECK, R. F., POHLSCHRODER, M., SPUDICH, J. L., JUNG, K. W., ALAM, M., FREITAS, T., HOU, S., DANIELS, C. J., DENNIS, P. P., OMER, A. D., EBHARDT, H., LOWE, T. M., LIANG, P., RILEY, M., HOOD, L. & DASSARMA, S. 2000. Genome sequence of *Halobacterium* species NRC-1. *Proc Natl Acad Sci USA* 97**,** 12176-81.
NICHOLS, M. D., DEANGELIS, K., KECK, J. L. & BERGER, J. M. 1999. Structure and function of an archaeal topoisomerase VI subunit with homology to the meiotic recombination factor Spo11. *EMBO J* 18**,** 6177-88.
NISHIDA, H., KIYONARI, S., ISHINO, Y. & MORIKAWA, K. 2006. The closed structure of an archaeal DNA ligase from *Pyrococcus furiosus*. *J Mol Biol* 360**,** 956-67.
NORAIS, C., HAWKINS, M., HARTMAN, A. L., EISEN, J. A., MYLLYKALLIO, H. & ALLERS, T. 2007a. Genetic and physical mapping of DNA replication origins in *Haloferax volcanii*. *PLoS Genet* 3**,** e77.
NORAIS, C., HAWKINS, M., HARTMAN, A. L., EISEN, J. A., MYLLYKALLIO, H. & ALLERS, T. 2007b. Genetic and physical mapping of DNA replication origins in *Haloferax volcanii*. *PLoS Genet*, 3**,** e77.
OESTERHELT, D. & STOECKENIUS, W. 1971. Rhodopsin-like protein from the purple membrane of *Halobacterium halobium*. *Nat New Biol*, 233**,** 149-52.
OHTANI, N., HARUKI, M., MORIKAWA, M., CROUCH, R. J., ITAYA, M. & KANAYA, S. 1999. Identification of the genes encoding Mn2+-dependent RNase HII and Mg2+-dependent RNase HIII from Bacillus subtilis: classification of RNases H into three families. *Biochemistry* 38**,** 605-18.
OHTANI, N., YANAGAWA, H., TOMITA, M. & ITAYA, M. 2004. Identification of the first archaeal Type 1 RNase H gene from *Halobacterium* sp. NRC-1: archaeal RNase HI can cleave an RNA-DNA junction. *Biochem J* 381**,** 795-802.
OREN, A. 1999. Bioenergetic aspects of halophilism. *Microbiol Mol Biol R: MMBR* 63**,** 334-48.
OREN, A. 2008. Microbial life at high salt concentrations: phylogenetic and metabolic diversity. *Saline Syst* 4**,** 2.
OREN, A. 2010. Industrial and environmental applications of halophilic microorganisms. *Environ Technol* 31**,** 825-34.

OYAMA, T., ISHINO, S., FUJINO, S., OGINO, H., SHIRAI, T., MAYANAGI, K., SAITO, M., NAGASAWA, N., ISHINO, Y. & MORIKAWA, K. 2011. Architectures of archaeal GINS complexes, essential DNA replication initiation factors. *BMC Biol* 9**,** 28.
PACEK, M., TUTTER, A. V., KUBOTA, Y., TAKISAWA, H. & WALTER, J. C. 2006. Localization of MCM2-7, Cdc45, and GINS to the site of DNA unwinding during eukaryotic DNA replication. *Mol Cell* 21**,** 581-7.
PALMER, J. R. & DANIELS, C. J. 1995. *In-Vivo* Definition of an Archaeal Promoter. *J Bacteriol* 177**,** 1844-1849.
PAN, M., SANTANGELO, T. J., LI, Z., REEVE, J. N. & KELMAN, Z. 2011. *Thermococcus kodakaraensis* encodes three MCM homologs but only one is essential. *Nucleic Acids Res* 39**,** 9671-80.
PAPE, T., MEKA, H., CHEN, S., VICENTINI, G., VAN HEEL, M. & ONESTI, S. 2003. Hexameric ring structure of the full-length archaeal MCM protein complex. *Embo Rep* 4**,** 1079-83.
PAPKE, R. T., KOENIG, J. E., RODRIGUEZ-VALERA, F. & DOOLITTLE, W. F. 2004. Frequent recombination in a saltern population of *Halorubrum*. *Science* 306**,** 1928-9.
PAPKE, R. T., ZHAXYBAYEVA, O., FEIL, E. J., SOMMERFELD, K., MUISE, D. & DOOLITTLE, W. F. 2007. Searching for species in haloarchaea. *Proc Natl Acad Sci USA* 104**,** 14092-7.
PASCAL, J. M., TSODIKOV, O. V., HURA, G. L., SONG, W., COTNER, E. A., CLASSEN, S., TOMKINSON, A. E., TAINER, J. A. & ELLENBERGER, T. 2006. A flexible interface between DNA ligase and PCNA supports conformational switching and efficient ligation of DNA. *Mol Cell* 24**,** 279-91.
PAYTUBI, S., MCMAHON, S. A., GRAHAM, S., LIU, H., BOTTING, C. H., MAKAROVA, K. S., KOONIN, E. V., NAISMITH, J. H. & WHITE, M. F. 2012. Displacement of the canonical single-stranded DNA-binding protein in the Thermoproteales. *Proc Natl Acad Sci USA* 109**,** E398-405.
PENG, N., AO, X., LIANG, Y. X. & SHE, Q. 2011. Archaeal promoter architecture and mechanism of gene activation. *Biochem Soc Trans* 39**,** 99-103.
PERLER, F. B., KUMAR, S. & KONG, H. 1996. Thermostable DNA polymerases. *Adv Protein Chem* 48**,** 377-435.
PESTRYAKOV, P. E. & LAVRIK, O. I. 2008. Mechanisms of Single-Stranded DNA-Binding Protein Functioning in Cellular DNA Metabolism. *Biochem-Moscow* 73**,** 1388-1404.
PFAFFL, M. W. 2001. A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res* 29**,** e45.
PFAFFL, M. W., HORGAN, G. W. & DEMPFLE, L. 2002. Relative expression software tool (REST) for group-wise comparison and statistical analysis of relative expression results in real-time PCR. *Nucleic Acids Res* 30**,** e36.
PFEIFFER, F., SCHUSTER, S. C., BROICHER, A., FALB, M., PALM, P., RODEWALD, K., RUEPP, A., SOPPA, J., TITTOR, J. & OESTERHELT, D. 2008. Evolution in the laboratory: the genome of *Halobacterium salinarum* strain R1 compared to that of strain NRC-1. *Genomics*, 91**,** 335-46.
PIARD, K., BALDACCI, G. & TRATNER, I. 1998. Single point mutations located outside the inter-monomer domains abolish trimerization of *Schizosaccharomyces pombe* PCNA. *Nucleic Acids Res* 26**,** 2598-605.
PISANI, F. M., DE FELICE, M., CARPENTIERI, F. & ROSSI, M. 2000. Biochemical characterization of a clamp-loader complex homologous to eukaryotic replication factor C from the hyperthermophilic archaeon *Sulfolobus solfataricus*. *J Mol Biol* 301**,** 61-73.

POSTOW, L., CRISONA, N. J., PETER, B. J., HARDY, C. D. & COZZARELLI, N. R. 2001. Topological challenges to DNA replication: conformations at the fork. *Proc Natl Acad Sci USA* 98**,** 8219-26.
PRESTON, C. M., WU, K. Y., MOLINSKI, T. F. & DELONG, E. F. 1996. A psychrophilic crenarchaeon inhabits a marine sponge: *Cenarchaeum symbiosum* gen. nov., sp. nov. *Proc Natl Acad Sci USA* 93**,** 6241-6.
PRICE, C. M., BOLTZ, K. A., CHAIKEN, M. F., STEWART, J. A., BEILSTEIN, M. A. & SHIPPEN, D. E. 2010. Evolution of CST function in telomere maintenance. *Cell cycle* 9**,** 3157-65.
RAMILO, C., GU, L., GUO, S., ZHANG, X., PATRICK, S. M., TURCHI, J. J. & LI, G. M. 2002. Partial reconstitution of human DNA mismatch repair in vitro: characterization of the role of human replication protein A. *Mol Cell Biol* 22**,** 2037-46.
RAMPAKAKIS, E., GKOGKAS, C., DI PAOLA, D. & ZANNIS-HADJOPOULOS, M. 2010. Replication initiation and DNA topology: The twisted life of the origin. *J Cell Biochem* 110**,** 35-43.
RAO, H. G., ROSENFELD, A. & WETMUR, J. G. 1998. *Methanococcus jannaschii* flap endonuclease: expression, purification, and substrate requirements. *J Bacteriol* 180**,** 5406-12.
REARDON, J. T. & SANCAR, A. 2005. Nucleotide excision repair. *Prog Nucleic Acid Res Mol Biol* 79**,** 183-235.
REEVE, J. N. 2003. Archaeal chromatin and transcription. *Mol Microbiol* 48**,** 587-598.
REITER, W. D., HUDEPOHL, U. & ZILLIG, W. 1990. Mutational analysis of an archaebacterial promoter: essential role of a TATA box for transcription efficiency and start-site selection in vitro. *Proc Natl Acad Sci USA* 87**,** 9509-13.
RICHARD, D. J., BOLDERSON, E., CUBEDDU, L., WADSWORTH, R. I., SAVAGE, K., SHARMA, G. G., NICOLETTE, M. L., TSVETANOV, S., MCILWRAITH, M. J., PANDITA, R. K., TAKEDA, S., HAY, R. T., GAUTIER, J., WEST, S. C., PAULL, T. T., PANDITA, T. K., WHITE, M. F. & KHANNA, K. K. 2008. Single-stranded DNA-binding protein hSSB1 is critical for genomic stability. *Nature* 453**,** 677-81.
RICHARD, D. J., CUBEDDU, L., URQUHART, A. J., BAIN, A., BOLDERSON, E., MENON, D., WHITE, M. F. & KHANNA, K. K. 2011. hSSB1 interacts directly with the MRN complex stimulating its recruitment to DNA double-strand breaks and its endo-nuclease activity. *Nucleic Acids Res* 39**,** 3643-51.
ROBBINS, J. B., MCKINNEY, M. C., GUZMAN, C. E., SRIRATANA, B., FITZ-GIBBON, S., HA, T. & CANN, I. K. O. 2005. The euryarchaeota, nature's medium for engineering of single-stranded DNA-binding proteins. *J Biol Chem* 280**,** 15325-15339.
ROBBINS, J. B., MURPHY, M. C., WHITE, B. A., MACKIE, R. I., HA, T. & CANN, I. K. 2004a. Functional analysis of multiple single-stranded DNA-binding proteins from *Methanosarcina acetivorans* and their effects on DNA synthesis by DNA polymerase BI. *J Biol Chem* 279**,** 6315-26.
ROBBINS, J. B., MURPHY, M. C., WHITE, B. A., MACKIE, R. I., HA, T. & CANN, I. K. O. 2004b. Functional analysis of multiple single-stranded DNA-binding proteins from *Methanosarcina acetivorans* and their effects on DNA synthesis by DNA polymerase BI. *J Biol Chem* 279**,** 6315-6326.
RODRIGUEZ, A. C., PARK, H. W., MAO, C. & BEESE, L. S. 2000. Crystal structure of a pol alpha family DNA polymerase from the hyperthermophilic archaeon *Thermococcus* sp. 9 degrees N-7. *J Mol Biol* 299**,** 447-62.

RODRIGUEZ, A. C. & STOCK, D. 2002. Crystal structure of reverse gyrase: insights into the positive supercoiling of DNA. *EMBO J* 21**,** 418-26.
ROLLAND, J. L., GUEGUEN, Y., PERSILLON, C., MASSON, J. M. & DIETRICH, J. 2004. Characterization of a thermophilic DNA ligase from the archaeon *Thermococcus fumicolans*. *FEMS Microbiol Lett* 236**,** 267-73.
ROTHENBERG, E., TRAKSELIS, M. A., BELL, S. D. & HA, T. 2007. MCM forked substrate specificity involves dynamic interaction with the 5'-tail. *J Biol Chem* 282**,** 34229-34.
ROY, R., KOZLOV, A. G., LOHMAN, T. M. & HA, T. 2007. Dynamic structural rearrangements between DNA binding modes of E. coli SSB protein. *J Mol Biol* 369**,** 1244-57.
RUSSELL, H. J., RICHARDSON, T. T., EMPTAGE, K. & CONNOLLY, B. A. 2009. The 3'-5' proofreading exonuclease of archaeal family-B DNA polymerase hinders the copying of template strand deaminated bases. *Nucleic Acids Res* 37**,** 7603-11.
SAKAGUCHI, K., ISHIBASHI, T., UCHIYAMA, Y. & IWABATA, K. 2009. The multi-replication protein A (RPA) system – a new perspective. *FEBS J* 276**,** 943-963.
SAKAKIBARA, N., KELMAN, L. M. & KELMAN, Z. 2009. Unwinding the structure and function of the archaeal MCM helicase. *Mol Microbiol* 72**,** 286-96.
SAKURABA, H., GODA, S. & OHSHIMA, T. 2004. Unique sugar metabolism and novel enzymes of hyperthermophilic archaea. *Cheml Rec* 3**,** 281-7.
SCHAFER, G., ENGELHARD, M. & MULLER, V. 1999. Bioenergetics of the Archaea. *Microbiol Mol Biol R: MMBR* 63**,** 570-620.
SCHLEPER, C., JURGENS, G. & JONUSCHEIT, M. 2005. Genomic studies of uncultivated archaea. *Nat Rev Microbiol* 3**,** 479-88.
SCHLEPER, C. & NICOL, G. W. 2010. Ammonia-oxidising archaea--physiology, ecology and evolution. *Adv Microb Physiol* 57**,** 1-41.
SEDGWICK, B., BATES, P. A., PAIK, J., JACOBS, S. C. & LINDAHL, T. 2007. Repair of alkylated DNA: recent advances. *DNA repair* 6**,** 429-42.
SEYBERT, A. & WIGLEY, D. B. 2004. Distinct roles for ATP binding and hydrolysis at individual subunits of an archaeal clamp loader. *EMBO J* 23**,** 1360-71.
SHAN, B., XU, J., ZHUO, Y., MORRIS, C. A. & MORRIS, G. F. 2003. Induction of p53-dependent activation of the human proliferating cell nuclear antigen gene in chromatin by ionizing radiation. *J Biol Chem* 278**,** 44009-17.
SHEN, Y., MUSTI, K., HIRAMOTO, M., KIKUCHI, H., KAWARABAYASHI, Y. & MATSUI, I. 2001. Invariant Asp-1122 and Asp-1124 are essential residues for polymerization catalysis of family D DNA polymerase from *Pyrococcus horikoshii*. *J Biol Chem* 276**,** 27376-83.
SHIMMIN, L. C. & DENNIS, P. P. 1996. Conserved sequence elements involved in regulation of ribosomal protein gene expression in halophilic archaea. *J Bacteriol* 178**,** 4737-4741.
SKOWYRA, A. & MACNEILL, S. A. 2012. Identification of essential and non-essential single-stranded DNA-binding proteins in a model archaeal organism. *Nucleic Acids Res* 40**,** 1077-90.
SLAYMAKER, I. M. & CHEN, X. S. 2012. MCM Structure and Mechanics: What We Have Learned from Archaeal MCM. *Subcell Biochem* 62**,** 89-111.
SLEIGH, M. J. 1976. The mechanism of DNA breakage by phleomycin *in vitro*. *Nucleic Acids Res* 3**,** 891-901.
SOPPA, J. 2011. Ploidy and gene conversion in Archaea. *Biochem Soc Trans* 39**,** 150-4.

SOPPA, J., BAUMANN, A., BRENNEIS, M., DAMBECK, M., HERING, O. & LANGE, C. 2008. Genomics and functional genomics with haloarchaea. *Arch Microbiol*, 190**,** 197-215.
SOPPA, J. & LINK, T. A. 1997. The TATA-box-binding protein (TBP) of *Halobacterium salinarum*. Cloning of the tbp gene, heterologous production of TBP and folding of TBP into a native conformation. *Eur J Biochem / FEBS* 249**,** 318-24.
STILLER, J. W. & HALL, B. D. 2002. Evolution of the RNA polymerase II C-terminal domain. *Proc Natl Acad Sci USA* 99**,** 6091-6.
STILLMAN, B. 2008. DNA polymerases at the replication fork in eukaryotes. *Mol Cell* 30**,** 259-60.
STROUD, A., LIDDELL, S. & ALLERS, T. 2012. Genetic and Biochemical Identification of a Novel Single-Stranded DNA-Binding Complex in *Haloferax volcanii*. *Fron Microbiol* 3**,** 224.
STRZALKA, W. & ZIEMIENOWICZ, A. 2011. Proliferating cell nuclear antigen (PCNA): a key factor in DNA replication and cell cycle regulation. *An Botany* 107**,** 1127-40.
SUN, J. & KONG, D. 2010. DNA replication origins, ORC/DNA interaction, and assembly of pre-replication complex in eukaryotes. *Acta Bioch Biophys Sci* 42**,** 433-9.
SWIATEK, A. & MACNEILL, S. A. 2010. The archaeo-eukaryotic GINS proteins and the archaeal primase catalytic subunit PriS share a common domain. *Biol Direct* 5**,** 17.
TAKAHASHI, T. S., WIGLEY, D. B. & WALTER, J. C. 2005. Pumps, paradoxes and ploughshares: mechanism of the MCM2-7 DNA helicase. *Trends Biochem Sci* 30**,** 437-44.
TAKAYAMA, Y., KAMIMURA, Y., OKAWA, M., MURAMATSU, S., SUGINO, A. & ARAKI, H. 2003. GINS, a novel multiprotein complex required for chromosomal DNA replication in budding yeast. *Genes Dev* 17**,** 1153-65.
THEOBALD, D. L., MITTON-FRY, R. M. & WUTTKE, D. S. 2003. Nucleic acid recognition by OB-fold proteins. *Annu Rev Biophys Biomol Struct* 32**,** 115-133.
THOMAS, T. & CAVICCHIOLI, R. 2000. Effect of temperature on stability and activity of elongation factor 2 proteins from Antarctic and thermophilic methanogens. *J Bacteriol* 182**,** 1328-32.
THOMSEN, J., DE BIASE, A., KACZANOWSKI, S., MACARIO, A. J., THOMM, M., ZIELENKIEWICZ, P., MACCOLL, R. & CONWAY DE MACARIO, E. 2001. The basal transcription factors TBP and TFB from the mesophilic archaeon *Methanosarcina mazeii*: structure and conformational changes upon interaction with stress-gene promoters. *J Mol Biol* 309**,** 589-603.
TIRANTI, V., ROCCHI, M., DIDONATO, S. & ZEVIANI, M. 1993. Cloning of human and rat cDNAs encoding the mitochondrial single-stranded DNA-binding protein (SSB). *Gene* 126**,** 219-25.
VAN DE VOSSENBERG, J. L., DRIESSEN, A. J. & KONINGS, W. N. 1998. The essence of being extremophilic: the role of the unique archaeal membrane lipids. *Extremophiles: life under extreme conditions* 2**,** 163-70.
VIVONA, J. B. & KELMAN, Z. 2003. The diverse spectrum of sliding clamp interacting proteins. *FEBS Lett* 546**,** 167-72.
WALTERS, A. D. & CHONG, J. P. 2009. *Methanococcus maripaludis*: an archaeon with multiple functional MCM proteins? *Biochem Soc Trans* 37**,** 1-6.
WALTERS, A. D. & CHONG, J. P. 2010. An archaeal order with multiple minichromosome maintenance genes. *Microbiology* 156**,** 1405-14.
WARBRICK, E. 1998. PCNA binding through a conserved motif. *BioEssays* 20**,** 195-9.

WASEEM, N. H., LABIB, K., NURSE, P. & LANE, D. P. 1992. Isolation and analysis of the fission yeast gene encoding polymerase delta accessory protein PCNA. *EMBO J* 11**,** 5111-20.
WENDOLOSKI, D., FERRER, C. & 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**,** 959-64.
WERNER, F. & WEINZIERL, R. O. 2002. A recombinant RNA polymerase II-like enzyme capable of promoter-specific transcription. *Mol Cell* 10**,** 635-46.
WESSEL, R., SCHWEIZER, J. & STAHL, H. 1992. Simian virus 40 T-antigen DNA helicase is a hexamer which forms a binary complex during bidirectional unwinding from the viral origin of DNA replication. *J Virol* 66**,** 804-15.
WHITE, M. F. 2011. Homologous recombination in the archaea: the means justify the ends. *Biochem Soc Trans* 39**,** 15-9.
WILKINSON, A., DAY, J. & BOWATER, R. 2001. Bacterial DNA ligases. *Mol Microbiol* 40**,** 1241-8.
WILLIAMS, G. J., JOHNSON, K., RUDOLF, J., MCMAHON, S. A., CARTER, L., OKE, M., LIU, H., TAYLOR, G. L., WHITE, M. F. & NAISMITH, J. H. 2006. Structure of the heterotrimeric PCNA from *Sulfolobus solfataricus*. *Acta Crystallogr Sect F Struct Biol Crystalliz Commun* 62**,** 944-8.
WINTER, J. A., CHRISTOFI, P., MORROLL, S. & BUNTING, K. A. 2009. The crystal structure of *Haloferax volcanii* proliferating cell nuclear antigen reveals unique surface charge characteristics due to halophilic adaptation. *BMC Struct Biol* 9**,** 55.
WINTER, J. A., PATOLI, B. & BUNTING, K. A. 2012. DNA Binding in High Salt: Analysing the Salt Dependence of Replication Protein A3 from the Halophile *Haloferax volcanii*. *Archaea* 2012**,** 719092.
WOESE, C. R. & FOX, G. E. 1977. Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proc Natl Acad Sci USA* 74**,** 5088-90.
WOESE, C. R., KANDLER, O. & WHEELIS, M. L. 1990. Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proc Natl Acad Sci USA* 87**,** 4576-9.
WOYCHIK, N. A. & HAMPSEY, M. 2002. The RNA polymerase II machinery: structure illuminates function. *Cell* 108**,** 453-63.
XIE, Y. & REEVE, J. N. 2004. Transcription by *Methanothermobacter thermautotrophicus* RNA polymerase in vitro releases archaeal transcription factor B but not TATA-box binding protein from the template DNA. *J Bacteriol* 186**,** 6306-10.
XU, H., ZHANG, P., LIU, L. & LEE, M. Y. 2001. A novel PCNA-binding motif identified by the panning of a random peptide display library. *Biochemistry* 40**,** 4512-20.
YANG, W. 2011. Surviving the sun: repair and bypass of DNA UV lesions. *Protein Sci* 20**,** 1781-9.
YOSHIMOCHI, T., FUJIKANE, R., KAWANAMI, M., MATSUNAGA, F. & ISHINO, Y. 2008a. The GINS complex from *Pyrococcus furiosus* stimulates the MCM helicase activity. *J Biol Chem* 283**,** 1601-1609.
YOSHIMOCHI, T., FUJIKANE, R., KAWANAMI, M., MATSUNAGA, F. & ISHINO, Y. 2008b. The GINS complex from *Pyrococcus furiosus* stimulates the MCM helicase activity. *J Biol Chem* 283**,** 1601-9.
ZHANG, F., WU, J. & YU, X. 2009. Integrator3, a partner of single-stranded DNA-binding protein 1, participates in the DNA damage response. *J Biol Chem* 284**,** 30408-15.

ZHANG, G., CAMPBELL, E. A., MINAKHIN, L., RICHTER, C., SEVERINOV, K. & DARST, S. A. 1999. Crystal structure of *Thermus aquaticus* core RNA polymerase at 3.3 A resolution. *Cell* 98**,** 811-24.
ZHANG, J., ROUILLON, C., KEROU, M., REEKS, J., BRUGGER, K., GRAHAM, S., REIMANN, J., CANNONE, G., LIU, H., ALBERS, S. V., NAISMITH, J. H., SPAGNOLO, L. & WHITE, M. F. 2012. Structure and mechanism of the CMR complex for CRISPR-mediated antiviral immunity. *Mol Cell* 45**,** 303-13.
ZHANG, R. & ZHANG, C. T. 2003. Multiple replication origins of the archaeon *Halobacterium species* NRC-1. *Biochem Biophys Res Commun* 302**,** 728-34.
ZHANG, X. & WIGLEY, D. B. 2008. The 'glutamate switch' provides a link between ATPase activity and ligand binding in AAA+ proteins. *Nat Struct Mol Biol* 15**,** 1223-7.
ZHAO, A., GRAY, F. C. & MACNEILL, S. A. 2006. ATP- and NAD+-dependent DNA ligases share an essential function in the halophilic archaeon *Haloferax volcanii*. *Mol Microbiol* 59**,** 743-52.
ZHAO, Y., JERUZALMI, D., MOAREFI, I., LEIGHTON, L., LASKEN, R. & KURIYAN, J. 1999. Crystal structure of an archaebacterial DNA polymerase. *Structure* 7**,** 1189-99.
ZHU, W., ZENG, Q., COLANGELO, C. M., LEWIS, M., SUMMERS, M. F. & SCOTT, R. A. 1996. The N-terminal domain of TFIIB from *Pyrococcus furiosus* forms a zinc ribbon. *Nat Struct Biol* 3**,** 122-4.
ZHUANG, Z., YODER, B. L., BURGERS, P. M. & BENKOVIC, S. J. 2006. The structure of a ring-opened proliferating cell nuclear antigen-replication factor C complex revealed by fluorescence energy transfer. *Proc Natl Acad Sci USA* 103**,** 2546-51.
ZOU, Y., LIU, Y. Y., WU, X. M. & SHELL, S. M. 2006. Functions of human replication protein A (RPA): From DNA replication to DNA damage and stress responses. *J Cell Physiol* 208**,** 267-273.

Appendix

Table A1 Oligonucleotides used in this study

No.	Oligonucleotide	Sequence (5'→3')	Restriction site
Construction and screening of plasmids for regulated gene expression (pNMP01-series)			
P1	NPM01-5Bcl	GGTTGGTTGTGATCAGCTCGTGGCGAGAACGGAAC	BclI
P2	NPM01-3NPN	GTTGTTGTTGGCGGCCGCTTAATTAACATATGCGCAATAGGTCCGCG	NotI, PacI, NdeI
P3	HfxPolD2-5Nde	GTGTGTGTGTCATATGCGCGAGGAGGAAACCCGGT	NdeI
P4	HfxPolD2-3V	GTGTGTGTGTGATATCCTCGGTGAGGAGGGCGACG	EcoRV
P5	HfxPolD1-5Nde	GTGTGTGTGTCATATGCCACTGGAGACGCCGGCGC	NdeI
P6	HfxPolD1-3V	GGTGGTGGTTGATATCGTTGTCGTGCGATGTTGCA	EcoRV
P7	HfxPolB-5Nde	GTGTGGTTGTCATATGACGCAGACGGGTCTGACCG	NdeI
P8	HfxPolB-3V	GGTGGTGTGTGATATCGAGGTCGCCGTCGTCGAGG	EcoRV
P9	HfxPriL-3V	ACCTATTGCGCATATGCGCCCGCTCCACGC	NdeI
P10	HfxPriL-TNA-R	CGGCGTCGAAGTCGGCGAGC	N/A
P11	HfxRfcA	ACCTATTGCGCATATGAGCGAGGCCGCGGA	NdeI
P12	HfxRfcA-TNA-R	GTCGTGACCGCCGAACGACG	N/A
P13	HfxMCM-5Nde	GTGTGGTTGTCATATGGCGCAGGCCCCCCAGAACCGAGAC	NdeI
P14	HfxMCM-TNA-R	CCGAGATGAGCGTCCCGATG	N/A
P15	HfxRpaA1-5Nde	GTGTGTGGTTCATATGGAACTCGACCAGCATGCCG	NdeI
P16	HfxRpaA1-3V	GTGTGGTTGTGATATCTCCCTCGCGGATGGTCTGT	EcoRV
P17	HfxRpaB1-5Nde	GGTTGGTTGTCATATGACTGATTTGCGAACCCATG	NdeI
P18	HfxRpaB1-5Nde	GTGTGTGGTTGATATCGCTCCGGGGTTCCCAGAGT	EcoRV
P19	HfxRpaC-5Nde	GGTTGGTTGTCATATGGGCGTCATCCGGGAGGTCT	NdeI
P20	HfxRpaC-5Nde	GTGTGTGGTTGATATCGTCGGCGACCTCGATGTTG	EcoRV
P21	Hfx-PolB-R500	CGCCGCCTTTGATGTCCTTG	N/A
P22	Hfx-PolD1-R500	GCCCGCGGAGCAGTTTCGAG	N/A
P23	Hfx-PolD2-R500	GGGCTTTGTACTCGTCGATG	N/A
P24	Hfx-PriS-R750	GTTGCCCTCGCGGACGGCGT	N/A
P25	Hfx-PriL-R500	GCCAGTCCGCGCCCCAGTCG	N/A
P26	Hfx-RpaA-R500	GCAGGTCGATGAGGTCGGCC	N/A
P27	Hfx-RpaB-R500	TGGTCGAGTCGTCGCCGACT	N/A
P28	Hfx-RpaC-R500	CGTCCGAGAGGCCGAGCGCG	N/A
P29	Hfx-RfcA-R500	CCGCGCAGCGCGACTGGATG	N/A
P30	Hfx-RfcB-R750	GGGCGTCTCGTCCACGTCGT	N/A
P31	Hfx-RfcC-R750	ACGCGGAGGATGTCAGCCAG	N/A
P32	Hfx-Mcm-R500	GGTCGAAGTCGATGCGGAAC	N/A
P33	Hfx-PcnA-R750	GCCGAGTTCGACGGTGACCT	N/A
P34	Hfx-LigA-R750	GTGCCGGCCTGTGCCAGCAT	N/A
Construction of plasmids for constitutive high-level expression			

P35	Hsa-Fdx-5’	GGTGGTGGTTGGATCCGTCGACGGCCGGCAGCACCTGAAC	BamHI
P36	Hsa-Fdx-3’	GGTTGGTTGGCATATGTGCAGAGTTCGGCTTCCGAACGCA	NdeI
Construction and screening of a stable p*tna*- strains			
P37	HfxPolB- ForBcl	AGGAAGAGGATGATCACGTCCTCGCCGGGGATGCCGACGC	BclI
P38	HfxPolB-RevBcl	GGAGAAGAAGTGATCACACCCGGAACCACGGAAAGACATT	BclI
P39	HfxRpaC-ForBcl	AAGAGGAAGATGATCAGACGGCGACGGGGACGTCCCGCCG	BclI
P40	HfxRpaC-RevBcl	AGGAAGAAGATGATCAACAGGGAAGGGCGGCGCTTAAGCG	BclI
P41	HfxPriS-ForBcl	AGGAAGAGAATGATCATGGTCGAGCGCCGATAGCGCCTCC	BclI
P42	HfxPriS-RevBcl	AAGGAGAAGGTGATCACGTGGTCGTTTCTCTCACTACCGT	BclI
P43	Hfx650upPolBfor	ACGACGCTCGCCCAGAACAG	N/A
P44	Hfx-650uRpaCfor	GATTCACGTCGGGCGAGAAC	N/A
P45	Hfx-650upPriSfor	CGTGAACTGGTACGACGACATC	N/A
P46	HfxEcoRV220rev	GCTGCGCAACTGTTGGGAAG	N/A
P47	Hfx-PolB-R300	ATGAGCAGGCGGTTCGGAAAG	N/A
P48	Hfx-RpaC-R300	GCCGAAAATCTTCGTCAGCTCC	N/A
P49	Hfx-PriS-R300	ACACCGCCTGCATCTCCTC	N/A
P50	Hfx-PriS-Bcl20for	CGAGCGCCGATAGCGCCTCC	N/A
Construction of plasmids for truncated/mutated *tna* promoter			
P51	pNPM2- 5Bcl	AGAAGGAGAGTGATCAAGAACGGAACAGCCGGCGA	BclI
P52	pNPM3- 5Bcl	GAGAGGAGAATGATCAAGCCGGCGACACCGATGCA	BclI
P53	pNPM4- 5Bcl	AGAGGAAGAGTGATCAACCGATGCACACACCAGTC	BclI
P54	pNPM5- 5Bcl	GAGGAAGAGATGATCAACACCAGTCCACGAGCGCC	BclI
P55	Hfx-PolB- R200	CTCGTCGCGGCCCGCGCCCT	N/A
P56	Hfx-PolD1- R200	CGGTTCCCGACGGGTCCGCG	N/A
P57	Hfx-MCM- R200	GCGCCTCCTCCGCGTACTCC	N/A
P58	Hfx-L11-for	TTATAATATATGATCAGGTACCGACTTCGACTA	BclI
P59	Hfx-L11-rev	ATAATATTAACATATGCGCAATAGGTCCGCGAATGT	NdeI
P60	pNPM 1511 for	GCGAACGACCTACACCGAACTGAG	N/A
P61	pNPM 2284 rev	GCGTAACCACCACACCCGCC	N/A
P62	pNMP M1 for	TTTCCGCCTGCCGGTTACTTCACATTCGC	N/A
P63	pNPM M1 rev	GCGAATGTGAAGTAACCGGCAGGCGGAAA	N/A
P64	pNPM M2 for	TTTCCGCCTGCCGACTACTTCACATTCGC	N/A
P65	pNPM M2 rev	GCGAATGTGAAGTAGTCGGCAGGCGGAAA	N/A
P66	pNPM M3 for	TTTCCGCCTGCCGATGACTTCACATTCGC	N/A
P67	pNPM M3 rev	GCGAATGTGAAGTCATCGGCAGGCGGAAA	N/A
qRT-PCR			
P68	Hfx-24PolB-for	CGACGAGTATGTCTGGTTCGGC	N/A
P69	Hfx-24-PolB-rev	GTGCTGCACCATCTTCGCTGTCG	N/A
P70	Hfx-24-PolD1-for	GAGCGGCCACTGGATTATCGACC	N/A
P71	Hfx-PolD1-rev	GCCCGCGATGTCCTTGTCCTTCAT	N/A
P72	Hfx2-PolD2-for	CCGCGGCGGCATGTGTCTCGTCAT	N/A

P73	Hfx3-PolD2-rev	GTGCCGTCGATGAGGTCTTGGAGCC	N/A
P74	Hfx-24-RpaC-for	TCGGCCTCTCGGACGTGAACCTCG	N/A
P75	Hfx-24-RpaC-rev	CGGTGGCCGTCGTCGCGGTCGAAC	N/A
P76	Hfx-50S-for	GACAGGACGGCTCCCAGAA	N/A
P77	Hfx-50S-rev	CGCAGATCCAGTTGAGGTCG	N/A
Construction of plasmids for rpa genes deletions			
P78	RpaA-up-5’	GAGAGGAGGAGAATTCTGGCAATTCGTCCGGCCGGTCGCC	EcoRI
P79	RpaA-up-3’	AGGAGGAGGAGGATCCCGAGTCGGGATTCGCCCCCGTCGG	BamHI
P80	RpaA-dw-5’	GAAGAGAGAGGGATCCCTCGGGTCGTCGCGGTGGTCGCGG	BamHI
P81	RpaA-dw-3’	GAAGAGAGGAACTAGTATCGGCACCGCGGCGGCGGCCGCC	SpeI
P82	RpaB-up-5’	AGAGGGAAGAGAATTCGGTGGTGCCACAGCCTCAAAACCG	EcoRI
P83	RpaB-up-3’	AGGAGAGAGAGGATCCGGTCTACCTCGCCTCCACCTTGGT	BamHI
P84	RpaB-dw-5’	AAGAGAGGGAGGATCCGCGGGCGTCTCGCTCCGTTTCGTT	BamHI
P85	RpaB-dw-3’	AGGAGAGAGAACTAGTGTTTTCGACCTGCTCGCGCAGGTA	SpeI
P86	RpaC-up-5’	AGAGAAGAGAGGATCCGACGGATAGACCGCCGGTCGCCGG	BamHI
P87	RpaC-up-3’	AAGAGAGGAGACTAGTAGGGCGATTCGAGCGTCCATCGCG	SpeI
P88	RpaC-dw-5’	AGGAGAGAGAGAATTCTGCTCGGCGGACGGCGGGGCGCTC	EcoRI
P89	RpaC-dw-3’	AGGAGAGAGAGGATCCCAGGCGTCACCTCCCGGAACGAAG	BamHI
Construction of plasmids for expression of full-length RpaC proteins			
P90	Hfx-rpaC-NdeI-5’	AAGGAGAAGACATATGATGGGCGTCATCCGGGAGGTCTAC	NdeI
P91	HfxrpaCHindIII3’	GAAGAGAAGGAAGCTTGTAGACCTCCCGGATGACGCCCAT	HindIII
P92	Hgm-rpaC-5’	GGTGGTGGTTCATATGGGTGCCATCGAGGACGTATACGAC	NdeI
P93	Hgm-rpaC-3’	GGTGGTGTGAAGCTTCCGGCTCACCGAACGGAGAGAAAAG	HindIII
P94	Hqr-rpaC-5’	GTTGTTGGTTCATATGAGTTCATTGACTGACATTTACGAC	NdeI
P95	Hqr-rpaC-3’	GGTGGTGGTTAAGCTTTAGATGGGTTGACAGCTATCCACA	HindIII
Construction of plasmids for expression of mutated RpaC proteins			
P96	Hgm-RpaC-Δ1F	GAAGTCGAGGGTATCGCCTCGCTCGGCCTTTCGGAC	N/A
P97	Hgm-RpaC-Δ1R	GTCCGAAAGGCCGAGCGAGGCGATACCCTCGACTTC	N/A
P98	Hgm-RpaC-Δ2F	TCGTGTCGAAGACCTCGAAACGGCCGATATCGCGTC	N/A
P99	Hgm-RpaC-Δ2R	GACGCGATATCGGCCGTTTCGAGGTCTTCGACACGA	N/A
P100	Hgm-RpaC-Δ3F	AACGTGGAGTACGTCCCCTCACCGGCAGGTGCCGCT	N/A
P101	Hgm-RpaC-Δ3R	AGCGGCACCTGCCGGTGAGGGGACGTACTCCACGTT	N/A
P102	Hgm-RpaC-Δ12F	GAAGTCGAGGGTATCGCCGAAACGGCCGATATCGCG	N/A
P103	Hgm-RpaC-Δ12R	CGCGATATCGGCCGTTTCGGCGATACCCTCGACTTC	N/A
P104	Hgm-RpaC-Δ23F	TATCGTGTCGAAGACCTCTCACCGGCAGGTGCCGCT	N/A
P105	Hgm-RpaC-Δ23R	AGCGGCACCTGCCGGTGAGAGGTCTTCGACACGATA	N/A
P106	Hgm-RpaCΔ123F	GAAGTCGAGGGTATCGCCTCACCGGCAGGTGCCGCT	N/A
P107	Hgm-RpaCΔ123R	AGCGGCACCTGCCGGTGAGGCGATACCCTCGACTTC	N/A
P108	Hgm-RpaCΔNTD	GGTTGGTTGGCATATGGACGATGTGAAGTTCCTC	N/A
P109	Hgm-RpaCTD-for	AATTGATAGGATCCG	N/A
P110	Hgm-RpaCTDrev	AATTCGGATCCTATC	N/A

Construction of plasmids for expression of Flag epitope-tagged RpaC proteins			
P111	Flag-NdeI-F	TATGGACTACAAGGACGACGACGACAAGCA	N/A
P112	Flag-NdeI-R	TATGCTTGTCGTCGTCGTCCTTGTAGTCCA	N/A
Diagnostic PCR for gene deletion			
P113	HfxRpaA-for	ACCAGCATGCCGAGGAGCTCGCCT	N/A
P114	HfxRpaA-R500	GCAGGTCGATGAGGTCGGCC	N/A
P115	HfxRpaB-for	TGCGGCTGAGATAGCCGACCAGTT	N/A
P116	HfxRpaB-R500	TGGTCGAGTCGTCGCCGACT	N/A
P117	HfxRpaC-for	ATGGGCGTCATCCGGGAGGTCTAC	N/A
P118	RpaCR500new	AGGTCCTCGACGCGGTAGCTATCG	N/A
P119	HfxTrpA-for	TCGGCGTGGTCGCCG	N/A
P120	HfxTrpA-rev	CTTCCGTTCTCGGAG	N/A
P121	HfxHdrB-for	CGTGGCACCTCCCCG	N/A
P122	HfxHdrB-rev	GTCGCCTTCGTGTGC	N/A
Construction of plasmids for protein expression			
P123	CCBD- for	AAGCCTGACCATATGGATATCGCTAGCGCAAATACACCGGTATCAGG	NdeI, EcoRVNheI
P124	CCBD- rev	TGAGAACGAAGGTACCTTACTATACTACACTGCCACCGGGTT	KpnI
P125	Hfx-FLCCBD for	AAGGAGAAGACATATGGGCGTCATCCGGGAGGTCTAC	NdeI
P126	HfxFL-CCBD-rev	GAAGAGAAGAGCTAGCCTGCTGTGCGCCAGTGTGGAC	EcoRV
P127	HfxNTDCCBDre	TTGAGCGAGAGCTAGCGCCGGGTTCGATGTCGGCGA	NheI
P128	NCBD- for	GCATGACTACCATATGGCAAATACACCGGTATCAGG	NdeI
P129	NCBD- rev	GCATCTAGATGATATCTACTACACTGCCACCGGGTT	EcoRV
P130	PriS-CTDCBDfor	GTACGTACGTGATATCGACATCAGCGTGAACGTGAC	EcoRV
P131	PriS-CTDCBDrev	GTACTACGTAGGTACCTCATTCTGTCACCTTCTCGG	KpnI
P132	CR-for	AGGGAAGAAAGCGAAAGGAGCG	N/A
P133	CR-rev	AGAGCGCCCAATACGCAAACC	N/A
Construction of plasmids for PriS-GINS complementation assay			
P134	Hfx-Pris1021-for	GACATCAGCGTGAACGTGACCGAC	N/A
P135	Hfx-Pris1320-rev	GTAGTGGGTGCCGAGCTGTTCGGC	N/A
P136	Hfx-PrisNdeI-for	AAGAGGAAGGCATATGATGGACGGGCGCACACGCGAGTAC	NdeI
P137	PriS-Dde-for	GGAGAGAGAACTCAGGAGCGTGCTTCGGACGGAGCGACAG	DdeI
P138	PriS-Dde-rev	AGGAAGAGAACTCAGGTCGTCCACGTTCATTCATTCTGT	DdeI
P139	HfxGinS-for	AGAGGAGAAGCATATGATGAACGTGGACGACCTCAGGAGC	NdeI
P140	HfxGinS-rev	GGAAGAGAAGAAGCTTCTGTTCAGGCGGCGGAAAGAAGGG	HindIII
P141	HboPriS-5Nde	GTGTGGTGGTCATATGGAGGCGCACACCCGGGA	NdeI
P142	HboPriS-3H	GGTTGGTTGGAAGCTTCTGTCGCTCTGTCCGCCGGA	HindIII
P143	HboGinS-5Nde	GGTTGGGTTTCATATGAATTTAGACGAGTTACG	NdeI
P144	HboGinS-3H	GGTTGGTTGGAAGCTTCAGTCGGCCGAGGCTTGCC	HindIII

Table A2 SSB-encoding genes in completely sequences haloarchaeal genomes

	RpaA1, RpaA2 and Rpe ORFs			RpaB1, RpaB2 and Ral ORFs			RpaC ORF
Species	RpaA1	RpaA2	Rpe	RpaB1	RpaB2	Ral	RpaC
Halobacterium NRC-1	ID:1447913	ID:1447912	ID:1447911	ID:1448599	ID:1448600	ID:1448601	ID:1447061
Haloarcula marismortui	ID:3128313	ID:3128235	ID:3127962	ID:3126669	ID:3126671	ID:3126672	ID:3130475
Natronomonas pharaonis	ID:3701922	ID:3701923	ID:3701924	ID:3701099	ID:3701098	ID:3701097	ID:3703116
Haloquadratum walsbyi	ID:4194596	ID:4192923	ID:4192922	ID:4193367	ID:4193368	ID:4193369	ID:4194595
Haloferax volcanii	ID:8924887	ID:8925705	ID:8926148	ID:8924234	ID:892568	ID:8924278	ID:8926823
Halorubrum lacusprofundi	ID:7401758	ID:7401759	ID:7401760	ID:7401657	ID:7401658	ID:7401659	ID:7401631
Halomicrobium mukohataei	ID:8410544	ID:8410545	ID:8410546	ID:8410544	ID:8410545	ID:8410546	ID:8410070
Halorhabdus utahensis	ID:8382811	ID:8382812	ID:8382813	ID:8383908	ID:8383909	ID:8383910	ID:8384542
Halogeometricum borinquense	ID:9993716	ID:9993717	ID:9993718	ID:9994699	ID:9994700	ID:9994701	ID:9994501
Haloterrigena turkmenica	ID:8745744	ID:8745743	ID:8745742	ID:8741033	ID:8741032	ID:8741031	ID:8740984
Halopiger xanaduensis	ID:10797641	ID:10797642	ID:10797643	ID:10796754	ID:10796755	ID:10796758	ID:10799284
Natrialba magadii	ID:8825942	ID:8825943	ID:8825944	ID:8825069	ID:8825070	ID:8825079	ID:8825052

Halorhabdus tiamatea	HLRTI_09532	HLRTI_09527	HLRTI_09522	HLRTI_08850	HLRTI_08855	HLRTI_08860	HLRTI_17725
Halalkalicoccus jeotgali	ID:9419694	ID:9419693	ID:9419692	ID:9418159	ID:9418158	ID:9418156	ID:9420762

Notes: The table shows Gene IDs for SSB genes in haloarchaeal genomes with the exception of Halorhabdus tiamatea SARL4B where Gene IDs are yet to be assigned and locus tags are shown instead (see Table A3 for protein information). Orange and blue shaded boxes indicate adjacent ORFs encoding RpaA1, RpaA2 and Rpe proteins or RpaB1, RpaB2 and Ral proteins, respectively. Brown shaded boxes are Ral-encoding ORFs that are separated from RpaB2 by on to eight unrelated ORFs (or predicted ORFs).

Table A3 SSB-encoding proteins in fourteen completely sequenced haloarchaeal genomes

	RpaA1 and RpaA2 ORFs		RpaB1and RpaB2 ORFs		RpaC ORF
Species	RpaA1	RpaA2	RpaB1	RpaB2	RpaC
Halobacterium NRC-1	GI:15790309 226 4e-71 34% 52%	GI:15790308 172 1e-40 41% 63%	GI:15790993 353 8e-69 49% 67%	GI:15790994 204 7e-70 54% 71%	GI:15789450 363 8e-63 24% 43%
Haloarcula marismortui	GI:55379364 134 7e-27 34% 50%	GI:55379365 140 2e-42 45% 64%	GI:55376151 349 8e-72 51% 69%	GI:55376152 207 4e-70 53% 68%	GI:55378982 363 3e-65 39% 60%
Natronomonas pharaonis	GI:76802614 239 8e-77 32% 54%	GI:76802615 204 8e-48 43% 62%	GI:76801292 350 6e-70 49% 69%	GI:76801291 207 1e-70 54% 69%	GI:76800811 410 2e-65 40% 58%
Haloquadratum walsbyi	GI:110668327 214 3e-67 33% 53%	GI:110668326 178 5e-43 39% 55%	GI:110669144 346 7e-75 51% 70%	GI:110669145 201 1e-68 53% 70%	GI:110667397 512 1e-70 35% 58%
Haloferax volcanii	GI:292655492 252 9e-82 36% 55%	GI:292655491 194 7e-47 41% 59%	GI:292654472 379 4e-78 54% 72%	GI:292654471 222 1e-76 57% 73%	GI:292654691 228 3e-71 38% 60%

Halorubrum lacusprofundi	GI:222479058 218 2e-68 34% 52%	GI:222479059 194 3e-46	GI:222478575 355 8e-73 49% 70%	GI:222478576 210 1e-71 56% 69%	GI:222478550 358 1e-67 33% 58%
Halomicrobium mukohataei	GI:257387088 149 6e-28 33% 50%	GI:257387089 139 4e-42 44% 64%	GI:257387088 372 5e-77 52% 73%	GI:257387089 209 2e-71 55% 72%	GI:257386633 353 2e-59 37% 59%
Halorhabdus utahensis	GI:257051631 254 6e-79 35% 53%	GI:257051632 194 2e-47 45% 61%	GI:257052703 369 6e-75 49% 70%	GI:257052704 201 3e-68 52% 68%	GI:257053314 378 7e-66 36% 59%
Halogeometricum borinquense	GI:313126639 237 6e-76 35% 53%	GI:313126640 187 5e-45 42% 62%	GI:313127597 375 5e-79 54% 72%	GI:313127598 217 2e-74 57% 74%	GI:313127408 439 1e-74 38% 60%
Haloterrigena turkmenica	GI:284176223 239 9e-77 33% 54%	GI:284176222 204 8e-46 47% 64%	GI:284163751 370 1e-72 49% 69%	GI:284163750 213 8e-73 54% 70%	GI:284163709 229 3e-71 41% 63%
Halopiger xanaduensis	GI:336254076 247 7e-80 33% 54%	GI:336254077 207 3e-51 42% 61%	GI:336253189 339 7e-65 46% 77%	GI:336253190 207 2e-70 52% 68%	GI:336253183 401 4e-73 38% 60%

Natrialba magadii	GI:289582730 246 3e-79 33% 53%	GI:289582731 216 6e-49 47% 64%	GI:289581883 344 6e-67 42% 66%	GI:289581884 217 2e-72 54% 69%	GI:289581867 404 5e-72 38% 59%
Halorhabdus tiamatea	GI:335437337 234 6e-75 34% 53%	GI:335437336 194 4e-47 45% 61%	GI:334896332 368 9e-76 49% 71%	GI:334896333 199 1e-67 51% 68%	GI:335440024 385 7e-67 38% 59%
Halalkalicoccus jeotgali	GI:300711246 236 9e-76 34% 54%	GI:300711245 162 6e-35 45% 61%	GI:300709731 360 3e-72 50% 69%	GI:300709730 214 2e-73 54% 75%	GI:300712301 383 4e-66 36% 56%

Data in each table cell (from top to bottom): accession number (GenBank identifier, GI), BLAST score, E value, protein sequence identity and protein sequence similarity to M.acetivorans C2A query sequences. Query sequences: MA0590 (MacRPA2, GI:20089479, RpaA1 orthologue), MA0591 (GI:20089480, RpaA2 orthologue), MA3019 (MacRPA3, GI:20091837, RpaB1 orthologue), MA3018 (GI:20091836, RpaB2 orthologue), MA4645 (MacRPA1, GI:20093424, RpaC orthologue). BLAST parameters: algorithm Blastp (protein-protein BLAST); matrix – BLOSUM62; gap costs – existence 11, extension 1; database – non-redundant protein sequences.

Figure A1 Multiple sequence alignment of RpaC proteins from fourteen haloarchaeal species