An Examination of Novel Archaeal Stress Response Genes

The University of Waikato
Te Whare Wānanga o Waikato

A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy at The University of Waikato by

Marcel Eduard Dinger

March 2003

Abstract

The regulation of gene expression is the fundamental system employed by cells and organisms to respond to changes in their surroundings. To further understand the origins and nature of the processes underlying gene regulatory systems, the stress response in the thermophilic archaeon, *Thermococcus zilligii*, was examined. Microorganisms respond to stress by triggering physiological and morphological changes, which are largely brought about by adaptive changes in gene expression. The use of a thermophilic archaeon in this research is significant, since this group of organisms is widely regarded to resemble ancient life more closely than other extant life on Earth. Moreover, since the informational processing systems in archaea are thought to be closely related to those of the eukaryal nucleus, the study of the archaeal stress-response may further the understanding of the evolutionary relationship between the Archaea and the Eucarya. The stress response of *T. zilligii* was examined by analysing the total cytoplasmic protein complements from cultures ranging from mid-log to late stationary phase. Using N-terminal sequencing and BLASTP searches, homologues of four low molecular weight proteins that varied in level were identified in the genome sequences of members of the closely related genus, *Pyrococcus*. The genes originated from four distinct families, which were named *srb*, *src*, *srd* and *sre*.

The *srb* family represents an essentially undescribed group of genes (COG1433), that occurs primarily amongst members of the Euryarchaeota. Due to the sporadic distribution of *srb* genes amongst extant life, the *srb* genes are suggested to have undergone horizontal transfer across large phylogenetic distances on several occasions. A sequence homologous to *srb* family members was identified as part of a hypothetical ORF in *Clostridium perfringens*, which encodes a protein that consists of a fusion of a transcriptional regulator belonging to the ArsR family and an SRB homologue. Several *srb* genes were found to occur as part of putative operons, the majority of which included genes homologous to the MinD family of P-loop ATPases.

The *src* family encodes the highly conserved Sac10b family of DNA-binding proteins (COG1581). *src* genes occur exclusively amongst archaeal genomes, but were absent in *Halobacterium* sp. NRC-1, *Methanopyrus kandleri*, *Methanosarcina mazei* and *M. acetivorans*. The consensus secondary structure of the SRC proteins consists of a relatively

long α-helix flanked by two small β-strands, followed by a second α-helix, and then three further β-strands.

The *srd* family encodes proteins that are homologous to the C-terminal half of members of the Lrp/AsnC family of transcriptional regulators (COG1522). It is suggested here that the SRD family forms a logical subgroup that is structurally, and presumably functionally, distinct to the other members of the Lrp/AsnC family. *srd* genes occur in all non-methanogenic archaea and in some bacteria. It is suggested that the *srd* family originated in the archaeal lineage and was later transferred to members of the Bacteria. The *srd* genes often appear as part of operon-like arrangements, which seem to be unrelated to one another, and consequently it is suggested that the SRD proteins are involved in a relatively generic function, such as transcriptional regulation.

The *sre* family represents an undescribed group of highly conserved genes (COG1698) that occur exclusively amongst archaeal genomes. One *sre* gene was found in every archaeal genome sequenced to date, except for *Pyrobaculum aerophilum* and *Halobacterium* sp. NRC-1. Comparison of the SRE and 16S rRNA distance trees suggests that the *sre* genes were not prone to horizontal transfer. The secondary structure predicted from the SRE sequences consists of four large helical regions separated by short coiled regions.

Recombinant proteins encoded by the *srbPab1*, *srbPab2*, *srdPab2* and *srePab* genes cloned from *Pyrococcus abyssi* were produced in *Escherichia coli* and purified. An examination of the predicted quaternary associations of the recombinant proteins revealed that each of the proteins, with the exception rSRBPab2, formed multimeric structures. rSRBPab1 formed dimer-sized multimers that occurred at 80°C, but not at room temperature. rSRDPab2 formed dimer-sized molecules that occurred preferentially at 80°C and in the presence of DNA. rSREPab formed structures that were consistent in size with those predicted for dimers and tetramers. Neither rSRBPab1, rSRDPab2 or rSREPab bound to DNA in the conditions examined.

Acknowledgements

There are a number of people that I would like to acknowledge for their contributions to this study. Primarily, I wish to thank my supervisor, Dr. David Musgrave, for his intellectual and critical input, as well as his continued enthusiasm, support, reliability and patience throughout this research. Second, I wish to extend my sincerest gratitude to Ms. Raewyn Towers for always finding a way of getting everything that I needed and for making the lab a safe, yet enjoyable place to work. Third, I thank Dr. Ray Cursons for his interest and many helpful technical tips, insights and ideas. I also wish to acknowledge Peter Charlton for useful discussions and technical help, Bridget Douglas for proofreading and collating, and Prof. Patrick Forterre (Université Paris-Sud, France) for the *Pyrococcus abyssi* culture and access to preliminary genomic sequence data.

I also wish to thank the rest of my lab and inter-lab colleagues over the years for their friendship, support and memorable moments: Dick Wilkins, Greg Jacobson, Greg Baillie, Mark Walker, Mark Barber, Debbie Steel, Gethin Morgan, Andrew McNaughton, Tao Xie, Ivy Zhang, Linda Peters, Jo Haines, Quanah Hudson, Mahinda Gunatillake, Angela van Diepen, Rachel Dunn, Ron Ronimus, Dion Thompson, Jason McGrath, Jason Ward and Cameron Evans. Outside of the lab, I also wish to acknowledge, for all manner of reasons, Ben Albahari, Maarten Dinger, Andy Haigh, and Gavin Scandlyn.

Financially, I gratefully acknowledge the support of The University of Waikato for provision of a scholarship over a period of this study. I also wish to thank the Department of Biological Sciences for a generous Student Travel Award, which allowed me to attend the Thermophiles conference in Brest, France.

Finally, I wish to express my sincere "dankbaarheid" to my parents for aiding in the supply of many of life's essentials, such as een huis (met rozen en tulpen), kleren, lekker eeten, wijn, bier, kaas, chocolade, koffie, en stroopwafels.

Table of Contents

APPENDICES 125

REFERENCES 158

List of Figures

List of Tables

List of Abbreviations

aa	amino acid(s)
APS	ammonium persulfate
ATP	adenosine 5'triphosphate
AWE	acid-wash extract
bp	base pair(s)
BLAST	basic local alignment search tool
BSA	bovine serum albumin
CHAPS	3-[(3-Cholamidopropyl) dimethylammonio]-1-propanesulfonate
COG	clusters of orthologous groups (of proteins)
DNA	deoxyribonucleic acid
DTT	dithiothreitol
EDTA	ethylenediamine tetra-acetic acid
EMSA	electrophoretic mobility shift assay
ERMA	Environmental Risk Management Authority
HEPES	N-2-hydroxyehtylpiperazine-N'-2-ethanesulfonic acid
IPTG	isopropyl-β-D-thiogalactopyranoside
LB	Luria-Bertani
LUCA	last universal common ancestor
MW	molecular weight
MWCO	molecular weight cut-off
NCBI	National Center for Biotechnology Information
OD	optical density
ORF	open reading frame
PAGE	polyacrylamide gel electrophoresis
PBS	phosphate-buffered saline
PCR	polymerase chain reaction
PIPES	piperazine-N,N'-bis-(ethanesulfonic acid)
PMSF	phenylmethylsulfonyl fluoride
PVDF	polyvinylidene fluoride
RNA	ribonucleic acid
rpm	revolutions per minute
SDS	sodium dodecyl sulfate
TEMED	N,N,N',N'-tetramethyl-ethylenediamine
TP	total cytoplasmic proteins
Tris	tris(hydroxymethyl)amino-methane
UV	ultraviolet
v/v	volume to volume
w/v	weight to volume

Preface

Life on Earth is thought to have started evolving approximately 4 billion years ago, about 500 million years after the Earth was formed (Ernst 1983; Walter 1983). Since the discovery of ribozymes (Guerrier-Takada *et al.* 1983), RNA molecules capable of catalysing chemical reactions, the most widely held view is that life evolved from an RNA world (Gilbert 1986; Joyce 1989). The resulting primitive RNA-based life forms are hypothesised to have given rise to DNA-based life (Benner and Ellington 1987; Benner *et al.* 1989), which ultimately lead to the inception of the last universal common ancestor (LUCA) of all life on Earth today (Woese 1987; Woese 1998).

The oldest morphological evidence for life on Earth dates back approximately 3.5 billion years (Schopf and Packer 1987; Schopf 1993). These remarkably preserved microfossils indicate the existence of microbial life, particularly cyanobacteria, that are thought to have already been widespread by this time (Schopf *et al.* 2002). The cyanobacteria, with their capacity to harness the energy of the sun to produce organic matter and oxygen, were responsible for the critical changes in the Earth's atmosphere that set the stage for the evolution of a tremendous diversity of new life-forms (Mojzsis *et al.* 1996; Brocks *et al.* 1999). However, prior to the emergence of the cyanobacteria, methanogenic microorganisms or methanogens, are also thought to have played an important role in modifying the Earth's atmosphere to allow life to proliferate (Kasting and Siefert 2002). It is believed that the production of methane by methanogens caused a greenhouse effect on the primitive Earth that was required to prevent the oceans of the dimly lit planet from freezing, and consequently maintain an environment in which life could further evolve (Kasting and Siefert 2002).

Today, life on Earth is divided into three primary domains, the Bacteria, the Eucarya, and the Archaea (Woese *et al.* 1990). The realisation that all known extant organisms can be organised into these three distinct groups, originated from the examination of the small subunit (16S) ribosomal RNA (rRNA) sequences (Woese and Olsen 1986; Woese 1987). The comparison of the 16S rRNA sequences provided a systematic approach to determining the phylogenetic relationship between all living organisms, and created an entirely new perspective on the universal ancestor (Woese 1987). However, the rooting of the universal tree of life, that is determining the nature and order in which the bifurcations

of the three domains occurred, remains a matter of considerable controversy and speculation (Forterre and Philippe 1999; Lopez-Garcia *et al.* 1999; Caetano-Anolles 2002; Cavalier-Smith 2002). Similarly, the characteristics of the universal common ancestor itself are also a subject of continued debate (Doolittle 2000), such as whether it was DNA- or RNA-based (Poole *et al.* 1998; Forterre 2001), mesophilic, thermophilic or hyperthermophilic (Penny and Poole 1999; Glansdorff 2000; Moulton *et al.* 2000; Brochier and Philippe 2002), and how complex it was compared to microorganisms today (Forterre and Philippe 1999; Kyrpides *et al.* 1999; Woese 2000; Castresana 2001; Woese 2002).

Since the proposition of the three-domain classification of life on Earth, considerable physiological, biochemical, and genetic evidence has continued to accumulate that support the distinctiveness of the three groups (Olsen and Woese 1996; Fitz-Gibbon and House 1999; Snel *et al.* 1999; Graham *et al.* 2000). From a molecular viewpoint, one of the most interesting findings concerning the characteristics of the three groups, was the realisation that the transcriptional apparatus of the Eucarya and the Archaea are closely related and distinct to that in the Bacteria (Zillig *et al.* 1979; Langer *et al.* 1995; Thomm 1996; Soppa 1999). Moreover, the fundamental proteins involved in chromosomal organisation in the Eucarya, histones, were also found to occur and exhibit a similar role in members of a major subgroup of the Archaea, the Euryarchaeota (Sandman *et al.* 1990; Pereira *et al.* 1997). These discoveries have reignited speculations concerning the evolutionary relationship between the Eucarya and the Archaea, and the origins of the eukaryal nucleus (Lake and Rivera 1994; Martin and Muller 1998; Sandman and Reeve 1998; Horiike *et al.* 2001; Hartman and Fedorov 2002).

The fundamental determinant in the regulation of gene expression relates to the controlled interaction of the transcriptional machinery with the DNA and its associated chromosomal proteins (Urnov and Wolffe 2001; Wolffe 2001). The regulation of gene expression is the central process that governs cellular differentiation in multicellular organisms, and consequently its understanding is important to numerous areas of biology and medicine. The availability of an homologous, though considerably simplified, transcriptional machinery and chromosomal architecture in some Archaea provides a unique opportunity to not only appreciate the functional essence of these systems, but also exploit them as a simplified model in understanding these complex structures and processes (Olsen and Woese 1997).

Although microorganisms usually do not undergo cellular differentiation like multicellular eukaryotes, they do undergo physiological and morphological changes in response to changes in the environment, such as osmolarity, pH, temperature and nutrient availability (Csonka 1989; Kolter *et al.* 1993; Mager and De Kruijff 1995). Much like eukaryotes, microorganisms coordinate the appropriate response required for the conditions through a diverse range of molecular machineries that ultimately regulate protein levels (Kolter *et al.* 1993; Emerson 2002). These processes can act upon virtually every stage in the process of protein production, including the transcriptional and translational efficiency, as well as RNA and protein turnover. In either microorganisms or multicellular eukaryotes, control at the transcriptional level can occur through a variety of different mechanisms, including modification of RNA polymerase (Hengge-Aronis 2002), chemical modification of eukaryal histones (Berger 2002), DNA methylation (Lewin 1999a; Attwood *et al.* 2002), and regulatory protein binding (Lewin 1999b).

Another perhaps more subtle mechanism for controlling gene expression is via the modification of the DNA topology itself (Wang and Syvanen 1992). Due to the double helical structure of DNA (Watson and Crick 1953), when its ends are unable to freely rotate, such as in a circular molecule or a chromosomal domain, any factor that either coils the DNA or alters its twist will introduce tension into the molecule that will be manifested as changes in the overall twist and writhe (Crick 1976; Bates and Maxwell 1993). Various intracellular factors can affect the topology of DNA, such as proteins that bend, wrap or otherwise contort DNA (Drlica and Rouvière-Yaniv 1987; Kornberg and Lorch 1992); topoisomerases, which are a special class of enzymes that can twist or untwist DNA by temporarily cutting one or both strands of DNA (Champoux 2001; Wang 2002); and the transcription process itself (Wang and Lynch 1993). Changes in DNA topology can also be brought about by a number of environmental parameters, such as temperature, pH and osmolarity (Depew and Wang 1975; Drlica 1992; Duguet 1993; Dorman 1996; Lòpez-Garcìa and Forterre 1999). In general, changes in DNA topology can affect gene expression in two distinct ways: (1) changes in twist can alter the presentation of a DNA regulatory sequence element; and (2) changes in writhe can alter the relative proximity of remote regulatory DNA sequence elements.

Despite the many similarities between the archaeal and eukaryal transcriptional machinery, the Archaea possess some regulatory factors that are also found in Bacteria (Bell and Jackson 1998; Aravind and Koonin 1999; Kyrpides and Ouzounis 1999) and a number of

other major DNA-binding proteins that are unique to certain archaeal families (Dijk and Reinhardt 1986; Chartier *et al.* 1988; Choli *et al.* 1988; Searcy *et al.* 1998). How these different machineries interact with the common archaeal transcriptional machinery is presently unknown, however, several groups are investigating these processes (Voorhorst *et al.* 1999; Bell *et al.* 2001; Thomsen *et al.* 2001). Moreover, since many families of the Archaea have only been recently discovered, the very presence of other regulatory proteins has not been investigated at all.

The objective of this thesis is to identify new families of regulatory and DNA architectural proteins in Archaea, particularly in those species which also contain archaeal histones. By obtaining a greater appreciation of the actual factors involved in the regulation of gene expression in Archaea, it is hoped that this research will broaden the landscape for the study of their eukaryal-like nucleoid systems. Ultimately, armed with sufficient biochemical and molecular evidence concerning the fundamental processes that drive extant life, it may be possible to recreate a model of some of the key events and processes that paved our most distant evolutionary past.

CHAPTER ONE

Identification of Low Molecular Weight Stress Response Proteins in *Thermococcus zilligii*

Summary

The microbial stress-response has been extensively studied in bacteria, and many stress-response proteins have been characterised. Although most of these proteins are widespread amongst members of the bacterial domain, few homologues have been identified in the complete genome sequences of members of the Archaea. Consequently, little is known about how archaea regulate gene expression. To identify potential regulatory proteins in members of the Archaea, the total cytoplasmic protein complements from cultures of varying age were compared from the thermophilic archaeon, *Thermococcus zilligii*. A total of 13 proteins were detected to vary in level with changing growth-phase. The N-terminal sequences of four different low-molecular weight proteins, named SRBTzi, SRCTzi, SRDTzi and SRETzi, which changed significantly in level with growth-phase, were determined. Using BLASTP searches at GenBank, homologues of the proteins were identified in the complete genome sequences of other members of the Thermococcales and other archaea. SRETzi was found to belong to an undescribed, highly conserved family of proteins found, exclusively, in many archaeal genomes. SRBTzi belonged to an undescribed protein family that occurs in most archaeal genomes, as well as in the bacterial genomes, *Thermotoga maritima* and *Clostridium acetobutylicum*. SRCTzi was identified as a member of the Sac10b family, which is a highly conserved family of archaeal-specific DNA-binding proteins. Finally, SRDTzi belonged to a group of hypothetical undescribed proteins, mostly archaeal in origin, that possess significant sequence homology to the C-terminal region of the leucine-responsive regulatory proteins (Lrps).

1.1 Introduction

In their natural environment, microorganisms must endure multiple stresses, including changes in temperature, osmolarity and nutrient availability (Kolter *et al.* 1993). Microorganisms respond to such stresses by undergoing adaptive changes in gene expression, which result in the subsequent physiological and morphological differentiation of the cells (Hengge-Aronis 1996b).

The regulation of gene expression has been studied extensively amongst members of the Bacteria, and many proteins involved in regulatory systems and chromosomal organisation have been identified. The induction and repression of transcription of particular genes with changing growth phase has been attributed to the effects of RNA polymerase sigma factors (Hengge-Aronis 1996a), DNA- and RNA-binding proteins (Hengge-Aronis 2002), and environmental factors that affect DNA topology (Drlica 1992; Lòpez-Garcìa and Forterre 1997; 1999). Although various sigma factors and many DNA- and RNA-binding proteins are relatively ubiquitous in members of the Bacteria, the majority have no homologues in the archaeal genomes sequenced so far. This is most likely due to the considerable differences in the basal transcriptional apparatus of bacteria compared to archaea. Indeed, the transcriptional apparatus in archaea has been suggested to more closely resemble that of eukaryotes (Langer *et al.* 1995; Leigh 1999). Archaea feature TATA box-like promoter sequences, transcription factors including TATA binding protein, TFIIB and TFIIS, and RNA polymerases similar to eukaryotes (Rowlands *et al.* 1994; Soppa 1999).

In any cell, whether it be archaeal, bacterial or eukaryal, the DNA must be extremely well-compacted in order to fit within it. Indeed, it has been calculated that for *Escherichia coli* to fit its genome inside its cell, the volume of the DNA must be compacted by about 1000-fold (Schmid 1990). With the continuing advancement in the understanding of the regulation of gene expression, it is perhaps not surprising that it has become increasingly apparent that an organism's DNA-packaging system is also intimately involved with its transcriptional regulatory systems (D'Ari *et al.* 1993; Loewen and Hengge-Aronis 1994; McGovern *et al.* 1994; Pereira *et al.* 1997; Wolffe 2001). DNA-packaging systems can influence the regulation of gene expression in a number of different ways, including; (a) changing the topological state of the DNA such that the presentation of recognition sequences to regulatory proteins is altered; (b) modulating the architecture of the DNA to effectively localise remote regulatory sequence elements; (c) directly obscuring or limiting accessibility to regulatory sequences by regulatory proteins; or (d) positively mediating the

interaction of regulatory proteins to enhance transcription (Drlica 1992; Kornberg and Lorch 1992; Wang and Syvanen 1992; Wang and Lynch 1993). Consequently, the study of abundant DNA-binding proteins is an important aspect in understanding and modelling the regulation of transcriptional control systems. A number of major DNA-binding proteins, which are thought to be involved in chromosomal organisation in some manner, have been identified amongst members of both the Archaea and the Bacteria.

To date, the major groups of DNA-binding proteins that have been implicated in chromosomal organisation in various members of both the Archaea and the Bacteria are the HU, H-NS, FIS, Dps, Lrp, Sac7, Sac10b, MC1, HTa, and archaeal histone families. A brief summary of each of these families of DNA-binding proteins follows.

HU. The HU (heat-unstable nucleoid protein) and closely-related IHF (integration host factor) proteins are ubiquitous amongst members of the Bacteria, and to date, the only archaeal genomes found to contain genes encoding HU are those from members of the Thermoplasmales. HU occurs as both a homodimer or heterodimer (encoded by *hupA* and *hupB*), with a molecular weight of approximately 20 kDa, and binds RNA or DNA (single- or double-stranded) with little or no sequence specificity (Bonnefoy and Rouviere-Yaniv 1991). There are an estimated 60,000 copies of HU per *E. coli* cell, and approximately one dimer of HU per 200 bp of DNA (Hayat and Mancarella 1995). Like eukaryal histones, HU preferentially binds negatively supercoiled DNA, and in the presence of topoisomerase I, HU is able to introduce negative supercoils in relaxed plasmid DNA (Le Hégarat *et al.* 1993). Furthermore, HU condenses DNA and forms nucleosome-like particles, consisting of 8-10 dimers of HU and 275 bp of DNA (Broyles and Pettijohn 1986). In addition to compacting DNA, HU also participates in several other molecular events, such as the initiation of DNA replication (Dixon and Kornberg 1984; Bahloul *et al.* 2001), translation (Balandina *et al.* 2001), transcriptional regulation (Aki and Adhya 1997), transposition (Lavoie and Chaconas 1993), and recombination (Kamashev and Rouviere-Yaniv 2000).

H-NS. The H-NS (histone-like nucleoid-structuring) proteins, and their closely related homologues, the StpA family, are neutral DNA-binding proteins with a molecular mass of approximately 15 kDa. Genes encoding H-NS have been identified in all sequenced genomes of enterobacteria, as well as some other members of the Proteobacteria, such as *Xylella fastidiosa* and *Vibrio cholorae*. H-NS proteins bind double stranded DNA in a relatively non-specific fashion, showing only a preference for intrinsically curved DNA regions (Hulton *et al.* 1990). Although H-NS does not wrap DNA, H-NS binding does

result in significant compaction of DNA and may be involved indirectly in the control of DNA topology (Owen-Hughes *et al.* 1992). In *E. coli*, the overproduction of H-NS results in the strong compaction of the chromosomal DNA, affects the transcription of many genes, and is detrimental to cell growth (McGovern *et al.* 1994). Mutation of the gene encoding H-NS, *hns*, is also detrimental for cell growth, and affects the production of many proteins (Bertin *et al.* 1990). Recent findings suggest that H-NS and StpA stimulate the expression of stringently regulated genes by their effect on local DNA topology (Johansson *et al.* 2000).

FIS. FIS (factor for inversion stimulation) is a nucleoid-associated protein dimer, with a molecular weight of about 24 kDa, which occurs at a frequency dependent on the growth phase of the cell, ranging from 25,000 molecules per cell in stationary-phase to 100,000 in log-phase (Ali Azam *et al.* 1999). In exponentially growing *E. coli*, FIS is thought to be the most abundant nucleoid-associated protein (Ishihama 1999). FIS regulates recombination and replication and is involved in both enhancing and inhibiting gene expression (Xu and Johnson 1995). Moreover, FIS has been implicated in coupling DNA topology with the physiological changes that occur in stationary phase (Schneider *et al.* 1999). FIS binds DNA with a relatively low sequence specificity, and upon binding results in significant bending and deformation of the DNA (Pan *et al.* 1996). Beyond the effect of FIS interaction with specific promoters and regulatory elements, FIS is suggested to influence global DNA architecture by its controlled stabilisation of writhe, which results in the formation of particular loops and branching in DNA that in turn influences gene expression (Muskhelishvili and Travers 1997; Schneider *et al.* 2001).

Dps. Dps (DNA-binding protein from starved cells) was initially identified as a starvation-inducible DNA-binding protein, with a molecular mass of approximately 19 kDa, in *E. coli* (Almirón *et al.* 1992). Genes encoding Dps have since been identified in virtually every bacterial genome, as well as the archaeal genome of *Halobacterium* sp. NRC-1. In *E .coli*, Dps levels steadily accumulate from 6000 molecules per cell in log-phase to approximately 200,000 molecules per cell in stationary phase (after 48 h growth), making it the most abundant DNA-binding protein in the cell (Ali Azam *et al.* 1999). Dps has been shown to co-crystallise with DNA, to form a highly compacted structure that is able to protect the DNA from various environmental assaults (Grant *et al.* 1998; Wolf *et al.* 1999). Apart from compacting DNA in stationary phase, Dps has also been shown to provide a direct

role in protection of DNA against the toxic and mutagenic effects of oxidative damage in metabolically active cells (Martinèz and Kolter 1997).

Lrp. Lrp (leucine-responsive regulatory protein) is a small basic protein composed of two identical 19 kDa subunits, which was initially discovered for its role in the regulation of several leucine-regulated genes (Lin *et al.* 1990). Lrp is now regarded as a global transcription factor that has been found to regulate, positively or negatively, at least 75 genes in *E. coli* (Calvo and Matthews 1994; Newman and Lin 1995). In *E. coli*, expression of the *lrp* gene is up to four-fold higher in minimal media compared to rich media, and consequently Lrp is considered a positive factor for transcription enhancement of genes that are induced in starvation conditions (Ali Azam *et al.* 1999). Furthermore, the relatively high cellular abundance of Lrp, approximately 3000 molecules per cell, together with evidence that Lrp binding facilitates the formation of higher-order nucleoprotein complexes (Wang and Calvo 1993), leads to the suggestion that Lrp also plays a role in chromosomal organisation (D'Ari *et al.* 1993). Genes encoding Lrp are widely distributed in extant life, and occur in all sequenced archaeal genomes, and the majority of sequenced bacterial genomes. Since the Lrp family is the only group of DNA-binding proteins to have widespread occurrence in both the Archaea and the Bacteria, the interaction of the Lrp proteins with the eukaryal-like transcription apparatus in archaea, is currently a subject of considerable interest (Napoli *et al.* 1999; Bell and Jackson 2000; Brinkman *et al.* 2000).

Sac7 and Sac10b. First identified in *Sulfolobus acidocaldarius*, the Sac proteins consist of a group of low molecular-weight DNA-binding proteins, named according to their size and basicity as Sac10a, Sac10b, Sac8a, Sac8b, Sac7a, Sac7b, Sac7c, Sac7d and Sac7e (Dijk and Reinhardt 1986; Grote *et al.* 1986; Choli *et al.* 1988). The most abundant, and best-characterised, of these is the Sac7 family, which make up almost 5% of the total cytoplasmic proteins in members of the *Sulfolobus* genus. Sac7d has been shown to create sharp kinks in DNA on binding in the minor groove, and can constrain negative supercoils in DNA (Mai *et al.* 1998; Robinson *et al.* 1998; Napoli *et al.* 2002). The Sac10b family, which are unrelated to the Sac7 proteins, are also highly abundant and make up almost 4% of the total cytoplasmic protein. Using electron microscopy, Sac10b has been shown to bind DNA and form regular, though non-compacted, structures (Lurz *et al.* 1986). However, SSh10b from *Sulfolobus shibatae*, an orthologue of Sac10b, has been demonstrated to constrain DNA in negative supercoils in a temperature-dependent fashion (Xue *et al.* 2000). Sac10b homologues have been identified in every archaeal genome

sequenced to date, and have so far not been found in any bacterial genomes (Forterre *et al.* 1999). However, the Sac7 proteins have so far only been found within the *Sulfolobus* genus. The impact that these proteins have on the stress-response and the regulation of gene expression in members of the *Sulfolobus* genus is unknown.

MC1. MC1 (methanogen chromosomal protein 1) is the major chromosomal protein present in various species of *Methanosarcina* (Chartier *et al.* 1988; Imbert *et al.* 1990), and apart from the three sequenced genomes of the Methanosarcinales, is only present in the genome of *Halobacterium* sp. NRC-1 (Ng *et al.* 2000). MC1 binds relatively non-specifically to double-stranded DNA as a monomer (molecular weight of approximately 11 kDa) and protects DNA against thermal denaturation (Imbert *et al.* 1990). Moreover, MC1 facilitates DNA-bending which brings distal regions of the chromosome into close contact, which suggests possible roles in transcription and recombination (Cam *et al.* 1999). Additionally, the bends and topological changes introduced into the DNA by MC1 binding results in significant compaction of the DNA, although no nucleosome-like structures have been observed (Toulmè *et al.* 1995).

HTa. The HTa protein family, first identified in the archaeon *Thermoplasma acidophilum* (DeLange *et al.* 1981), and to date only found in the three genomes of the Thermoplasmales, are a group of nucleoid-associated DNA-binding proteins that form nucleosome-like structures (Searcy 1986). HTa has a molecular weight of approximately 10 kDa, and associates to form a tetrameric form, which has been indicated to wrap about 40 bp of DNA (Searcy 1986). As well as its role in compacting DNA, HTa is also suggested to protect DNA from the thermal (59°C) and acidic (pH optima of 1-2) environments, to which the Thermoplasmales are adapted (Searcy 1986). Although HTa was originally suggested to be distantly related to the eukaryal histones H2A, H3 and the bacterial HU (Searcy 1986), recent comparisons using the available sequence data for these proteins indicates that the similarity between these sequences is not statistically significant and that the HTa family of DNA-binding proteins forms a distinct family (Dinger, unpublished result).

Archaeal histones. Amongst the Euryarchaeota, the most commonly occurring DNA-binding proteins are the archaeal histones, which have primary sequence and structural similarity to the eukaryal histone-fold domain (Sandman *et al.* 1990; Arents and Moudrianakis 1995; Starich *et al.* 1996). Genes encoding archaeal histones occur in every euryarchaeal genome sequenced to date, with the exception of the genomes belonging to

members of the Thermoplasmales (Dinger, unpublished results). Archaeal histones have been shown to wrap DNA toroidally into structures that resemble those formed by the histone $(H3 + H4)_2$ tetramer at the centre of the eukaryal nucleosome, and consequently, are thought to resemble the archetypal ancestor of the eukaryal nucleosome core histones (Musgrave *et al.* 1991; Pereira *et al.* 1997). Like their eukaryal counterparts (Hamiche *et al.* 1996; Hamiche and Richard-Foy 1998), archaeal nucleosomes are also able to constrain DNA in both positive and negative supercoils (Musgrave *et al.* 2000). In a salt environment that resembles *in vivo* conditions, archaeal nucleosomes constrain DNA in negative supercoils, but in low salt conditions the tetramer is suspected to alter its conformation, and cause DNA to be constrained in positive supercoils (Musgrave *et al.* 2000). Furthermore, the observation that the DNA affinity of nucleosomes composed of different histone monomers is sequence-dependent, has lead to the suggestion that the specific assembly of archaeal histone-DNA complexes could be localized and used to regulate gene expression (Marc *et al.* 2002).

Previous investigations in our laboratory have revealed that the histones in the sulfur-metabolising thermophilic archaeon, *Thermococcus zilligii* (Klages 1991; Klages and Morgan 1994), are expressed and degraded in a growth phase-dependent fashion (Baillie 1994; Dinger 1998). Using antibodies that bound specifically to archaeal histones, it was shown that the levels of histones in *T. zilligii*, which constitute approximately 5% of the total soluble protein in exponentially growing cells, dramatically decreased as the cells entered stationary phase and were undetectable by late stationary phase. This surprising observation supported the notion that archaeal histones are involved in more than just DNA compaction. Moreover, in light of the accumulation of Dps in the stationary-phase of many bacteria, the hypothesis was raised that a replacement DNA-binding protein to organise and compact the DNA in stationary phase may also occur in archaea.

In order to identify potential regulatory or chromosomal-organising proteins in archaea, the total protein complements of *T. zilligii* were examined at various stages of growth by SDS-PAGE. Since the majority of known chromosomal organisation proteins in both archaea and bacteria are less than 20 kDa in mass, abundant low-molecular weight proteins were targeted for further examination. N-terminal sequence analysis in conjunction with existing genome data was then used to identify the putative candidates.

In this chapter, the identity of several low molecular-weight proteins that vary in response to growth-phase, are determined. Two of the identified proteins are found to exhibit

significant similarity with DNA-binding proteins characterised previously in other microorganisms. Two further proteins that belong to hitherto undescribed protein families are also identified.

Some of the results from this chapter were published as part of a paper in *Molecular Microbiology* (Dinger *et al.* 2000) (Appendix A) and presented at the *Thermophiles '98* conference in Brest, France (Appendix D).

1.2 Materials and Methods

1.2.1 Culturing of *Thermococcus zilligii*

(i) Media Preparation

Media for *Thermococcus zilligii* cells was prepared as described (Klages and Morgan 1994). *T zilligii* growth medium contains 8.6 mM (1.5 g/L) K_2HPO_4, 1.5 mM (0.3 g/L) $MgCl_2 \cdot 6H_20$, 42.8 mM (2.5 g/L) NaCl, 4.4 mM (0.5 g/L) sodium thioglycolate, 0.8% (w/v) trypticase peptone (Difco), 0.001% (w/v) resazurin, 0.1% (v/v) Wolin's vitamin solution (Wolin *et al.* 1963), 0.5% (v/v) Zeikus' trace elements solution (Zeikus *et al.* 1979) and 0.2% (w/v) sulfur (sterilised by baking at 100°C for at least two days). The total required volume of growth medium, without trace elements, vitamins and sulfur, was prepared in deionised water (>18 MΩ resistance), then divided into 850 mL aliquots in 1.0 L Schott bottles. Also, 100 mL was added to a 150 mL serum bottle containing 0.2 g sulfur and sealed with a rubber septum and metal cap. The bottles were then autoclaved for 20 min at 121°C, with loosened screw caps. After cooling to approximately 80°C, 1.7 g sulfur was added to each of the 1.0 L bottles, which were then sealed with rubber stoppers adapted with a rubber septum. Using a 5 mL syringe, with an 18-gauge hypodermic needle, 4.25 mL Zeikus' trace elements solution and 0.85 mL Wolin's vitamin solution were injected into the bottles. To the serum bottle, 500 µL trace elements and 100 µL vitamin solution were similarly added. To make the bottles anaerobic, the headspace was flushed with 0.2 µm filtered nitrogen gas for approximately five minutes or until the resazurin indicator turned fully yellow. The bottles were kept in a 75°C incubator, until inoculated (usually within an hour).

(ii) Culture inoculation and growth

The final inoculum was prepared in two steps. First, 100 μL of *T. zilligii* stock culture (obtained from the Thermophile Research Unit, The University of Waikato) was introduced aseptically into an anaerobically sealed Hungate tube containing 10 mL of *T. zilligii* growth medium. The cells were left to grow for 24 h, then examined by oil-immersion phase-contrast microscopy (Olympus BH-2) at 1000x magnification, to confirm that the culture was pure (i.e. no obvious contaminating species) and that it was in its log phase as indicated by a high proportion of dividing cells. The 150 mL serum bottle was inoculated with 2 mL of the log phase culture and incubated overnight at 75°C. The culture was examined to ensure it was in its log phase and its optical density measured at 650 nm. Each of the 1.0 L bottles containing 850 mL growth medium, were inoculated with 7.0 mL of the log phase culture and incubated at 75°C.

(iii) Measurement of cell concentration

Cell concentration was estimated by absorbance spectrophotometry. These measurements were converted to cell number by multiplying them by a conversion factor empirically determined by cell counting (as described below).

Cell densities were measured at least twice daily. Before removing a sample for measurement, the bottles were degassed (i.e. excess pressure created by hydrogen sulfide production was released) in a fume cupboard by inserting an 18-gauge hypodermic needle through the rubber septum. With a 2.5 mL syringe fitted with an 18-gauge hypodermic needle, 1.1 mL of culture was extracted and added to a 2 mL plastic cuvette, taking care to avoid bubbles forming. The absorbance was measured at 650 nm using a Shimadzu UV-250 spectrometer, with deionised water as the blank.

Cell counting was performed according to described methods (Koch 1994). A 250 μL aliquot of culture was transferred to a 1.5 mL Eppendorf tube. If the concentration was estimated to be greater than 3×10^8 cells per mL (equivalent to an absorbance of approximately 0.2) then the sample was diluted appropriately with a solution containing 0.05% (v/v) formaldehyde and 0.001% (w/v) sodium dodecyl sulfate (SDS; pH adjusted to 7.4 by addition of solid K_2HPO_4). An equal volume of 0.1 M hydrochloric acid was then added to this suspension. After gentle mixing, 3 μL of the diluted sample was placed on an A70 Helber (Hawksley) counting chamber (50 x 50 x 20 μm well dimensions or 0.05 nL

volume), using an autopipette, and a coverslip was carefully laid on top. The slide was allowed to stand for approximately 2 min, before being examined by oil-immersive phase-contrast microscopy (Olympus BH-2) at 1000x magnification. Since most of the cells adsorb to either the bottom or top interface, the depth of the wells could be verified, ensuring that the chamber had been filled correctly. The number of cells in at least 64 squares were scored systematically (a minimum of 250 cells). The number of cells per mL (cell density) of undiluted culture was then calculated, according to equation (1).

$$\frac{\text{Total cells counted} \times \text{dilution factor} \times 4 \times 10^8}{\text{Number of squares counted} \times 20} \qquad (1)$$

With the assumption that cell distribution was random, the standard error was calculated using the Poisson distribution, with a 95% confidence limit ($z_\alpha = 1.96$), as shown in equation (2).

$$\frac{1.96 \times \text{number of cells per mL}}{\sqrt{\text{Total cells counted}}} \qquad (2)$$

(iv) Cell harvesting

Cells were harvested from the entire contents of each bottle, at the appropriate times. The contents were decanted into two Beckman JA10 (450 mL) centrifuge bottles (leaving most of the sulfur behind), and centrifuged at 11,500 *g* (8,000 rpm in a JA10 rotor) at 15°C for 15 min. The supernatant was virtually all poured off, and the cell pellet resuspended in the remaining volume. The concentrated solution was transferred to two Beckman JA20 (25 mL) centrifuge tubes and phosphate-buffered saline (PBS; 100 mM K_2HPO_4, 0.05% (w/v) KCl, 0.8% (w/v) NaCl) was added to almost fill the tubes. The tubes were centrifuged at 17,400 *g* (12,000 rpm in a JA20 rotor) at 15°C for 15 min. The supernatant was poured off again, the cell pellet resuspended in 1.0 mL of PBS, and the cell concentrate transferred into two 1.5 mL Eppendorf tubes. The tubes were centrifuged at 18,600 *g* (12,000 rpm in a JA18.1 rotor) at 4°C for 10 min. The supernatant was carefully removed using an autopipette, and the cells were washed again in 1.0 mL of PBS and the cells pelleted again as described above. The supernatant was again removed and the cells stored at -70°C.

1.2.2 Preparation of total cytoplasmic proteins and acid wash extracts from *T. zilligii* cells

The *T. zilligii* cells were processed as described (Baillie 1994), with some alterations, to yield two relatively crude fractions of protein samples. The first fraction isolated consisted of the total soluble cytoplasmic protein content, since proteins associated with the cell wall are removed, and is referred to as the "total cytoplasmic proteins" or "TP". The second fraction consists of the acid-soluble proteins and is referred to as the "acid-wash extract" or "AWE".

Total cytoplasmic proteins were isolated by lysing the cells using sonication, followed by centrifugation to remove the cellular debris. Approximately 100 mg (wet-weight) of cells from each sample were transferred to separate 1.5 mL Eppendorf tubes and resuspended in 1.0 mL of sonication buffer (80 mM Tris-HCl pH 7.6, 4 mM spermidine, 20 mM ethylenediamine tetra-acetic acid (EDTA) pH 8.0, 50 mM NaCl, 1 mM PMSF). Each of these suspensions was then equally divided into two separate 1.5 mL tubes. The cells were disrupted by sonication for 2 min on ice at 40% power, using a Heat Systems Ultrasonics Sonicator (Ultrasonic Liquid Processor) fitted with a microtip. Lysis was verified by examining samples by oil-immersive phase-contrast microscopy at x1000 magnification. The two lysates from each sample were pooled and centrifuged at 15,900 *g* for 5 min at room temperature. Finally, the supernatants were transferred to new 1.5 mL tubes and stored at -70°C.

The AWE samples were prepared from the total cytoplasmic protein samples, by first precipitating out the acid-insoluble proteins, then removing them by centrifugation, and finally neutralising the acid-soluble fraction by dialysis. To a 1.5 mL Eppendorf tube, 500 µL of total cytoplasmic protein solution was transferred, then acidified by addition of 6.9 µL of 98% H_2SO_4 (final concentration 0.25 M). After mixing briefly, the tubes were incubated at room temperature for 15 min, then centrifuged at 15,900 *g* for 10 min at room temperature. The supernatant was neutralised by dialysis (3,500 Da molecular weight cut-off (MWCO)) against three changes of 4.0 L of 10 mM Tris-HCl pH 7.6, 150 mM KCl, and 7 mM β-mercaptoethanol at 4°C with stirring, allowing at least 3 h between changes. The contents of the dialysis tubing were transferred to 1.5 mL tubes, centrifuged at 15,900 *g* for 5 min. Finally, the supernatants were placed into new 1.5 mL tubes and stored at –70°C.

The amount of protein yielded in both the total cytoplasmic protein and AWE samples was subsequently determined using a modification of the Bradford assay (Bradford 1976; Bollag and Edelstein 1991). Finally, the volumes of the samples were estimated, and the samples were diluted appropriately to produce protein solutions of approximately equal concentrations.

1.2.3 Heparin-affinity column chromatography

The resolution of *T. zilligii* total cytoplasmic proteins according to their heparin-binding properties was performed according to a described method (Starich *et al.* 1996), with some modifications.

Total cytoplasmic proteins for heparin-affinity column chromatography were prepared as described in 1.2.2. Approximately 7.5 mg of total cytoplasmic protein was diluted with low salt buffer (30 mM potassium citrate and 50 mM Tris-HCl pH 8.0) to a final volume of 15 mL. The protein solution was then dialysed (3,500 Da MWCO) against three changes of 3 L of low salt buffer with stirring, allowing at least 3 h between changes. The dialysed solution was then filtered with a 0.45 μm filter (Sartorius), before being loaded onto the column.

Chromatography of the proteins was performed using a 5 mL HiTrap Heparin-affinity column (Amersham Biosciences). Sample and buffers were loaded onto the column using a Peristaltic Pump P-3 (Pharmacia Fine Chemicals), running at approximately 3 mL/min. First, the column was washed with five volumes of low salt buffer. Next, after loading the sample, the column was washed again with five volumes of low salt buffer. The proteins were eluted using a gradient of low salt buffer to high salt buffer (200 mM potassium citrate and 50 mM Tris-HCl pH 8.0). The gradient was formed using 30 mL of each buffer in a Model 385 Gradient Former (Model 385, Bio-Rad). A total of 40 fractions were collected in 1.5 mL tubes. Finally, the column was washed with a further two volumes of high salt buffer, then five volumes of low salt buffer, and stored in 20% ethanol.

The fractions were stored at -20°C and examined by SDS-PAGE as described in 1.2.5.

1.2.4 Ultrafiltration

To enrich low-molecular weight proteins for the purpose of N-terminal sequencing, total cytoplasmic proteins isolated from early stationary phase cultures of *T. zilligii* were ultrafiltrated. Approximately 40 mL 2 mg/mL total cytoplasmic proteins were loaded into a 50 mL ultrafiltration cell (Amicon), fitted with a YM100 ultrafiltration disc (MWCO 100 kDa; Millipore). The cell was placed under pressure using nitrogen gas, such that a flow rate of 1 mL/min was achieved. When the volume of the retentate was below 4 mL, the process was stopped and the filtrate transferred to a clean 50 mL ultrafiltration cell fitted with a YM3 ultrafiltration disc (MWCO 3 kDa; Millipore). This solution was then similarly concentrated down to approximately 4 mL. The filtrate was discarded, and the retentate stored at -70°C until it was analysed.

1.2.5 Electrophoresis and staining of proteins

(i) Preparation and electrophoresis of denaturing gels

For separation of the complex protein samples, discontinuous sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) was performed as described (Schägger and von Jagow 1987). For the large format gels (16 x 16 cm), a Protean II Slab Gel (Bio-Rad) electrophoresis system was used, and for the small format gels (8 x 8 cm), a Owl Model P8DS-1 Penguin Dual Gel System (Owl Separation Systems) was used. The process can be divided into three main stages: gel preparation, sample preparation and electrophoresis.

Gels were cast with freshly prepared gel solution made from stock solutions of acrylamide/bis-acrylamide (49.5% T, 3%C), Tris-HCl pH 8.45 (3.0 M), SDS (10%) and ammonium persulfate (APS; 10%; freshly-prepared). The final composition of the stacking gels was 4% T acrylamide/bis-acrylamide, 0.75 M Tris-HCl pH 8.45, 0.1% SDS, 0.08% APS, and 0.08% N,N,N',N'-tetramethyl-ethylenediamine (TEMED). The final composition of the separating gels was 16.5% T acrylamide/bis-acrylamide, 1.0 M Tris-HCl pH 8.45, 0.1% SDS, 13% w/v glycerol, 0.03% APS, and 0.03% TEMED. All solutions were prepared in deionised water (>18 MΩ resistance).

Separating gels were poured using a 60 mL or 10 mL syringe, for large and small format gels respectively, fitted with an 18-gauge hypodermic needle, then immediately overlaid carefully with a 1-2 cm layer of deionised water. The small format gels were cast using a

Joey™ Model JGC-2 Gel Casting System (Owl Separation Systems). Gels were allowed to polymerise for at least an hour, before pouring off the water layer and carefully removing any remaining liquid from the top of the gel with filter paper. Stacking gels were then poured similarly, and the comb (20 tooth or 12 tooth for large and small format gels, respectively) was put in place. The stacking gel was allowed to polymerise for at least 45 min before removing the comb, and rinsing any unpolymerised acrylamide from the wells with deionised water using a syringe with an 18-gauge hypodermic needle. The anode buffer consisted of 0.2 M Tris-HCl pH 8.9 (diluted from a 10x stock solution), and the cathode buffer consisted of 0.1 M Tris-HCl, 0.1 M Tricine, 0.1% (w/v) SDS (no pH adjustment).

The appropriate amount of protein sample was transferred to 0.2 mL tubes and, if required, diluted with sample buffer such that all samples were the same volume. An equal volume of 2x gel loading buffer (8% (w/v) SDS, 24% (w/v) glycerol, 100 mM Tris-HCl pH 6.8, 4% (v/v) β-mercaptoethanol, 0.02% (w/v) Coomassie Brilliant Blue G-250) was added to each tube, then the samples were incubated at 100°C for 10 min in a PTC-100™ Thermal Cycler (MJ Research, Inc.). The samples were then centrifuged, and loaded onto the gel using a 20 μL pipette. Samples were electrophoresed immediately after loading them.

Electrophoresis was performed using an ATTO Crosspower 1000 (Model AE-8400) power supply. The large format gels were electrophoresed initially at 30 V (constant voltage) for approximately 1 h (until the samples had completely entered the gel), then at 30 mA (constant current) for each 0.75 mm gel and 60 mA (constant current) for each 1.5 mm gel for approximately 15 h. Electrophoresis of the large format gels was carried out with cold tap water running continuously through the central cooling core of the unit. The small format gels were electrophoresed at 30 mA (constant current) per gel for approximately 2 h or until the dye began to run off the bottom of the gel.

Directly after electrophoresis was stopped, gels were placed in either fixative (for staining) or transfer buffer (for blotting).

(ii) Gel Staining

Proteins on polyacrylamide gels were visualised by staining using either Coomassie Blue or silver, depending on the desired sensitivity. In both methods, all solutions were prepared in deionised water (>18 MΩ resistance) and the stacking gel was discarded prior to staining.

Coomassie Blue staining was performed essentially as described (DeSilva *et al.* 1996). Gels were fixed in at least five gel volumes of a solution containing 50% (v/v) methanol and 10% (v/v) acetic acid, for 30 min. The fixative was then replaced with at least five gel volumes of staining solution containing 0.1% (w/v) Coomassie Blue G-250, 40% (v/v) methanol and 10% (v/v) acetic acid, for approximately 1 h. Finally, gels were destained in two changes of at least ten gel volumes of a solution containing 30% (v/v) methanol and 10% (v/v) acetic acid, for at least 3 h or until the background was essentially clear.

Silver staining was performed essentially as described (DeSilva *et al.* 1996). Each step was performed with ten gel volumes of solution. Gels were fixed in a solution containing 30% (v/v) ethanol and 10% (v/v) acetic acid for 15 min, washed in deionised water three times for 5 min each, then stained for 30 min in a freshly-prepared solution containing 2% (w/v) $AgNO_3$ and 0.037% (v/v) formaldehyde. After staining, gels were rinsed briefly in deionised water, then developed in a solution containing 3% (w/v) Na_2CO_3, 0.037% (v/v) formaldehyde, and 0.002% (w/v) sodium thiosulfate. Once gels were developed to the desired extent, staining was stopped by replacing the developing solution with a solution containing 7% (v/v) acetic acid.

After staining, gels were scanned (Microtek ScanMaker V300) and stored in hermetically sealed plastic bags.

(iii) Membrane blotting and staining

Proteins were transferred from polyacrylamide gels onto a 0.45 μm polyvinylidene fluoride (PVDF) membrane (Hybond-P, Amersham Biosciences) using an Owl Model HEP-1 Panther™ Semi-Dry Electroblotter (Owl Separation Systems), according to the manufacturer's instructions. Directly following electrophoresis, the stacking gel was removed, and the remaining separating gel washed in transfer buffer (25 mM Tris-HCl pH 8.3, 192 mM glycine, 20% (v/v) methanol) for 15 min. The arrangement for the blot is illustrated in Figure 1-1. Layers of filter paper were soaked in transfer buffer then placed individually, ensuring there were no trapped air bubbles. The membrane was wetted with 100% methanol. Gels were electroblotted using a Lightning Volt™ OSP-250L Power Supply (Owl Separation Systems), at 400 mA (constant current) for 50 min. Immediately following transfer, the membrane was stained with amido black as described (Best and Speicher 1996). The membrane was washed three times for 5 min each. The proteins were then stained by incubating the membrane for 1 min in a solution containing 0.1% (w/v)

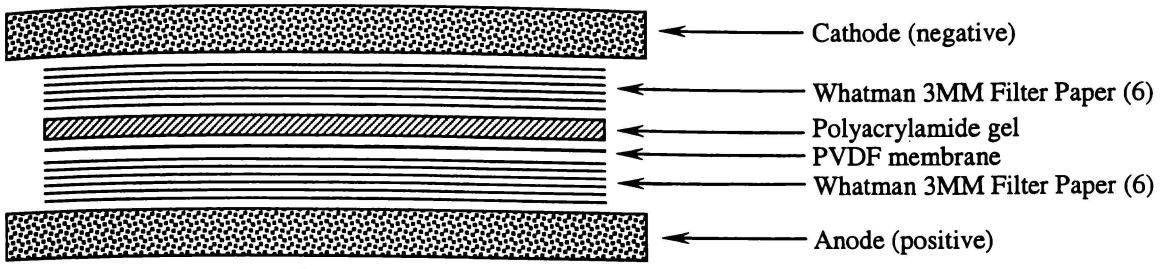


Figure 1-1. Arrangement of components for transferring protein from a polyacrylamide gel to a PVDF membrane.

amido black 10B and 10% (v/v) acetic acid. Background staining was removed by washing the membrane twice for 1 min each in 5% (v/v) acetic acid then rinsing in two changes of deionised water for 10 min each. Finally, the membrane was allowed to air dry at room temperature.

1.2.6 Protein sequencing and identification

Protein sequencing was performed using an Applied Biosystems Procise Sequencer (Model 610A) by the Protein Sequencing Facility, School of Biological Sciences, University of Auckland. The automated system uses Edman degradation to sequentially cleave and identify amino acids starting at the amino terminus (N-terminus) of the protein. Proteins to be sequenced were supplied on PVDF membrane in quantities of at least 30 pmol. To ensure that an unambiguous identification of the target protein could be made, a minimum of ten residues were sequenced from the N-terminus of each protein.

The N-terminal sequence data was then used to identify potential orthologues in GenBank. Searches were performed using BLASTP 2.2.2 (Altschul *et al.* 1997) at the National Centre for Biotechnology Information (NCBI), using the PAM30 scoring matrix, the Word Size set to 2 and the Expect threshold set to 100 or more depending on the length and complexity of the input sequence. To maximise the relevance of the results, searches were initially limited to sequences that originated from Archaea.

1.3 Results

1.3.1 The levels of five low molecular-weight proteins, SRATzi, SRBTzi, SRCTzi, SRDTzi and SRETzi, vary with changing growth-phase in *Thermococcus zilligii*

To identify potential proteins involved in the stress-response, changes in the total protein patterns of the archaeal isolate, *T. zilligii* were examined. *T zilligii* cell cultures were grown to various stages of their growth cycle, ranging from mid-log to late stationary phase (Figure 1-2). The mean cell densities of the cultures were determined regularly throughout the culturing period. Surprisingly, the cell density decreased dramatically from 6.5 x 10^8 to 4.5 x 10^8 between the 40 and 70 h time points and continued to drop to approximately 3.5 x 10^8 over the subsequent 40 h. Although the causes of this phenomenon were not examined in detail, the comparison of cells from early- to mid- stationary phase indicated increasing

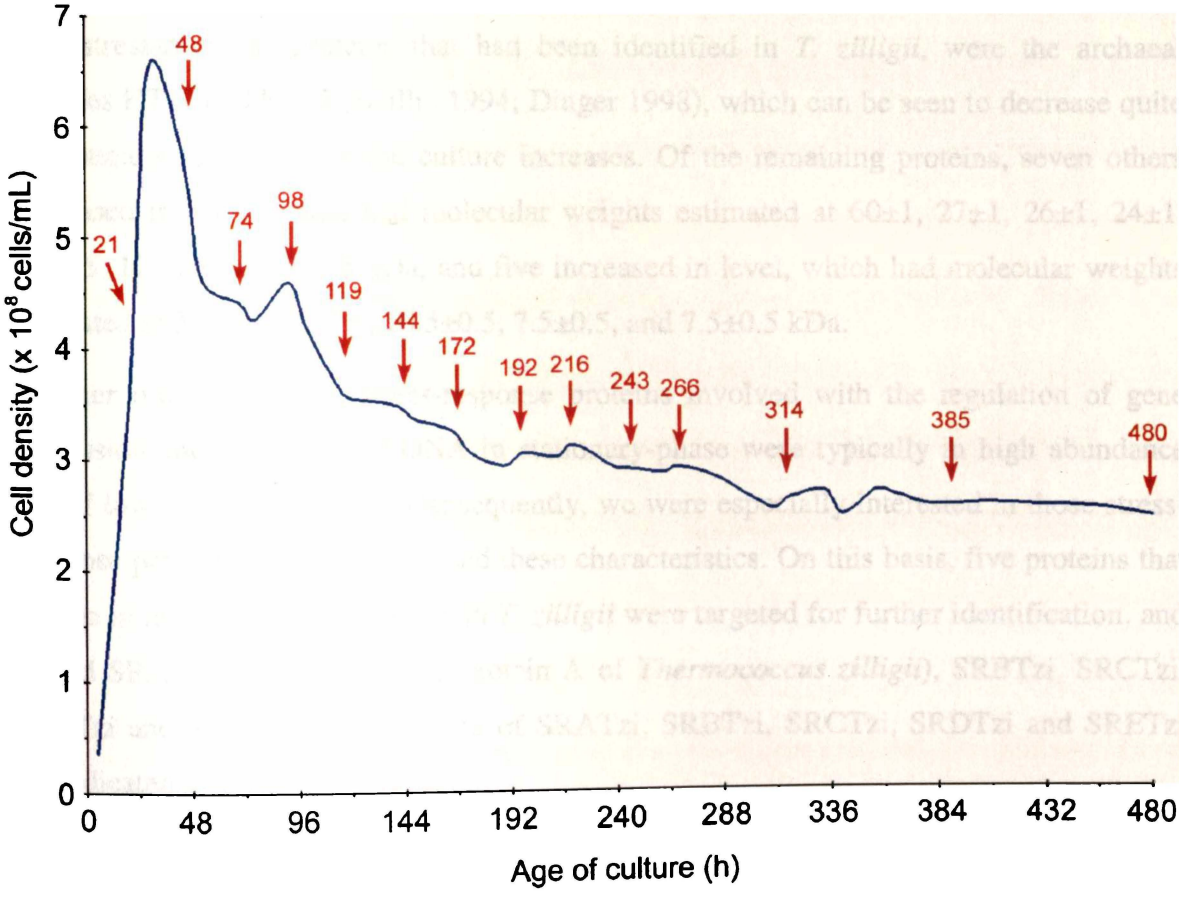


Figure 1-2. Mean cell density of *T. zilligii* cultures grown up to 480 h, showing the various points at which cells were harvested for further study. The cultures were grown to late stationary phase in rich medium at 75°C. Cell densities were obtained at least once every 12 h, by measuring the OD_{650} and converting the readings using conversion factors that had been empirically determined by direct cell counts at various stages of the growth cycle (Dinger 1998).

proportions of what appeared to be ruptured cells. Cell lysis could be caused by a number of factors, including build-up of toxic metabolic waste products or simply a nutrient or mineral deficiency. Further research into the metabolism and physiology of this microorganism would be required to understand the cause of this phenomenon. The viability of the cells from each of the harvesting points was analysed qualitatively, by taking an aliquot from the cultures and using them to inoculate fresh medium. In each case, rapidly growing cultures were obtained within 24 h (results not shown).

Total cytoplasmic proteins were extracted from the harvested cells, purified, then resolved by Tris-tricine SDS-PAGE (Figure 1-3). By using only the soluble cytoplasmic proteins, rather than whole cell extracts, the resolution of low-molecular weight proteins was found to be considerably enhanced (results not shown). The resolution of the total cytoplasmic proteins from cultures of various ages, shows that although the overall protein patterns appear very similar over the period of the growth cycle examined, at least 13 proteins (indicated by arrows) could be identified that varied significantly in level. Previously, the only stress-response proteins that had been identified in *T. zilligii*, were the archaeal histones HTz1 and HTz2 (Baillie 1994; Dinger 1998), which can be seen to decrease quite dramatically as the age of the culture increases. Of the remaining proteins, seven others decreased in level, which had molecular weights estimated at 60±1, 27±1, 26±1, 24±1, 18±0.5, 12±0.5, and 9±0.5 kDa, and five increased in level, which had molecular weights estimated at 36±1, 14.0±0.5, 10.5±0.5, 7.5±0.5, and 7.5±0.5 kDa.

In other microorganisms, stress-response proteins involved with the regulation of gene expression and protection of DNA in stationary-phase were typically in high abundance and of low molecular weight. Consequently, we were especially interested in those stress-response proteins that also possessed these characteristics. On this basis, five proteins that change in level with growth phase in *T. zilligii* were targeted for further identification, and named SRATzi (**S**tress **R**esponse protein **A** of ***Thermococcus zilligii***), SRBTzi, SRCTzi, SRDTzi and SRETzi. The positions of SRATzi, SRBTzi, SRCTzi, SRDTzi and SRETzi are indicated in Figure 1-3.

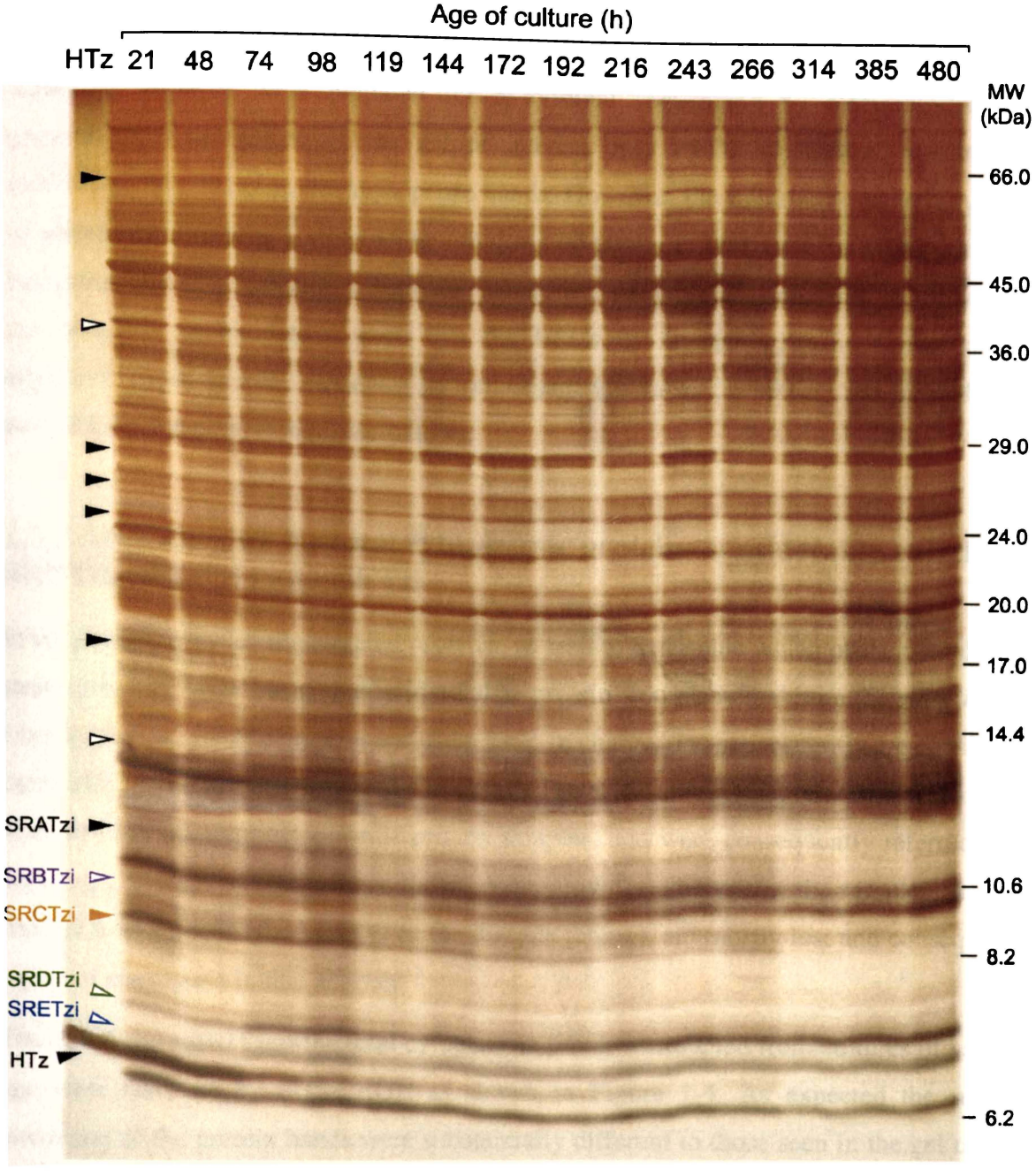


Figure 1-3. Changes in the total cytoplasmic protein complements of *T. zilligii* in response to age of culture. Total cytoplasmic proteins were extracted from *T. zilligii* cultures after various periods of growth, resolved by SDS-PAGE and stained with silver. The concentrations of the *T. zilligii* proteins samples were adjusted, so that each lane was loaded equally (10 μg). Open arrows indicate proteins that increase in level and closed arrows indicate proteins that decrease in level. The positions of five putative stationary phase response proteins (SRATzi, SRBTzi, SRCTzi, SRDTzi, and SRETzi) and HTz are indicated. Molecular weights (MW) were estimated from the 17S and VIIIL standards (Sigma-Aldrich). Lane HTz contains 0.5 μg of purified native HTz.

To determine whether SRATzi, SRBTzi, SRCTzi, SRDTzi and SRETzi, belonged to previously identified protein families, N-terminal sequencing of the proteins was attempted. Proteins were prepared for N-terminal sequencing by electroblotting the proteins from the region of the gel of interest onto PVDF membrane, staining the membrane, then excising the appropriate bands. However, the minimum required quantity of protein for N-terminal sequencing could not be attained. This was due to the technical limitation that the loading of greater quantities of proteins onto the gel caused a diminishment of resolution, such that the bands of interest were no longer clearly separated. Consequently, strategies to further enrich SRATzi, SRBTzi, SRCTzi, SRDTzi and SRETzi were developed (see below).

1.3.2 Acid-washing the total cytoplasmic proteins considerably enriches SRBTzi, SRDTzi and SRETzi

In an attempt to enrich the targeted proteins, as well to determine which of the cytoplasmic stress proteins were basic, the acid-soluble protein complements from *T. zilligii* were examined. Acid-soluble proteins are often basic in charge, and consequently are considered more likely to be associated with the nucleoid. The acid-soluble fractions were prepared by acid-washing the total cytoplasmic protein samples, and were consequently referred to as acid-wash extracts (AWEs). Since the procedure involves dialysis of the samples, proteins smaller than 3.5 kDa (the MWCO of the dialysis tubing), are mostly lost and consequently were not considered in this analysis.

The AWEs prepared from the total cytoplasmic proteins extracted from cultures of varying age were resolved by SDS-PAGE as shown in Figure 1-4. As expected the relative intensities of the protein bands were substantially different to those seen in the gel of total cytoplasmic proteins. Although, all of the five proteins of interest could be identified in the AWEs, only SRBTzi, SRDTzi and SRETzi were considerably enriched relative to the other proteins. SRATzi and SRCTzi both appeared somewhat diminished in concentration relative to the other proteins in the extract.

The maximal quantity of an AWE prepared from early stationary-phase total cytoplasmic proteins (derived from a 74 h culture of *T. zilligii*), was resolved by SDS-PAGE and the proteins from the region of interest were transferred to PVDF membrane and stained. The bands corresponding to SRBTzi, SRDTzi and SRETzi were excised from the membrane and processed for N-terminal sequence analysis. SRBTzi and SRETzi were transferred to

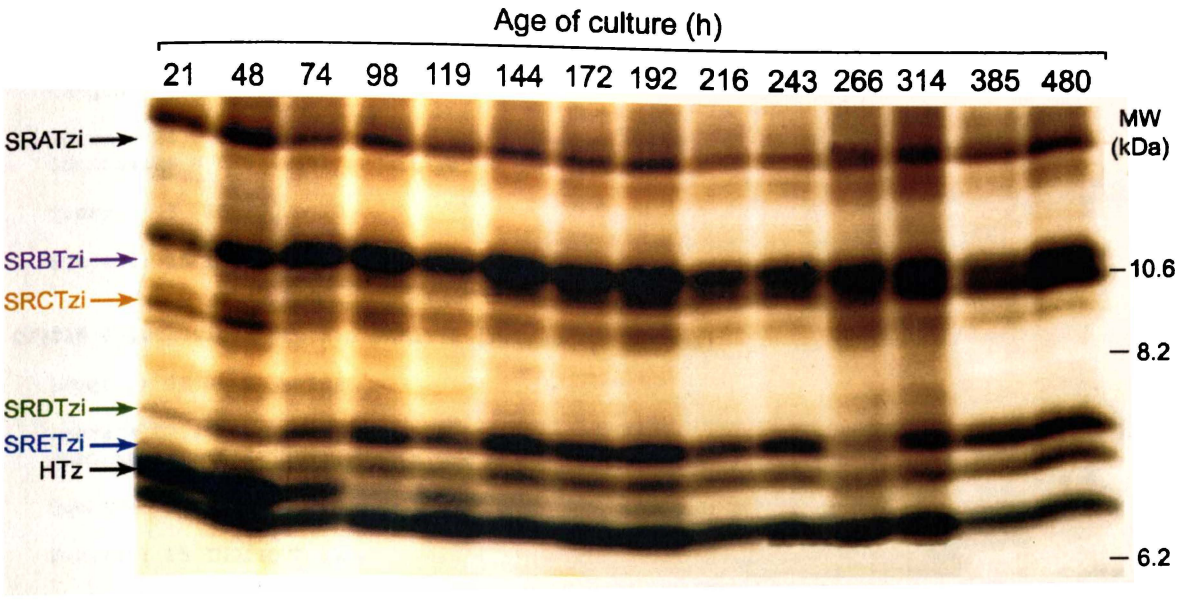


Figure 1-4. Changes in response to age of culture in the low-molecular weight, acid-soluble protein complements of *T. zilligii*. The acid-soluble protein fractions were prepared from *T. zilligii* cultures after various periods of growth, separated by SDS-PAGE and stained with silver. The molecular weights (MW) were determined using the 17S standard (Sigma-Aldrich). The positions of the five potential stationary phase response proteins (SRATzi, SRBTzi, SRCTzi, SRDTzi, and SRETzi) and HTz are indicated.

the membrane in sufficient quantity (approximately 10 pmol each) and purity to provide ten residues of reasonably accurate sequence data. The band corresponding to SRDTzi was not in sufficient quantity for N-terminal sequence analysis. Unfortunately, simply loading more concentrated sample was not possible, since this had the effect of decreasing the resolution of the low-molecular weight proteins such that they were no longer sufficiently separated. Consequently, a strategy to further selectively enrich this protein was required (see 1.3.7).

1.3.3 SRETzi belongs to a highly conserved, hitherto undescribed family of exclusively archaeal proteins

Using the SRETzi-enriched AWE, ten residues from the N-terminal sequence of SRETzi were elucidated. From the N-terminus, residues 3 to 12 read as E-L-I-Q-Q-I-V-Q-V-L. The first two residues could not be accurately determined.

Potential orthologues of SRETzi were identified by searching GenBank using the BLASTP program at the NCBI. Due to the short length and relatively low complexity of the input sequence, the Expect threshold was set to 100, the low-complexity filter was disabled, and the search was limited to sequences from Archaea. Amongst the top 20 matches (sorted by score), three sequences from three different species of *Pyrococcus* (a member of the

```
NP_143752 Hypothetical protein [Pyrococcus horikoshii]
O74108 Hypothetical protein PHS059 [Pyrococcus horikoshii]
A71207 Hypothetical protein PHS059 [Pyrococcus horikoshii]
BAA31048 93aa long hypothetical protein [Pyrococcus horikoshii]
   Length = 93 amino acids

   Score = 24.0 bits (49), Expect = 4.9
   Identities = 6/8 (75%), Positives = 8/8 (100%)

   Query:    3  IQQIVQVL 10
                I+QI+QVL
   Subject: 15 IEQIIQVL 22


Q9UY10 Hypothetical protein PAB1112 [Pyrococcus abyssi]

   Length = 93 amino acids

   Score = 22.3 bits (45), Expect = 16
   Identities = 6/8 (75%), Positives = 7/8 (87%)

   Query:    3  IQQIVQVL 10
                I QI+QVL
   Subject: 15 INQIIQVL 22


NP_577968 Hypothetical protein PF0239 [Pyrococcus furiosus DSM 3638]
AAL80363 Hypothetical protein PF0239 [Pyrococcus furiosus DSM 3638]

   Length = 86 amino acids

   Score = 22.3 bits (45), Expect = 16
   Identities = 6/8 (75%), Positives = 7/8 (87%)

   Query:    3  IQQIVQVL 10
                I QI+QVL
   Subject: 8  INQIIQVL 15
```

Figure 1-5. Top scoring GenBank search results for the N-terminal sequence of SRETzi using BLASTP. Searches were limited to sequences originating from Archaea and only those matches where the similar region occurs near the N-terminus are shown. The GenBank accession numbers for matched sequences are shown in bold print. Searches were performed using BLASTP 2.2.2 using the PAM30 scoring matrix, with a Word Size of 2.

Thermococcales) exhibited substantial homology with the input sequence in proximity to their N-terminus (Figure 1-5). Further analyses indicated that all three of these hypothetical proteins belong to the same protein family. Although the Expect values for the matches in the initial search are relatively high, the additional coincidences of their occurrence at the N-terminus, their similar theoretical molecular weight, and the close phylogenetic relationship between *Pyrococcus* and *Thermococcus*, strongly suggests that SRETzi also belongs to this protein family. The protein family has to date not been described, but initial searches at GenBank indicated that representatives of the family are present in virtually every archaeal genome sequenced so far and absent in all bacterial genomes. Details concerning the identification of the various SRE amino acid sequences, as well as a comprehensive computational analysis of the SRE family follows in Chapter Two.

1.3.4 SRBTzi belongs to an unknown family of proteins found in some Archaea and Bacteria

As stated in 1.3.2, the band corresponding to SRBTzi was excised from a membrane that contained proteins transferred from an early stationary-phase AWE preparation, which had been resolved by SDS-PAGE. The band corresponding to SRBTzi was excised and ten residues from the N-terminal sequence were successfully determined. From the N-terminus, residues 1 to 10 read as M-K-I-A-I-P-T-L-K-G.

Homologues of SRBTzi were identified by searching archaeal sequences in GenBank using the BLASTP program at the NCBI. As with SRETzi, the top scoring matches that exhibited homology with the input sequence in proximity to their N-terminus, were three sequences from three closely-related members of the Thermococcales (Figure 1-6). In addition, although with a considerably lower score, there was also a match with the N-terminus of a protein from *Archaeoglobus fulgidus*. The subsequent analysis revealed that all four of these hypothetical proteins belong to the same protein family. The theoretical molecular weight of these four proteins ranges from 12.5 to 19.0 kDa, which is of a similar order to the molecular weight of SRBTzi, which was estimated at 10.5±0.5 kDa. So, despite the Expect values of these matches being 0.84 or higher, the additional coincidences of the matches occurrence at the N-terminus, the comparable theoretical molecular weights, and the close phylogenetic relationship between *Pyrococcus* and *Thermococcus*, led to the conclusion that SRBTzi almost certainly belongs to the same family.

The protein family that SRBTzi is suggested to belong to, has to date not been well described, and its function remains unknown. Representatives of the family occur in one or more copies in many, but not all, of the archaeal genomes sequenced so far, and hitherto only in two bacterial genomes, *Thermotoga maritima* and *Clostridium acetobutylicum*. A comprehensive computational analysis of the SRB family follows in Chapter Two.

NP_579754 Hypothetical protein PF2025 [*Pyrococcus furiosus* DSM 3638]
AAL82149 Hypothetical protein PF2025 [*Pyrococcus furiosus* DSM 3638]

Length = 179 amino acids

Score = 26.5 bits (55), Expect = 0.84
Identities = 7/7 (100%), Positives = 7/7 (100%)

```
Query:   1 MKIAIPT 7
           MKIAIPT
Subject: 1 MKIAIPT 7
```

NP_142807 Hypothetical protein [*Pyrococcus horikoshii*]
H71076 Hypothetical protein PH0880 [*Pyrococcus horikoshii*]
BAA29974 116aa long hypothetical protein [*Pyrococcus horikoshii*]

Length = 116 amino acids

Score = 24.4 bits (50), Expect = 3.7
Identities = 9/11 (81%), Positives = 9/11 (81%), Gaps = 1/11 (9%)

```
Query:   1  M-KIAIPTLKG 10
            M KIAIPT KG
Subject: 1  MIKIAIPTSKG 11
```

H71051 Hypothetical protein PH1111 [*Pyrococcus horikoshii*]
BAA30210 135aa long hypothetical protein [*Pyrococcus horikoshii*]

Length = 135 amino acids

Score = 24.0 bits (49), Expect = 4.9
Identities = 8/12 (66%), Positives = 8/12 (66%), Gaps = 4/12 (33%)

```
Query:   1  MK----IAIPTL 8
            MK    IAIPTL
Subject: 1  MKPLSVIAIPTL 12
```

NP_071201 Conserved hypothetical protein [*Archaeoglobus fulgidus*]
B69547 Conserved hypothetical protein AF2378 [*Archaeoglobus fulgidus*]
AAB91286 Conserved hypothetical protein [*Archaeoglobus fulgidus*]

Length = 144

Score = 21.4 bits (43), Expect = 29
Identities = 7/10 (70%), Positives = 8/10 (80%)

```
Query:   1  MKIAIPTLKG 10
            M+IA  TLKG
Subject: 1  MRIAATTLKG 10
```

Figure 1-6. Top scoring GenBank search results for the N-terminal sequence of SRBTzi using BLASTP. Searches were limited to sequences originating from Archaea and only those matches where the similar region occurs near the N-terminus are shown. The GenBank accession numbers for matched sequences are shown in bold print. Searches were performed using BLASTP 2.2.2 using the PAM30 scoring matrix, with a Word Size of 2.

1.3.5 Heparin-affinity chromatography resolves HTz1/HTz2 and SRCTzi from early stationary-phase total cytoplasmic proteins

With the hypothesis in mind that one or more of the targeted low molecular weight stress-response proteins may replace the DNA-compaction role of HTz1 and HTz2, we considered that some of the proteins may have similar biochemical properties. In the purification of recombinant archaeal histones expressed in *Escherichia coli*, one of the crucial steps is the separation of the histones from the majority of the other proteins by heparin-affinity chromatography (Sandman *et al.* 1994; Starich *et al.* 1996). Based on this technique, early-stationary phase (48 h) total cytoplasmic proteins isolated from *T. zilligii* cultures were loaded onto a heparin-affinity column and the proteins eluted in fractions, using a potassium glutamate gradient. The fractions were then analysed by SDS-PAGE as shown in Figure 1-7. As expected, HTz1 and HTz2 started to elute at potassium glutamate concentrations of approximately 90 mM, but more interestingly, a larger protein started to elute at approximately 160 mM. By comparing this purified protein fraction with the original early-stationary phase total cytoplasmic proteins, it was found that this protein migrated at exactly the same position as the protein that had been designated as SRCTzi. Moreover, in the molecular weight range that this purified protein migrated at (approximately 10 kDa), SRCTzi is the only protein that is present at a concentration sufficient to have yielded the amount of protein that was eluted from the column. Consequently, it was concluded that the protein that had been purified, was, fortuitously, SRCTzi.

Since the heparin-affinity chromatography purified SRCTzi to near homogeneity (contaminating proteins only being detectable by silver staining), after concentration by ultrafiltration, it was very readily transferable to a membrane in sufficient concentration and purity for N-terminal sequencing.

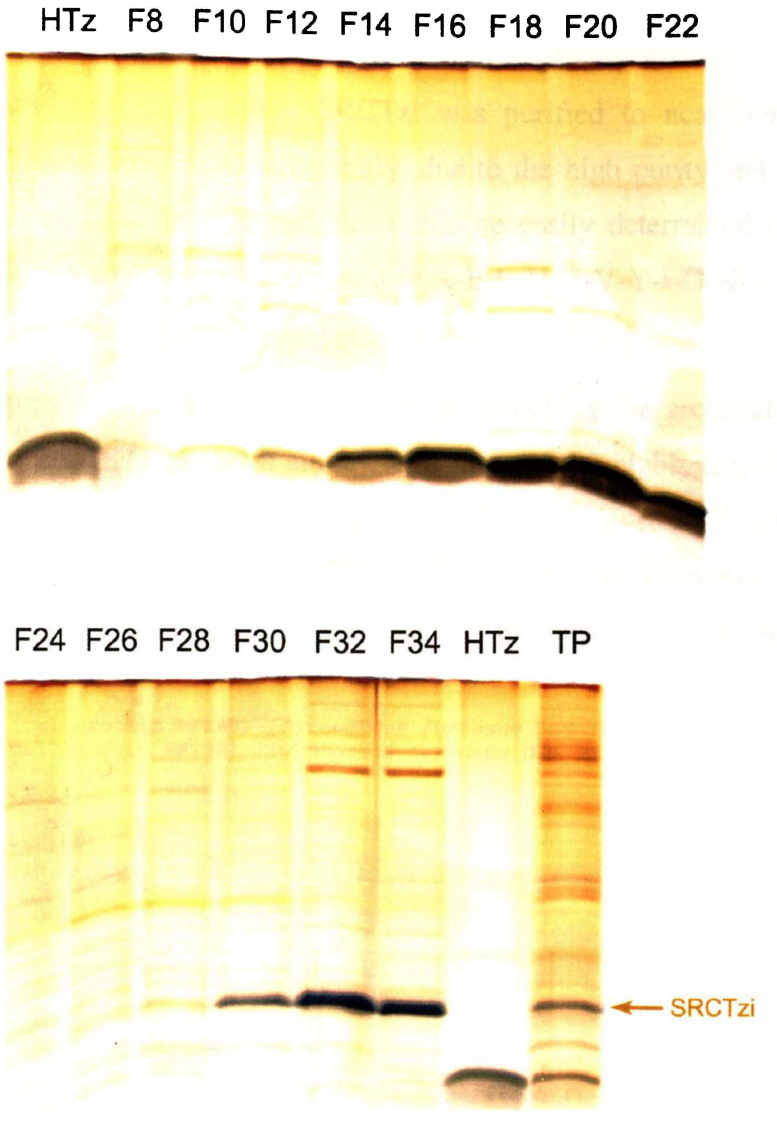


Figure 1-7. Heparin-affinity chromatography of total cytoplasmic proteins from early stationary-phase *T. zilligii* cells. Fractions, numbered F1 to F40, were eluted from a heparin-affinity column with a 30 to 200 mM potassium glutamate gradient, in 50 mM Tris-HCl pH 8.0. The fractions were examined by SDS-PAGE and stained with silver. Lane HTz contains 250 ng of purified native HTz as a standard, and Lane TP contains 5 μg of total cytoplasmic proteins isolated from early stationary-phase (48 h) *T. zilligii* cells, with the position of SRCTzi labelled.

1.3.6 SRCTzi belongs to the Sac10b family of archaeal-specific DNA-binding proteins

As described in the previous section, SRCTzi was purified to near homogeneity by heparin-affinity chromatography. Consequently, due to the high purity and concentration of the SRCTzi sample, the first 20 residues could be easily determined by N-terminal sequencing. Residues 1 to 20 were determined as A-E-E-H-V-V-Y-I-G-K-K-P-V-M-N-Y-V-L-A-V.

Potential orthologues of SRCTzi were identified by searching the archaeal sequences in GenBank using the BLASTP program at the NCBI. In total, 13 highly significant matches were found, with Expect values ranging from 2×10^{-13} to 7×10^{-5} (Figure 1-8). These matches all occurred at the N-terminus of the proteins and were members of the Sac10b family of DNA-binding proteins. Although Sac10b proteins have not been identified

NP_579610 Hypothetical protein PF1881 [*Pyrococcus furiosus* DSM 3638]
AAL82005 Hypothetical protein PF1881 [*Pyrococcus furiosus* DSM 3638]

Length = 93 amino acids

Score = 68.5 bits (154), Expect = 2e-13
Identities = 20/20 (100%), Positives = 20/20 (100%)

Query: 1 AEEHVVYIGKKPVMNYVLAV 20
AEEHVVYIGKKPVMNYVLAV
Subject: 2 AEEHVVYIGKKPVMNYVLAV 21

NP_143671 Hypothetical protein [*Pyrococcus horikoshii*]
A71196 Hypothetical protein PHS053 [*Pyrococcus horikoshii*]
BAA30960 93aa long hypothetical protein [*Pyrococcus horikoshii*]

Length = 93 amino acids

Score = 66.0 bits (148), Expect = 1e-12
Identities = 19/19 (100%), Positives = 19/19 (100%)

Query: 2 EEHVVYIGKKPVMNYVLAV 20
EEHVVYIGKKPVMNYVLAV
Subject: 3 EEHVVYIGKKPVMNYVLAV 21

NP_126085 Conserved hypothetical protein [*Pyrococcus abyssi*]
E75154 Conserved hypothetical protein PAB7094 [*Pyrococcus abyssi* (strain Orsay)]
CAB49316 Conserved hypothetical protein [*Pyrococcus abyssi*]

Length = 93 amino acids

Score = 66.0 bits (148), Expect = 1e-12
Identities = 19/19 (100%), Positives = 19/19 (100%)

Query: 2 EEHVVYIGKKPVMNYVLAV 20
EEHVVYIGKKPVMNYVLAV
Subject: 3 EEHVVYIGKKPVMNYVLAV 21

Figure 1-8. Top three scoring search results from GenBank for the N-terminal sequence of SRCTzi using BLASTP. Searches were limited to sequences originating from Archaea and only those matches where the similar region occurs near the N-terminus are shown. The GenBank accession numbers for matched sequences are shown in bold print. Searches were performed using BLASTP 2.2.2 using the PAM30 scoring matrix, with a Word Size of 2.

previously in any species of *Thermococcus*, members of the Sac10b family have been found in every archaeal genome sequenced to date and have not been found in bacterial genomes (Forterre *et al.* 1999). Finally, the molecular weight of members of the Sac10b family, approximately 10 kDa, agrees with that estimated for SRCTzi. These observations overwhelmingly support the suggestion that SRCTzi is a member of the Sac10b family. A comprehensive computational analysis of the Sac10b/SRC family is presented in the next chapter.

1.3.7 SRDTzi belongs to a family of proteins that is similar to the N-terminus of the Leucine-responsive regulatory proteins

Since SRDTzi was not sufficiently concentrated in the AWE for the purposes of N-terminal sequencing, a further enrichment strategy was developed. Ultimately, a crude, but effective, low-molecular weight protein enrichment system was established, which involved passing the total cytoplasmic proteins (isolated from early stationary-phase *T. zilligii* cells) through a cellulose ultrafiltration membrane (100 kDa MWCO), and retaining the filtrate. This procedure has the effect of removing the majority of the high molecular weight proteins, which subsequently allows a greater proportion of low molecular weight proteins to be loaded onto the gel. The effectiveness of the concentration of low molecular weight proteins using this system is demonstrated in Figure 1-9. The SRDTzi-enriched filtrate could then be accurately resolved on a large format denaturing polyacrylamide gel at sufficient concentration for N-terminal sequencing. To ensure that the protein originally identified as SRDTzi was the same as that excised from the low-molecular weight enriched filtrate, both samples were loaded on adjacent tracks on the gel that was transferred to the membrane (results not shown).

Due to the high purity and concentration of the band corresponding to SRDTzi, its N-terminal sequence could be determined accurately for at least the first 17 residues. Residues 1 to 17 read as V-R-A-Y-V-L-L-T-I-E-I-R-K-V-E-S-V.

Potential orthologues of SRDTzi were identified by searching archaeal sequences in GenBank using the BLASTP program at the NCBI. Surprisingly, only one of the results featured substantial similarity to the input sequence. As shown in Figure 1-10, the match occurred at the N-terminal end of a protein sequence from *Pyrococcus furiosus* (PF1893) and the low Expect value predicts that this match has only a 1 in 2500 chance of occurring randomly in the current database of archaeal sequences. Combined with the similarity in

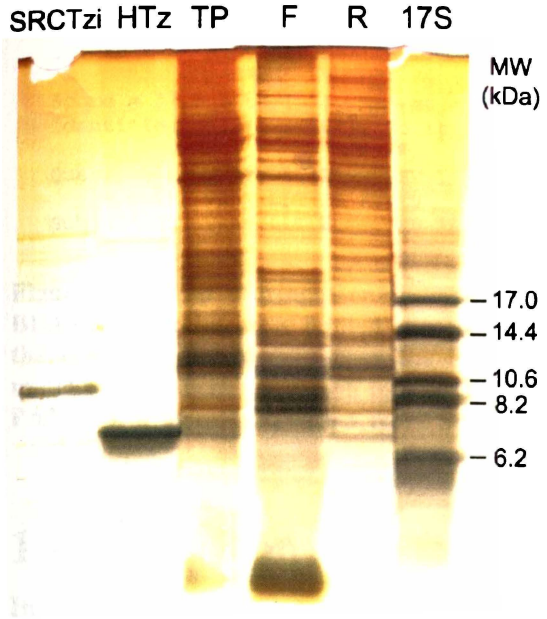


Figure 1-9. Enrichment of low molecular weight proteins by ultrafiltration. Total cytoplasmic proteins (TP) from early stationary-phase (48 h) *T. zilligii* cells were passed through a YM100 filter by ultrafiltration. The resulting filtrate (F) and retentate (R) fractions were resolved by SDS-PAGE and stained with silver. The 17S molecular weight (MW) marker and purified preparations of SRCTzi and HTz were loaded as standards.

the theoretical molecular weight of this protein and that estimated for SRDTzi, these observations provide considerable support to the suggestion that SRDTzi is a homologue of PF1893.

Interestingly, the GenBank entry for PF1893 describes it as a "transcriptional regulatory protein", belonging to the AsnC family. However, all characterised AsnC proteins are approximately twice the length of PF1893, and a comparison of the sequences reveals that PF1893 is similar to the C-terminal half of these proteins. In addition, the AsnC family are closely related to the larger family of Leucine-responsive regulatory proteins (Lrps) (Newman and Lin 1995), to which PF1893 also has significant similarity to the C-terminal portion thereof. Lrps have been extensively studied for their involvement in the stress-response in both Gram-positive and Gram-negative bacteria, and have been shown to increase in level in response to stress (Ali Azam *et al.* 1999). A comprehensive computational analysis of the SRD family and their relationship to the AsnC/Lrp family follows in Chapter Two.

```
NP_579622 Transcriptional regulatory protein, AsnC family PF1893 [Pyrococcus furiosus]
AAL82017 Transcriptional regulatory protein, AsnC family PF1893 [Pyrococcus furiosus]
  Length = 81 amino acids

  Score = 37.5 bits (81), Expect = 4e-04
  Identities = 12/17 (70%), Positives = 15/17 (87%)

  Query:   1  VRAYVLLTIEIRKVESV 17
              VRAY+LLT+EI KV+ V
  Subject: 2  VRAYILLTVEIGKVQKV 18
```

Figure 1-10 Top scoring GenBank search result for the N-terminal sequence of SRDTzi using BLASTP. Searches were limited to sequences originating from Archaea and only those matches where the similar region occurs near the N-terminus are shown. The GenBank accession numbers for matched sequences are shown in bold print. Searches were performed using BLASTP 2.2.2 using the PAM30 scoring matrix, with a Word Size of 2.

1.4 Discussion

In this chapter, the global protein patterns of *T. zilligii* were examined and at least 13 proteins were detected that varied significantly in level with changing growth phase. Having particular interest in proteins that may be involved in transcriptional regulation, chromosomal organisation or compacting DNA in stationary-phase, five relatively low molecular weight proteins that varied in response to changing growth phase were targeted for further analysis. These proteins were named SRATzi, SRBTzi, SRCTzi, SRDTzi, and SRETzi. Using several different enrichment strategies, sufficient purity and quantity was attained for the latter four proteins to allow N-terminal sequence analysis. Homologues of each of these four proteins were convincingly identified in GenBank using the BLASTP search program.

SRETzi is a protein with an estimated molecular weight of 7.5±0.5 kDa, that increased dramatically in level with the onset of stationary phase in *T. zilligii*. SRETzi was enriched in the acid-soluble fraction of total cytoplasmic proteins isolated from early stationary phase cells. The N-terminal sequence of SRETzi strongly suggests that it is a homologue of a family of highly conserved, exclusively archaeal proteins, that has to date not been described. Furthermore, homologues of SRETzi were found to occur in every archaeal genome sequenced to date, which suggests that the protein is involved in a fundamental cellular function. The fact that SRETzi is enriched in the acid-soluble fraction of proteins, suggests that the protein is basic in character, and therefore possibly part of the nucleoid. The overall biochemical characteristics of members of the SRE family, as inferred from their amino acid sequences, will be examined in the following chapter.

SRBTzi is a protein with an estimated molecular weight of 10.5±0.5 kDa, that also increased significantly in level at the onset of stationary phase. Like SRETzi, SRBTzi was also considerably enriched in the acid-soluble fraction of proteins. Convincing homologues of SRBTzi were readily identifiable in many, but not all, of the complete archaeal genomes, as well as in the bacterial genome of *Clostridium acetobutylicum* and in the archaeal gene-rich genome of the thermophilic bacterium, *Thermotoga maritima*. Recently, the complete solution structure of a homologue of SRBTzi, MTH1175 from *Methanobacterium thermoautotrophicum*, was determined by NMR-spectroscopy (Cort *et al.* 2000). The structure showed that MTH1175 adopts an α /β topology with a single mixed β-sheet, and contains two flexible loops and an unstructured C-terminal tail. Although the researchers found that the fold resembled that of Ribonuclease H and similar proteins, it differed from these in several important respects, and consequently, they concluded that it was unlikely that MTH1175 would exhibit nuclease activity. The structure of MTH1175 is discussed in more detail in Chapter Two.

SRCTzi is a protein with an estimated molecular weight of 9.0±0.5 kDa, that was found to occur at relatively high levels during log-phase, then proceeded to decline steadily as the age of the culture progressed, until it became undetectable after 266 h of incubation. SRCTzi was purified to near homogeneity by heparin-affinity chromatography, which consequently allowed the accurate determination of the first 20 residues from the N-terminus. Accordingly, homologues of SRCTzi were identifiable in GenBank with a very high degree of certainty. SRCTzi was found to be a member of the exclusively archaeal Sac10b family of DNA-binding proteins. Members of the Sac10b gene family have been identified in every archaeal genome sequenced to date (Forterre *et al.* 1999).

The DNA-protein complexes formed by Sac10b, isolated from the thermoacidophilic crenarchaeon *Sulfolobus acidocaldarius*, have been examined by electron microscopy (Lurz *et al.* 1986). Using a double-stranded DNA template, Sac10b was found to first bind cooperatively to both strands forming long, braided structures, and then, at higher protein:DNA ratios, the DNA strands were separated and a new structure was formed that evenly covered only one strand. Since neither structure showed any significant contraction in the length of the DNA, Sac10b was not considered to be involved in DNA packaging, but was suggested rather to play a role similar to that of RecA in recombination and repair processes.

The involvement of members of the Sac10b family in the stationary-phase response has to date not been examined. Consequently, the results presented here represent the first evidence of stationary-phase responsiveness in a member of the Sac10b family. Although the actual function of the Sac10b proteins remains unknown, the observation that the level of SRCTzi decreases as the culture ages, supports the notion that the Sac10b proteins are not involved in DNA-packaging or protection, as these capabilities would presumably still be required in stationary phase. Consequently, it would seem more likely that the Sac10b proteins play an active role in the nucleoid during cell growth, such as that suggested by Lurz and co-workers.

Finally, SRDTzi is another very low molecular weight protein, estimated at approximately 7.5±0.5 kDa, that also increased markedly in level as the culture aged. Although significantly enriched in the AWE, concentrations of SRDTzi suitable for N-terminal sequence analysis were obtained using ultrafiltration to remove the majority of the high molecular weight proteins and so concentrate those of low molecular weight. Convincing similarity to SRDTzi was found in a hypothetical protein identified in the complete genome sequence of *Pyrococcus furiosus*, PF1893. Interestingly, although PF1893 has been annotated in GenBank as a member of the AsnC family of transcriptional regulators, on closer inspection it was found that AsnC proteins are approximately twice the length of PF1893, and that PF1893 in fact only resembled their C-terminal portion. The AsnC protein family are also similar to the Lrp protein family. As will be shown in the following chapter, homologues of these "truncated" Lrp/AsnC proteins are found in the genomes of many other archaea and some bacteria.

Lrps have been extensively studied for their involvement in the stress-response in both bacteria and archaea, and can act as both activators or repressors on a number of different genes and operons (Newman and Lin 1995). Moreover, Lrps can also act as chromosomal organisers, inducing conformational changes in DNA and promoting the formation of higher-order DNA protein complexes (Wang and Calvo 1993). Although the precise DNA-binding mechanism of Lrps is unknown, mutational analysis of *E. coli* Lrp indicates that the DNA-binding region is in the N-terminal 40% of the protein and that the remaining 60% at the C-terminus contains a transcriptional activation domain and an overlapping leucine-response domain (Platko and Calvo 1993). These observations have since been supported by the crystal structure of LrpA, an Lrp protein from *P. furiosus*, which showed that the protein consists of two distinct domains; the N-terminal domain consisting of an

helix-turn-helix motif and the C-terminal domain consisting of a mixed α/β topology (Leonard *et al.* 2001). The suggestion that the C-terminal domain of the Lrp proteins is involved in transcriptional activation, raises the issue as to what role such a domain could play in the absence of a DNA-binding motif, or indeed whether it operates somehow in concert with other proteins to exert its activity. Finally, the phylogenetic relationship between the *srd* genes and the AsnC/Lrp family is also a matter of considerable interest; that is, did an ancestral gene of the *srd* family directly give rise to the AsnC/Lrp family, or is the SRD family a late adaptation that arose from a truncation of an AsnC/Lrp gene(s)? To fully address these questions, a further analysis of the function of the SRD proteins will be required.

To conclude, the screening approach of targeting low molecular weight proteins that vary with growth to identify proteins involved in transcriptional regulation, chromosomal organisation or protecting DNA proved to be successful. Two of the identified proteins were found to exhibit significant similarity with DNA-binding proteins characterised in other microorganisms. Since, the remaining two proteins belonged to hitherto undescribed protein families, further research will be required to examine their DNA-binding potential and their possible involvement in transcriptional regulation.

In the following chapter, the complete sequence analysis of the SRB, SRC, SRD and SRE gene families is presented. The examination of the phylogeny, DNA sequence elements, gene organisation, as well as the predicted biochemical properties of the proteins, is expected to provide valuable insights into the role of these genes.

CHAPTER TWO

Computational Analysis of the *srb*, *src*, *srd* and *sre* Gene Families

Summary

The availability of numerous completely sequenced archaeal and bacterial genomes, in conjunction with increasingly sophisticated bioinformatical tools, has made the computational analysis of DNA sequence data a pertinent aspect in the understanding and prediction of gene function. In the previous chapter, the identities of four stress response proteins, SRBTzi, SRCTzi, SRDTzi and SRETzi, were determined in *Thermococcus zilligii*. The proteins were found to belong to four different largely undescribed protein families, which were tentatively named as SRB, SRC, SRD and SRE and encoded by the *srb*, *src*, *srd* and *sre* gene families. This chapter describes the identification and computational analysis of the four gene families and the proteins they encode.

A total of 21 members of the *srb* family (COG1433) were identified from nine archaeal and two bacterial genomes, with the majority originating from members of the Euryarchaeota. Due to the sporadic distribution of *srb* genes amongst extant life, it is suggested that the *srb* genes have undergone horizontal transfer across large phylogenetic distances on several occasions. Intriguingly, a sequence homologous to *srb* family members was identified as part of a hypothetical ORF in *Clostridium perfringens*, which encodes a protein that effectively consists of a fusion of a transcriptional regulator belonging to the ArsR family, and an SRB homologue. In addition, several *srb* genes were found to occur as part of suspected operons. The majority of these putative operons included genes homologous to the MinD family of P-loop ATPases, which are characteristically known for their role in cell-division. Although no functional characterisation has been performed on any members of the

recently been determined (Cort *et al.* 2000). However, the structure of SRBMth1 does not currently provide any firm evidence for its possible function.

The SRC family was found to be equivalent to the highly conserved Sac10b family of DNA-binding proteins, which were previously described as being specific and ubiquitous amongst the Archaea (Forterre *et al.* 1999). However, in this study, although genes encoding SRC were found to occur exclusively amongst the Archaea, *src* genes were absent in the archaeal genomes of *Halobacterium* sp. NRC-1, *Methanopyrus kandleri*, *Methanosarcina mazei* and *M. acetivorans*. In total, 18 members of the *src* family (COG1581) were identified from 13 different species of archaea. The consensus secondary structure of the SRC proteins was determined, and found to consist of a relatively long α-helix flanked by two small β-strands, followed by a second α-helix, and then three further β-strands.

In the previous chapter, the closest homologues of SRDTzi were found to be similar to the C-terminal half of members of the Lrp/AsnC family of transcriptional regulators (COG1522) and consequently the SRD proteins are classified as also belonging to this group. The analysis of the Lrp/AsnC family revealed a total of 29 genes that encode only these reduced proteins. This subgroup was defined as the *srd* family. Mutational studies combined with the crystal structure of LrpA, a characterised Lrp protein from *Pyrococcus furiosus*, have indicated that the N-terminal half consists of a helix-turn-helix motif that is critical for DNA-binding (Leonard *et al.* 2001). The C-terminal half consists of a separate domain, of α/β topology, that is primarily involved in dimerisation. Consequently, it is suggested here that the SRD family forms a logical subgroup that is structurally and, almost certainly, functionally distinct to the other members of the Lrp/AsnC family. An examination of the phyletic distribution of the *srd* genes revealed that they occur in all non-methanogenic archaea, as well as in a few sparsely distributed bacteria. This led to the suggestion that the *srd* family originated in the archaeal lineage and was later transferred to members of the Bacteria. By inspecting the positions of the genes neighbouring *srd* in the various genomes, it was found that *srd* genes often appear as part of operon-like arrangements. Since the functions of the putative operons seemed to be unrelated to one another, it is suggested that the SRD proteins are involved in a relatively generic function, such as transcriptional regulation.

Genes encoding SRE were identified exclusively amongst members of the Archaea and occurred in every archaeal genome sequenced to date, with the exception of those of *Pyrobaculum aerophilum* and *Halobacterium* sp. NRC-1. The sequences of the SRE

family (COG1698) were found to be highly conserved, with just over 50% of amino acid residues being similar across all SRE proteins. The distance tree constructed from the SRE sequences was topologically similar to the 16S rRNA distance tree of the corresponding organisms, suggesting that the *sre* genes had not undergone any significant degree of horizontal transfer, or that the evolutionary constraints on the *sre* genes are similar to those of 16S rRNA genes. The consensus secondary structure computed from all the available SRE sequences, predicted that the protein consists of four large helical regions separated by short coiled regions. The predicted helix-turn-helix architecture of the SRE proteins suggests that the SRE proteins may bind DNA. Combined with the observation that SRETzi increases in level as the cells enter stationary phase, the SRE proteins are proposed as candidate regulators of gene expression in stationary phase.

The implications of the computational analyses of the *srb*, *src*, *srd* and *sre* gene families, and possible directions for future research, are discussed.

2.1 Introduction

In the previous chapter, four stress response proteins, named SRBTzi, SRCTzi, SRDTzi and SRETzi, were identified in the thermophilic archaeon, *Thermococcus zilligii*. The levels of three of these, SRBTzi, SRDTzi and SRETzi, were found to increase as the culture aged, and the level of SRCTzi was found to decrease. Using the N-terminal sequences determined for these proteins, homologues were successfully identified in the completely sequenced genomes of organisms closely related to *T. zilligii*, such as *Pyrococcus abyssi* and *P. furiosus*. As part of a genome-wide survey of protein structures in *Methanothermobacter thermautotrophicus*, the solution structure of MTH1175, a member of the SRB family, has been determined (Cort *et al.* 2000). However, despite some resemblance to the structure of Ribonuclease H and similar structures, the researchers concluded that MTH1175 was not likely to have nuclease activity and were not able to assign a function to the protein. SRCTzi was found to be homologous to the Sac10b family of DNA-binding proteins, which were first described in *Sulfolobus acidocaldarius* (Dijk and Reinhardt 1986). Although the role of the Sac10b proteins remains unknown, they have been shown to bind to DNA and form regular non-compacted structures (Lurz *et al.* 1986), and have also demonstrated a capacity to constrain DNA in negative supercoils (Xue *et al.* 2000). Finally, to date, no research has been published on any member of either

the SRD or SRE families, although the most similar homologues of SRDTzi were found to exhibit significant sequence similarity with the C-terminal half of the leucine responsive regulatory protein (Lrp) family.

With the advent of high-throughput DNA sequencing technology, the determination of complete genome sequences has become feasible (Fleischmann *et al.* 1995; Bult *et al.* 1996). As of May 2002, worldwide genome sequencing efforts have completely sequenced and annotated 75 prokaryotic and six eukaryotic genomes, and the genomes of approximately 200 other species are currently in progress, most of which are expected to be completed within the next five years (The Institute for Genomic Research, http://www.tigr.org/tdb/mdb/). With the aid of modern computational tools, a wide range of information can be inferred about a family of genes and the proteins they encode. Amino acid sequence data can be used to predict not only the structure and a range of biochemical characteristics of a family of proteins, but also provide insights into their possible functions. Amino acid sequences from related proteins can also be used to infer the evolutionary history of a family of proteins and indicate potential instances of horizontal gene transfer (Koonin *et al.* 2001). As well as providing further valuable information about the history of a gene family, the nucleotide sequences can also provide indicators to the transcriptional control of a gene, their organisation in the genome, and highlight possible functional relationships with other genes.

In this chapter, the results of a comprehensive computational analysis of the *srb*, *src*, *srd* and *sre* gene families, and the proteins they encode, are presented. Some of the results from this chapter were published in *Biochimica et Biophysica Acta* (Dinger and Musgrave 2000) (Appendix B) and *Astrobiology* (Appendix C), and presented at the *Annual Queenstown Molecular Biology Meeting* (Appendices E, F and G).

2.2 Materials and Methods

2.2.1 Identification of homologous genes

Homologues of the genes of interest were identified using a combination of the COGnitor program at NCBI (Tatusov *et al.* 1997; Tatusov *et al.* 2001) and GenBank BLAST searches (Altschul *et al.* 1997). COGnitor was used to retrieve the initial list of genes predicted to be orthologous to the target gene, then BLASTP queries were performed at GenBank to locate further homologues from genomes that were not represented in the

COG database. The target genes were found to most predominantly occur in archaea. A list of the archaeal genomes that were available in the GenBank and COG databases at the time of this study (May 2002) is shown in Table 2-1. Finally, TBLASTN searches were used to identify possible genes that had not been annotated in the genome data at GenBank. The corresponding open reading frames (ORFs) were located with the assistance of Genamics Expression 1.1 (Genamics).

Each gene was examined to check that the ORF that had been identified in the annotated genome data was indeed the most convincing. This was achieved by considering a number of factors, including the distance from or degree of overlap with neighbouring genes, the position of the putative ribosome binding site and transcription initiation elements, and the relative consensus start positions of the other homologues of the gene being examined. The sequence of the ribosome binding site was inferred from the sequence of the 3' end of the small subunit (16S) ribosomal RNA gene. In archaeal sequences, potential transcription initiation sites were identified by looking for AT rich regions, particularly those that corresponded best to the TATA box consensus sequence, TTTAWA (Reiter *et al.* 1990). These are positioned 25-30 bp upstream of a transcriptional initiator element. Transcription initiator elements are typically indicated in members of the Archaea by a pyrimidine-purine dinucleotide and are positioned approximately 20-25 bp upstream of the start codon (Hain *et al.* 1992).

Organism	*GenBank Accession Number*	*COG database?*	*Reference*
Archaeoglobus fulgidus	NC_000917	Yes	(Klenk *et al.* 1997)
Aeropyrum pernix	NC_000854	Yes	(Kawarabayasi *et al.* 1999)
Halobacterium sp. NRC-1	NC_002607	Yes	(Ng *et al.* 2000)
Methanococcus jannaschii	NC_000909	Yes	(Bult *et al.* 1996)
Methanopyrus kandleri	NC_003551	No	(Slesarev *et al.* 2002)
Methanothermobacter thermautotrophicus	NC_000916	Yes	(Smith *et al.* 1997)
Pyrococcus abyssi	NC_000868	Yes	(Unpublished)
Pyrobaculum aerophilum	NC_003364	No	(Fitz-Gibbon *et al.* 2002)
Pyrococcus furiosus	NC_003413	No	(Maeder *et al.* 1999)
Pyrococcus horikoshii	NC_000961	Yes	(Kawarabayasi *et al.* 1998)
Sulfolobus solfataricus	NC_002754	No	(She *et al.* 2001)
Sulfolobus tokodaii	NC_003106	No	(Kawarabayasi *et al.* 2001)
Thermoplasma acidophilum	NC_002578	Yes	(Ruepp *et al.* 2000)
Thermoplasma volcanium	NC_002689	Yes	(Kawashima *et al.* 2000)

Table 2-1. Archaeal genome sequences available from GenBank at the time of this study (May 2002). The genomes upon which the current version of the Clusters of Orthologous Groups (COG) database is based are also indicated.

Finally, the genes and their corresponding conceptual proteins were named according to their family name (i.e. SRB, SRC, SRD or SRE) and a three letter abbreviation of the species the gene originated from. In cases where more than one homologue of the gene was found in the same genome, the genes were essentially numbered arbitrarily, although likely corresponding homologues in other genomes were allocated the same number where possible.

2.2.2 Multiple alignment

The predicted protein sequences from each group of homologous genes were initially aligned using ClustalX 1.81 (Thompson *et al.* 1994; Thompson *et al.* 1997), with default parameters. The alignments were then imported into Microsoft Excel 2002 (Microsoft Corporation), where the sequences were further aligned manually. Residues were considered similar if at least 80% were either hydrophobic non-polar (A, V, I, L, P, M, F, W), polar non-charged (G, S, T, C, Y, N, Q), basic (H, R, K), or acidic (E, D) (Lehninger 1982; Karlin and Ghandour 1985).

The multiple alignments were used for the phylogenetic analyses and consensus secondary structure predictions.

2.2.3 Phylogenetic analysis

Phylogenetic trees were constructed from the multiple alignments of the complete protein sequences using neighbour-joining analysis. Neighbour-joining was performed using ClustalX 1.81 (Thompson *et al.* 1994; Thompson *et al.* 1997), with 100 bootstrap resamplings. The trees were rendered using TreeView for Windows 1.6.6 (Institute of Biomedical and Life Sciences, University of Glasgow).

2.2.4 Secondary structure prediction

The consensus secondary structure was predicted using the multiple alignments of protein sequences, with the neural network prediction algorithm, Jnet (Cuff and Barton 1999). Jnet was run using the Jpred 2 server (Cuff *et al.* 1998; Cuff and Barton 2000) at the European Bioinformatics Institute. The resulting data was reformatted and aligned with the original multiple alignments using Microsoft Excel 2002 (Microsoft Corporation). In the case of

the SRB family, the secondary structure was inferred from the tertiary solution structure, which was available for one of its members, and aligned appropriately.

2.3 Results

Part A: Computational Analysis of the *srb* Family

2.3.1 The *srb* genes occur primarily amongst the Euryarchaeota

Homologues of SRBTzi were identified using a combination of the COGnitor program and BLASTP searches at GenBank (NCBI). The proposed open reading frames for each of the genes was then examined to verify that the most substantive start position had been correctly identified. A total of 21 SRB homologues were identified from nine different archaeal genomes and two bacterial genomes. A complete list of all identified members of the SRB family is shown in Table 2-2.

To identify the potential homologues of SRBTzi, the theoretical amino acid sequence of the predicted SRBTzi homologue, PF2025 (named SRBPfu) from *P. furiosus*, was used to

Protein Name	*Originating Organism*	*Original Gene ID*	*Size (aa)*	*MW (Da)*	*pI*	*Genome Reference*
SRBAfu1	*Archaeoglobus fulgidus*	AF2378	144	14814	10.1	(Klenk *et al.* 1997)
SRBAfu2		AF2377	118	12661	6.0	
SRBAfu3		AF0668	101	11230	7.9	
SRBCac	*Clostridium acetobutylicum*	CAC3167	118	12882	6.6	(Nolling *et al.* 2001)
SRBMja1	*Methanococcus jannaschii*	MJ0580	109	12084	6.7	(Bult *et al.* 1996)
SRBMja2		MJ0327	101	11539	5.2	
SRBMth1	*Methanothermobacter*	MTH1175	124	13160	9.5	(Smith *et al.* 1997)
SRBMth2	*thermautotrophicus*	MTH1177	116	12779	5.3	
SRBPab1	*Pyrococcus abyssi*	PAB1692	194	20725	10.0	Unpublished
SRBPab2		PAB0517	138	14477	9.0	
SRBPfu	*Pyrococcus furiosus*	PF2025	179	19002	9.5	(Maeder *et al.* 1999)
SRBPho1	*Pyrococcus horikoshii*	PH1227	189	20086	10.3	(Kawarabayasi *et al.* 1998)
SRBPho2		PH1383	137	14528	9.2	
SRBPho3		PH0880	116	12491	8.7	
SRBPho4		PH0822	121	13668	6.1	
SRBSto	*Sulfolobus tokodaii*	ST2018	122	13769	6.2	(Kawarabayasi *et al.* 2001)
SRBTac1	*Thermoplasma*	Ta1041m	117	12493	5.8	(Ruepp *et al.* 2000)
SRBTac2	*acidophilum*	Ta1370	115	12634	5.9	
SRBTma1	*Thermotoga maritima*	TM1290	116	12538	6.2	(Nelson *et al.* 1999)
SRBTma2		TM1816	124	13625	4.9	
SRBTvo	*Thermoplasma volcanium*	TVN0554	119	13151	6.2	(Kawashima *et al.* 2000)

Table 2-2. Summary information for SRB homologues, including originating organism, original gene name in the annotated genome data, predicted size in amino acids, theoretical molecular weight (MW), and theoretical pI. Theoretical MW and pI were determined using the ProtParam tool (ExPASy).

query the COG database at NCBI. SRBPfu was found to be very closely related to members of COG1433, which contained a total of 21 genes as of May 2002. Since the *P. furiosus* genome was not actually included in the COG database at the time of the analysis, PF2025 itself is not formally a member of COG1433. A closer inspection of the genes from COG1433, revealed that three of the genes, *af2388*, *vc1323*, *mth1172*, possessed only weak similarity to the other members of the group. Furthermore, the products of *af2388* and *mth1172* were predicted to be 384 and 407 amino acid residues in length and were predicted to function as a transmembrane transporter and a possible cation transporter, respectively. *vc1323* had no proposed function, but was almost as closely related to members of two other COGs as it was to its most similar matches in COG1433 (results not shown). Consequently, these three members were not included as members of the SRB family.

To identify members of the SRB family from organisms that were not included in the COG database, BLASTP and TBLASTN searches were performed against the non-redundant database at GenBank (NCBI). The BLASTP searches revealed a further two *srb* homologues, *cac3167* and *st2018*, originating from *Clostridium acetobutylicum* and *Sulfolobus tokodaii*, respectively. Neither of these organisms are currently included in the COG database. TBLASTN searches did not reveal any further significant matches that were not already identified in the annotated genome data.

Since the genome data used was largely annotated by electronic means using relatively simplistic criteria, the proposed open reading frame for each of the genes was examined. Regions upstream of the proposed start codons were analysed for likely ribosome binding sites, transcription initiation elements, and for the presence of coding regions of neighbouring or overlapping genes (Figure 2-1). It was found that the start codon for three of the *srb* genes, *srbMja1*, *srbMja2*, and *srbPho2,* had most probably been identified incorrectly in the data at GenBank (NCBI). As can be seen in Figure 2-1, the proposed start codons feature more likely ribosome binding sites. In addition, the resulting coding sequences are more consistent with those of the other SRB homologues. The data shown in Table 2-2 was based on the ORFs determined in this analysis. The examination of the upstream regions also revealed that most of the archaeal *srb* genes possessed potential TATA boxes at the expected position of approximately 50 bp upstream of the start codon, or 25-30 bp upstream of a potential initiator element. Those genes that did not possess significant transcription initiation elements, were found to be those that had the coding

significant transcription initiation elements, were found to be those that had the coding region of the previous gene extending very close to the start codon. In these cases, it is suspected that *srb* gene is part of an operon (see 2.3.5).

The phyletic distribution of the genes amongst extant life was examined. The results were illustrated by highlighting genera that possessed at least one copy of the *srb* gene on a universal phylogenetic tree, based on small subunit ribosomal RNA sequences, indicating genera in which the genome of at least one member is known (Figure 2-2). Of the completely sequenced genomes available at the time of this study, 9 out of 15 archaeal genomes and 2 out of 60 bacterial genomes possessed at least one copy of *srb*. Furthermore, amongst the archaeal genomes that possess *srb*, eight of the nine belong to a cluster within Euryarchaeota. A search of the almost complete archaeal genome sequences of *Methanosarcina acetivorans, M. mazei* (members of the Methanosarcinales order) and *Ferroplasma acidarmanus* (a member of the Thermoplasmatales order) revealed that homologues were also present in each of these organisms. In light of the near absence of *srb* in both bacterial and crenarchaeal genomes, it would appear likely that the progenitor of *srb* first occurred after the divergence of the Euryarchaeota and Crenarchaeota. However, to add further credence to this suggestion, genome sequence data from a greater diversity of crenarchaeotes will be required.

The diverse distribution of the remaining genera in which *srb* occurs, leads to the hypothesis that the gene was passed to them by horizontal transfer. Indeed, it is especially noteworthy that *srb* is present in *Sulfolobus tokodaii*, but not in *S. solfataricus*. In the *Thermotoga maritima* genome, approximately 30% of the genes have homologues with

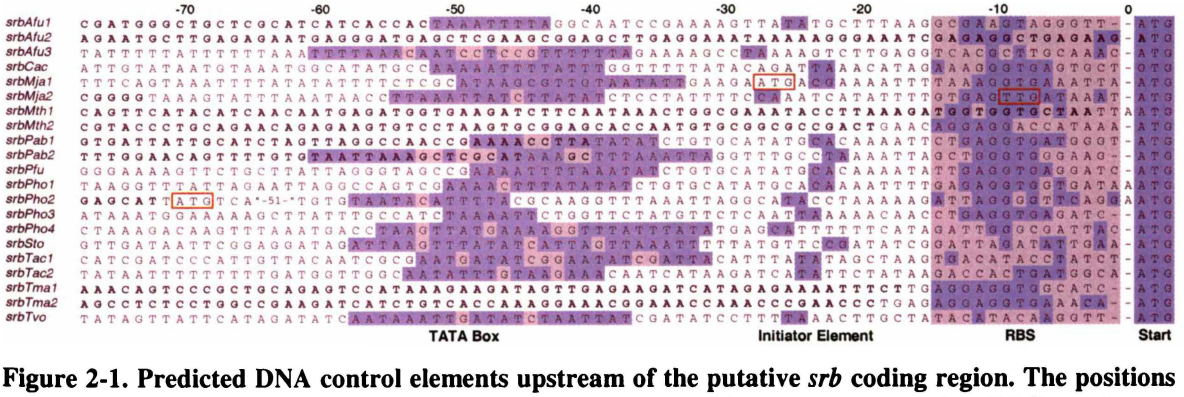


Figure 2-1. Predicted DNA control elements upstream of the putative *srb* coding region. The positions of the TATA box (in archaeal sequence), initiator element, ribosome binding site (RBS), and start codon (Start) are indicated. The regions are indicated by shading, with dark shading indicating identity with the consensus sequences for the respective control elements. Previously identified start codons, as predicted in the annotated genome data in GenBank (NCBI), are outlined in red. Coding regions of neighbouring or overlapping genes are in bold print. See Table 2-2 for the associated organism names and references.

archaeal genes, a substantially greater proportion than other bacterial genomes, and consequently it has been suggested to have been the recipient of numerous archaeal genes by horizontal transfer (Nelson *et al.* 1999). Interestingly, although the genome of the close relative of *Clostridium acetobutylicum*, *C. perfringens*, does not appear to contain a direct homologue of *srbCac*, it does contain a hypothetical gene encoding for a 244 aa protein, which has significant homology with some members of the *srb* family (see 2.3.4). No identifiable homologues of *srb* were found in the sequenced genomes of the next closest relatives of the Clostridia (according to 16S rRNA sequences), *Bacillus subtilis* and *B. halodurans*.

Finally, a phylogenetic pattern search using the COGnitor program, revealed that no other COG possesses an equivalent phylogenetic distribution to the *srb* gene family. This finding indicates that *srb* is not likely to be exclusively reliant on any other protein for its function, and unlikely to play a central role in a protein complex. Furthermore, this lack of reliance on other non-basal proteins is in accordance with the proposed capacity of the *srb* gene to

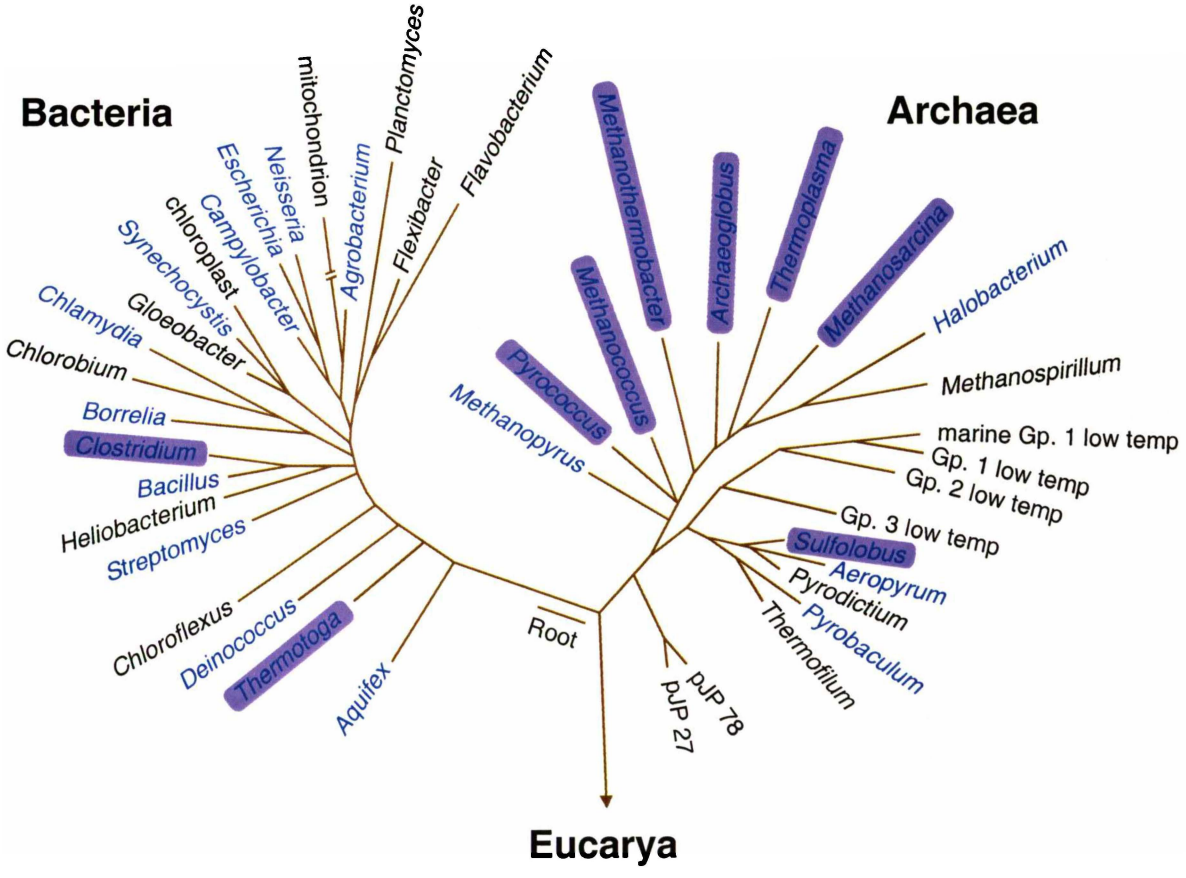


Figure 2-2. Phyletic distribution of the *srb* gene amongst extant life. Representatives from the Archaea and the Bacteria are shown on the universal phylogenetic tree based on small subunit ribosomal RNA sequences, as constructed by Norman Pace and colleagues (Pace 1997). Genera in which the complete genome has been sequenced for one or more of its members, at the time of this study, are shown in blue lettering. Genera in which one or more of its members is known to possess a gene encoding an SRB homologue are highlighted.

be transferred into the genomes of distantly related organisms.

2.3.2 *srb* genes are likely to have undergone multiple horizontal transfer events

To examine the relationship between the *srb* family members, the predicted amino acid sequences of the SRB proteins were aligned (Figure 2-4) and the resulting data used to construct a distance tree (Figure 2-3). Most of the branch points were able to be determined to a reasonable degree of statistical significance, as indicated by the majority of

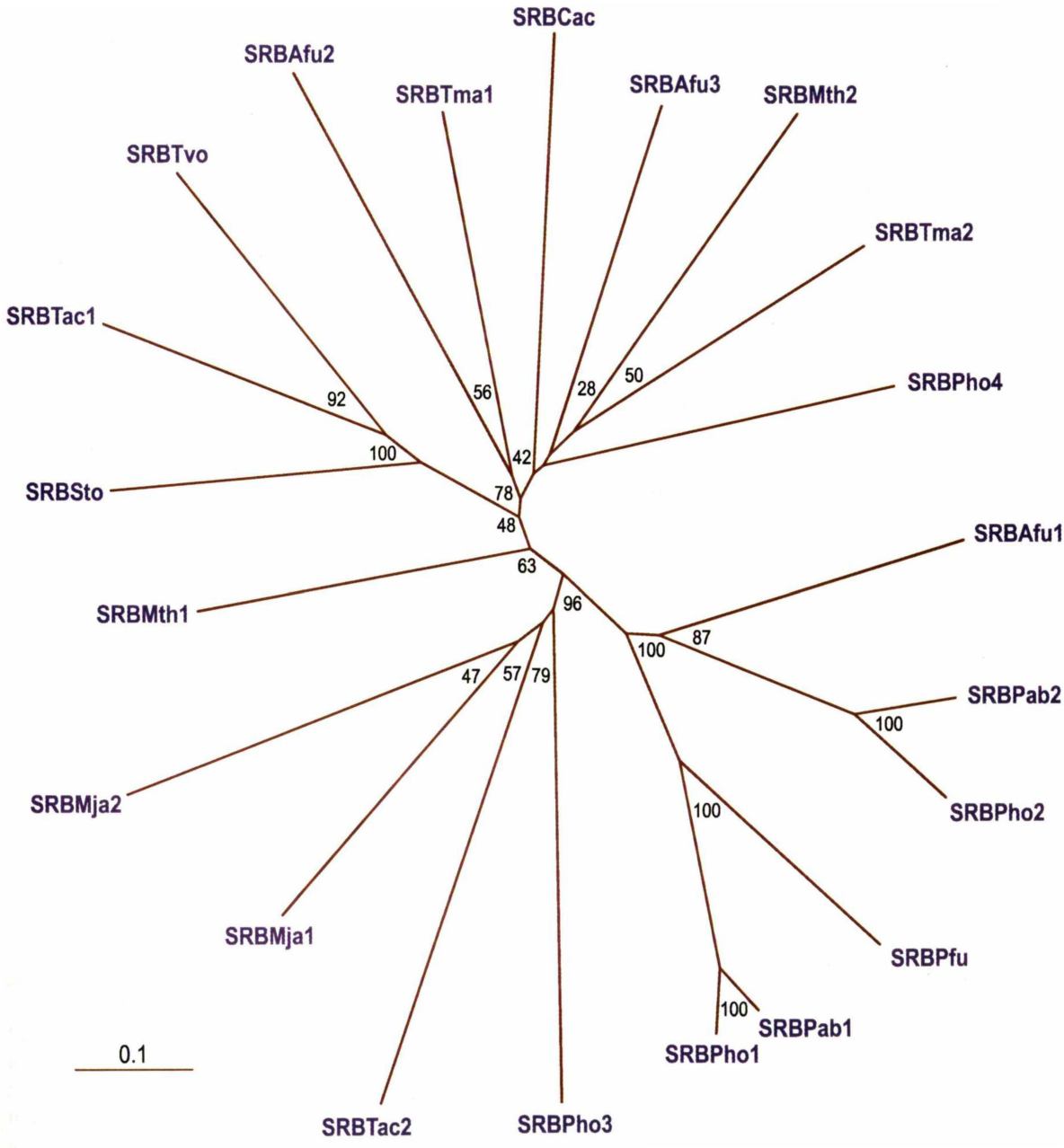


Figure 2-3. Distance tree of SRB protein sequences. The tree was inferred by neighbour-joining analysis. The percentage of 100 bootstrap resamplings that support each topological element is indicated at the branching points. The scale represents the expected number of amino acid changes per sequence position.

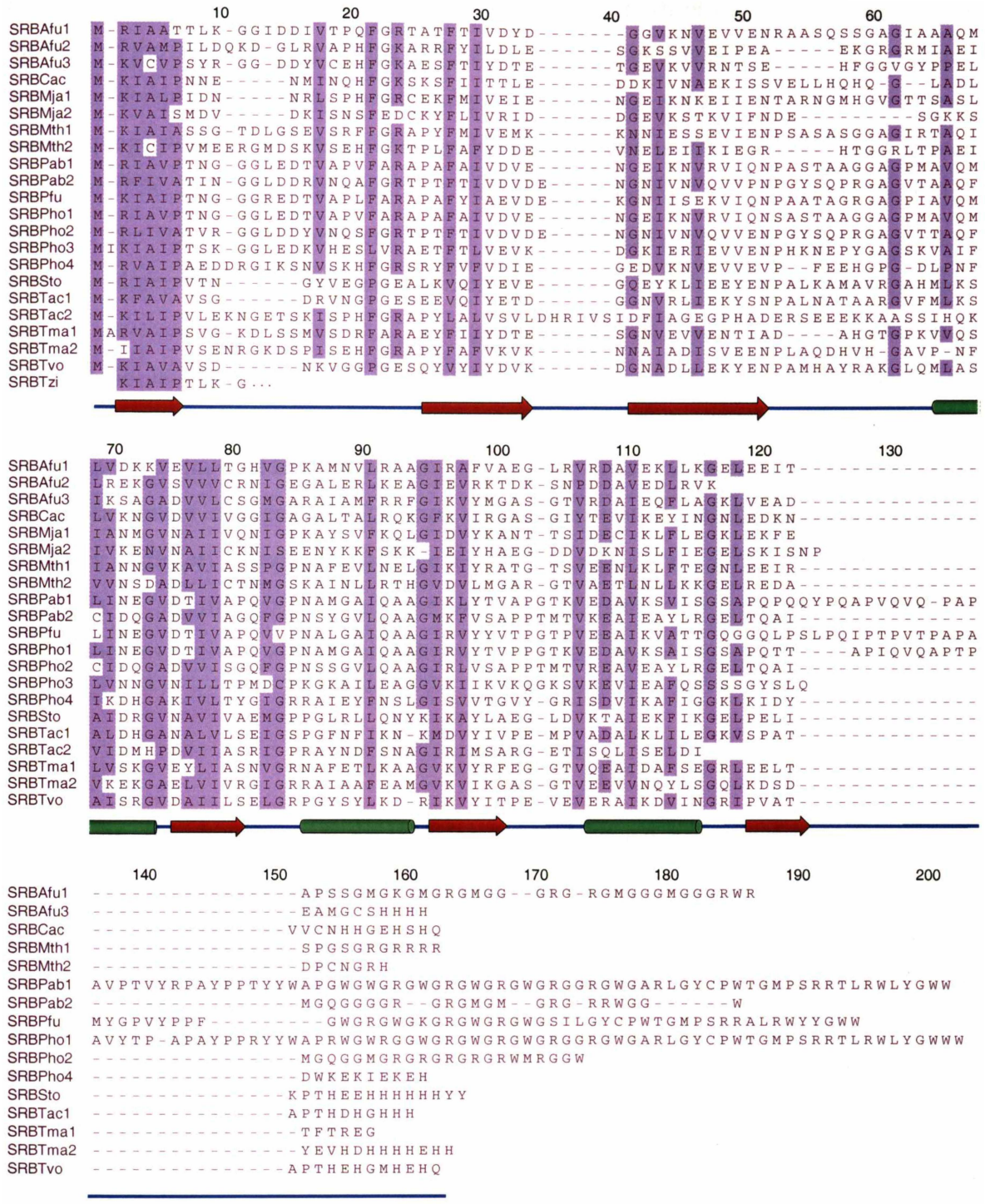


Figure 2-4. Multiple sequence alignment of predicted amino acid sequences of SRB and secondary structure of SRBMth1. Similar residues are shown highlighted. Residues were considered similar if they were hydrophobic non-polar (A, V, I, L, P, M, F, W), polar non-charged (G, S, T, C, Y, N, Q), basic (H, R, K), or acidic (E, D) (Lehninger 1982; Karlin and Ghandour 1985). See Table 2-2 for the associated organism names, original gene names, and references. The secondary structure of SRBMth1, determined by NMR (Cort *et al.* 2000), is aligned along the bottom of the sequences. α-helical residues, extended β-sheets and coiled regions are represented as green cylinders, red arrows and blue lines, respectively. The tertiary structure of SRBMth1 is shown in Figure 2-5.

bootstrap values of 50 and over. The close clustering of most of the branch points around the base of the tree, resulting in long individual branches, indicates that the SRB sequences are considerably divergent. Additionally, there is no support for any distinctive subgroupings within the family. The topology of the tree is almost entirely incongruent with that of the respective 16S ribosomal tree for these organisms. This observation further supports the notion that members of the *srb* gene family have undergone horizontal transfer. Moreover, in several cases, homologues of *srb* found within the same genome, such as *srbTac1* and *srbTac2* in *T. acidophilum* or *srbAful* and *srbAfu2*/*srbAfu3* in *A. fulgidus*, did not cluster together. This raises the possibility that these homologues are not in fact true paralogues. i.e. they are not the product of a gene duplication event, but rather a consequence of independent horizontal gene transfer events. However, the possibility that the genes simply changed rapidly after duplication, cannot be eliminated. Conversely, instances of likely paralogous *srb* genes are evident in *P. horikoshii* (*srbPho1* and *srbPho2*), *P. abyssi* (*srbPab1* and *srbPab2*), and *M. jannaschii* (*srbMja1* and *srbMja2*), where the respective pairs of homologues distinctly appear to group together on the distance tree.

2.3.3 The 3D structure of SRBMth1 resembles an RNase H motif

Although no functional studies of any SRB homologues have been performed previously, a search of the Molecular Modelling Database (MMDB) at the NCBI, indicated that the solution structure of SRBMth1 has recently been determined and analysed (Cort *et al.* 2000). The 3D structure data was used to compile a model of the protein (Figure 2-5), and the inferred secondary structure was aligned with the SRB multiple sequence alignment (Figure 2-3). The structure shows that SRBMth1 consists of a single mixed 6-stranded β-sheet flanked by three α-helices and contains two flexible loops and an unstructured C-terminal tail. Structural comparisons performed by Cort and coworkers, found that this folding topology most resembled the structures in the Ribonuclease H-like superfamily. In their analysis, they compared the SRBMth1 to three members of this family; RuvC resolvase (also known as Holliday junction resolvase), Ribonuclease H (RNase H), and the RNase H domain of HIV reverse transcriptase. In each case, the SRBMth1 structure lacked certain critical active regions found in these structures, which suggested that SRBMth1 was unlikely to exhibit any function analogous to these proteins. However, SRBMth1 does

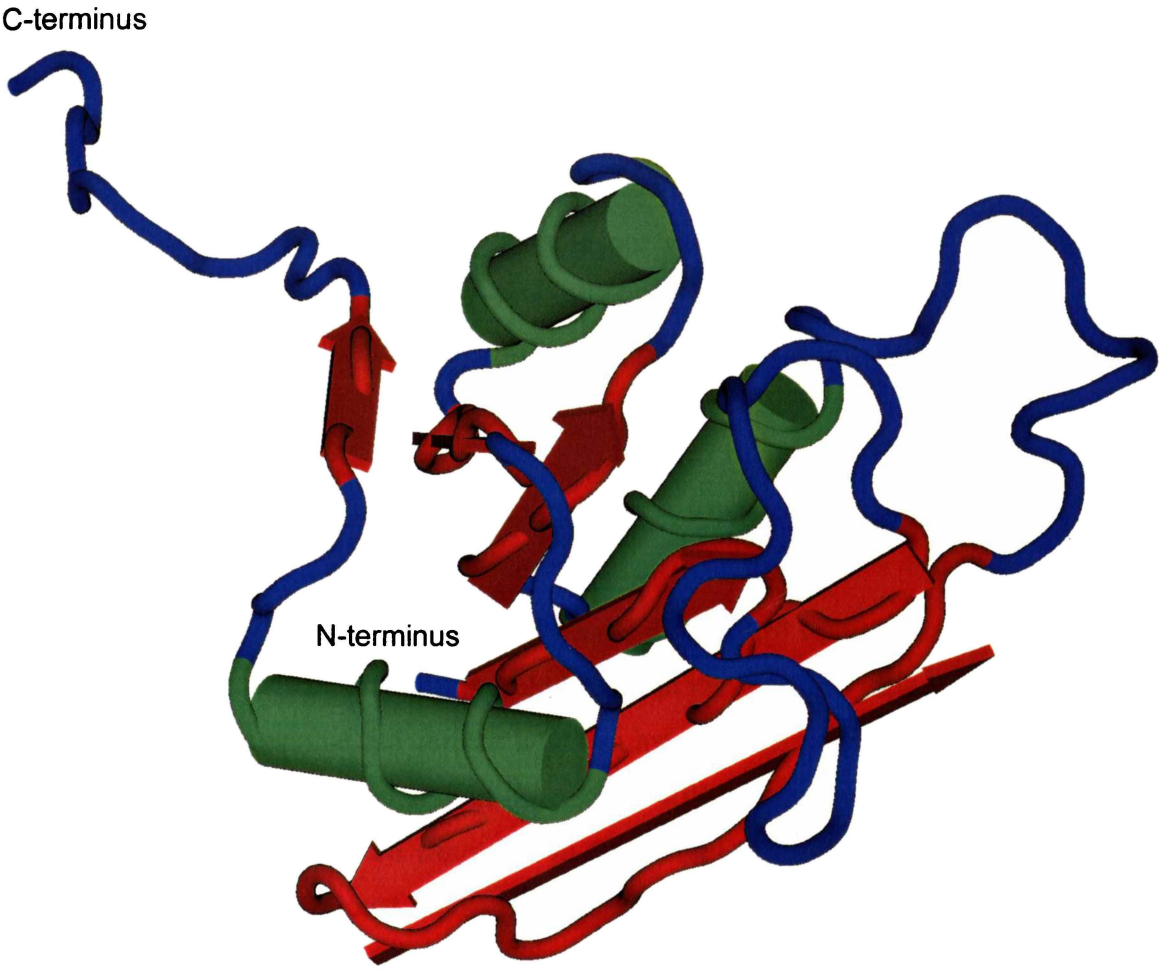


Figure 2-5. CN3D rendition of the solution structure of SRBMth1 (MTH1175). α-helical residues, extended β-sheets and coiled regions are represented as green cylinders, red arrows and blue curves, respectively. The C-terminus of the protein is indicated. Structural data was acquired from the MMDB at the NCBI (PDB Id 1EO1) and visualised with CN3D 4.0 (NCBI). The data originated from an NMR solution structure determination (Cort *et al.* 2000).

contain positively charged residues over its surface, similar to that seen the nucleic-acid binding region of RNase H. In addition, it is suggested that the distinctive large loop in the SRBMth1 structure forms a likely binding or active site (Cort *et al.* 2000).

Since the multiple alignment of the SRB sequences indicated that there was considerable variation between the primary structure of the homologues, the likelihood that the other SRB homologues would have a similar folding topology to SRBMth1 was assessed by examining some further theoretical biochemical properties of the proteins. First, the length, molecular weight and pI of the hypothetical proteins were calculated (Table 2-2). The calculations indicated that the SRB sequences vary considerably in size (11.5 to 20.7 kDa) and pI (4.9 to 10.3). However, this variability can be largely attributed to the arginine–rich C-terminal extensions that some SRB sequences possess, which contribute considerable size and basicity to the protein. When the calculations were performed excluding the

arginine-rich C-terminal regions, the theoretical pI ranged from 4.9 to 7.9. Second, using the neural network prediction algorithm, Jnet (Cuff and Barton 1999), the secondary structures of the SRB homologues were computed. The secondary structures of the SRB sequences were in good agreement with each other and consistently predicted the proteins to contain multiple β-sheet regions (results not shown). This was in reasonably good agreement with the actual structure determined for SRBMth1, with the most significant difference being the absence of substantial helical regions in the predicted structures.

2.3.4 A hypothetical ORF in *Clostridium perfringens* contains homologues of *arsR* and *srb*

Since the *C. acetobutylicum* genome contained a distinct *srb* homologue, it was somewhat surprising that *C. perfringens* did not. However, on closer examination, it was found that the *C. perfringens* genome did contain a 244 aa hypothetical protein (CPE1031), in which 108 residues in the C-terminal half possessed 40% similarity to either SRBMth2 or SRBPho4. CPE1031 had been annotated in GenBank (NCBI) as having similarity to the *arsR* family of transcriptional regulators. *arsR* is part of the arsenical resistance operon (*ars*), which has been studied in numerous bacteria, including *E. coli* (Carlin *et al.* 1995), *B. subtilis* (Sato and Kobayashi 1998), *Pseudomonas fluorescens* (Prithivirajsingh *et al.* 2001) and *Staphylococcus xylosus* (Rosenstein *et al.* 1992). *arsR* encodes a protein with a molecular weight of approximately 8 kDa, which has been suggested to function as a negative regulator of the *ars* operon (Rosenstein *et al.* 1992; Suzuki *et al.* 1998). BLAST searches at GenBank (NCBI) revealed that the closest match to CPE1031 was an ArsR homologue identified in the complete genome sequence of *B. subtilis* (yvbA), which shares 57% similarity over 87 residues at the N-terminal end of CPE1031. yvbA shares 53% similarity over 92 residues to the closest characterised ArsR protein, which was found in *S. xylosus*.

To show the relationship between CPE1031, ArsR and SRB, the amino acid sequences of ArsR from *S. xylosus*, CPE1031, yvbA and SRBMth2 were aligned (Figure 2-6). The figure illustrates that CPE1031 appears to consist of a fusion of an ArsR and a SRB homologue, with a spacer region of approximately 30 aa. To examine the likelihood that this coding region was not the result of a sequencing error, the nucleotide sequence of the space region was examined. Although several instances could be identified where a single

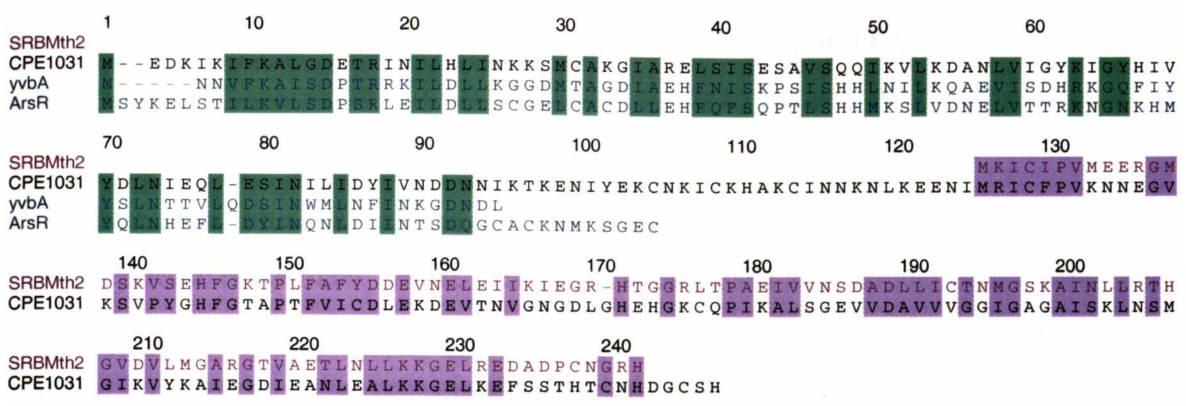


Figure 2-6. Multiple alignment showing the relationship between CPE1031 and SRBMth1, yvbA and ArsR. The sequences for CPE1031, SRBMth1, and yvbA were sourced from the annotated genome data of *Clostridium perfringens* (Shimizu *et al.* 2002), *M. thermautotrophicus* (Smith *et al.* 1997) and *Bacillus subtilis* (Kunst *et al.* 1997), respectively. ArsR is a negative regulator of the antimonite, arsenite and arsenate resistance operon of *Staphylococcus xylosus* ((Rosenstein *et al.* 1992)). Similar residues between SRBMth1 and CPE1031 are shaded purple, while similar residues between CPE1031, yvbA and ArsR are shaded green. Residues were considered similar if they were hydrophobic non-polar (A, V, I, L, P, M, F, W), polar non-charged (G, S, T, C, Y, N, Q), basic (H, R, K), or acidic (E, D) (Lehninger 1982; Karlin and Ghandour 1985).

base change would create a stop codon, there was no evidence of a terminator region or a substantive ribosome binding site for the SRB portion of the gene (results not shown). However, to confirm whether this fusion protein is actually synthesised, further research would need to be carried out.

To assess the likelihood that *cpe1031* is involved in arsenical resistance, the functions of genes in close proximity to *cpe1031* were examined and BLAST searches for other members of the *ars* operon were performed. An examination of the genes flanking *cpe1031*, indicated that no other members of the *ars* operon were nearby, and also that *cpe1031* was not likely to be part of an operon (results not shown). Using BLASTP searches, a likely *ars* operon was found in the *C. perfringens* genome, approximately 200 kb proximal to *cpe1031*, which contains coding regions for homologues of ArsR and arsenate reductase. Consequently, it would appear unlikely that *cpe1031* is involved in arsenical resistance, and it is suggested that the ArsR transcriptional regulatory domain has been recruited for another purpose.

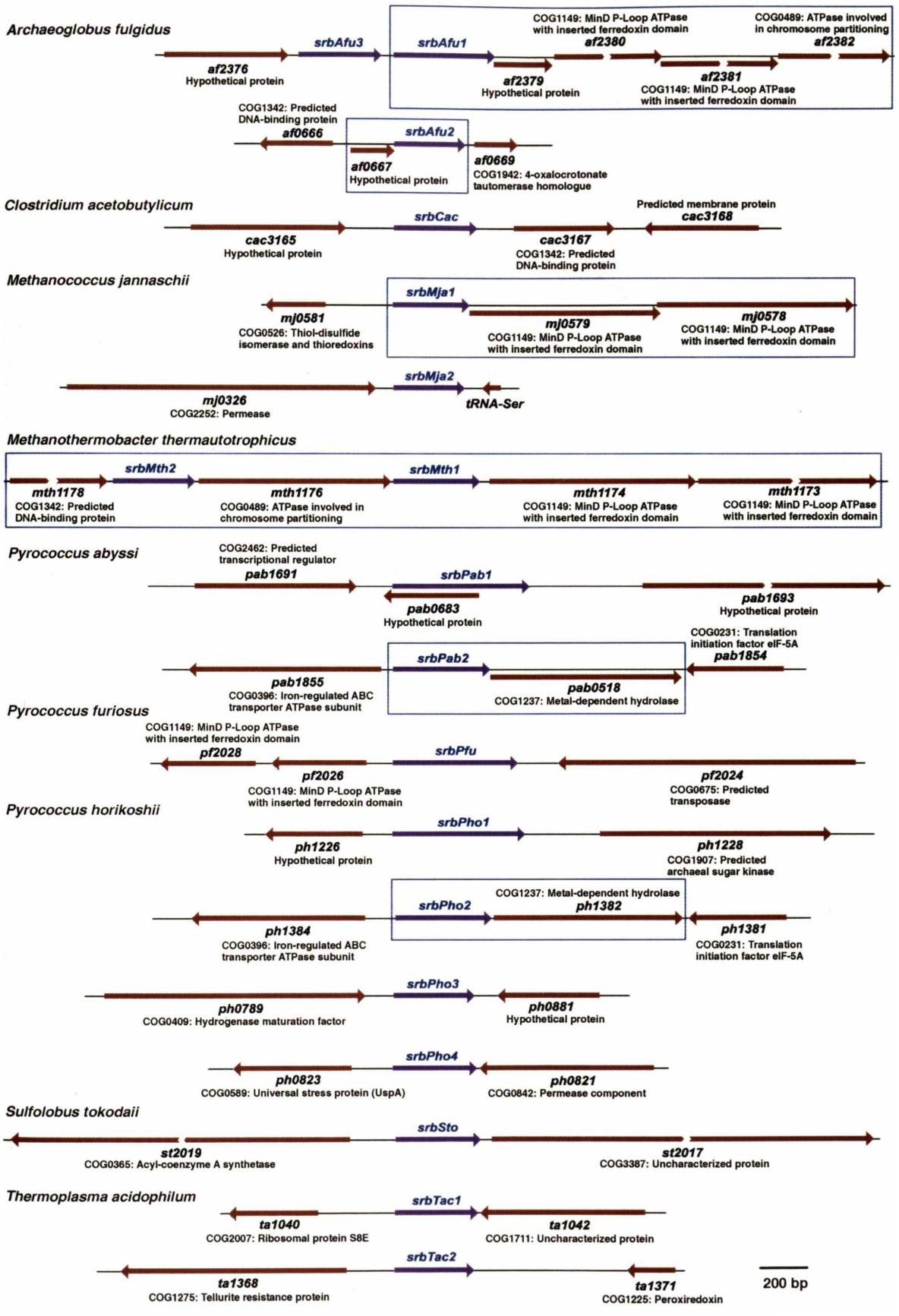
2.3.5 *srb* occasionally occurs in an operon-like arrangement with putative cell-division related genes

To gather insight into the possible role of SRB, the relative positions and functions of the genes neighbouring *srb* were studied in the available genomes. A graphical map comparing

the genomic neighbourhood of the various *srb* genes is shown in Figure 2-7. Two different gene families, COG1149 and COG1342, were found to regularly occur either adjacent or in close proximity to *srb*.

COG1149 is described in the COG database at NCBI as "a MinD P-loop ATPase with an inserted ferredoxin domain". MinD has been characterised as a critical component of the cell division machinery in many different species of bacteria and eukaryotes, and sequence analyses have reported that genes encoding homologues to MinD are found in virtually all living organisms (Shahrestanifar *et al.* 1994; Vitale *et al.* 1996; Bernander 1998; Gerard *et al.* 1998). In *E. coli*, MinD is localised in the inner membrane region of the cell envelope and although its precise role remains unknown, it has been shown to be intimately involved in organising the correct placement of the site for cell division (de Boer *et al.* 1991). The function of members of COG1149, however, has to date not been studied, and consequently the role of the inserted ferredoxin domain is unknown, or indeed whether or not members of COG1149 are actually involved in cell division. However, in a recent comprehensive attempt at classifying all known P-loop GTPases and related ATPases, members of COG1149 were placed in the Mrp/MinD-Etk superfamily (Leipe *et al.* 2002). Genes encoding members of COG1149 (hereafter referred to as *minDfer*) were found in the genomes of *A. fulgidus*, *M. jannaschii*, *M. thermautotrophicus*, *P. abyssi*, *P. horikoshii*, *P. furiosus*, and *T. maritima*.

A closer examination of the relative positions of the *srb* and *minDfer* genes, and their associated sequence elements, indicated that these genes were likely to be part of an operon. The start and end points of the operon were inferred by the spacing and/or orientation of the flanking genes, as well as the presence of potential promoter and terminator sites. The genes included in putative operons are indicated in Figure 2-7. The constituents and order of the genes in the operons varied from genome to genome. Operons containing *srb* and *minDfer* were identified in *A. fulgidus*, *M. jannaschii*, *M. thermautotrophicus*, and *T. maritima*. Interestingly, the *minDfer* gene was found to always occur in adjacent homologous pairs, even in *P. abyssi*, *P. horikoshii*, and *P. furiosus*, where the genes did not appear to be part of an operon. In both *A. fulgidus* and *M. thermautotrophicus*, the putative operon also included a gene encoding an ATPase homologue implicated in chromosome partitioning (COG0489). The functional connection between COG0489 and COG1149 sparks the hypothesis that this operon may have a role in cell division.



(continues on next page)

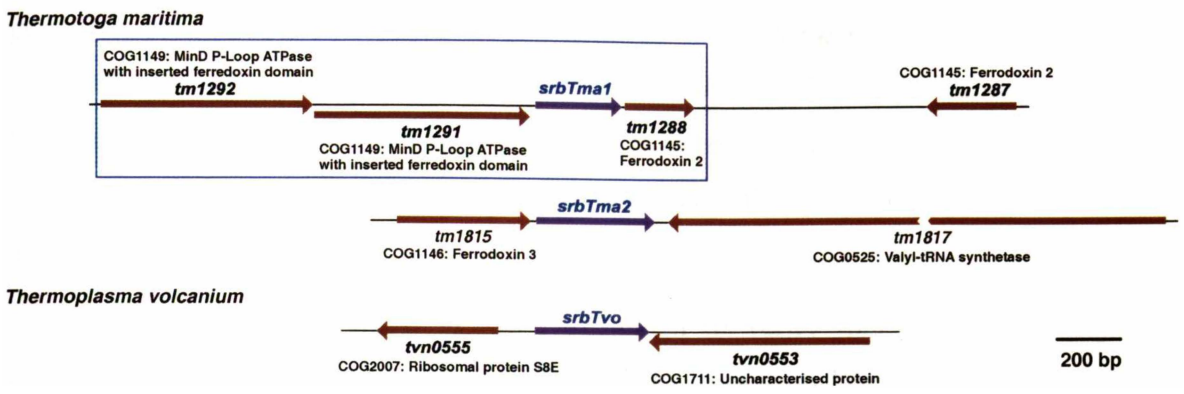


Figure 2-7. Comparison of relative positions, orientations and proposed functions of genes neighbouring or overlapping with *srb* in various genomes. The arrows represent the relative size, position and orientation of *srb* and the nearby genes. Overlapping genes are indicated by offset arrows. Functions for nearby genes were assigned using the Cognitor program at NCBI. Functions for genes that did not belong to any known COG were assigned according to the annotated data for the respective genomes from GenBank (NCBI). Genes are illustrated according to the scale shown, with the exception of those indicated with a break-point where the gene was resized to fit the figure. Boxed regions indicate groups of genes that are suspected to be part of an operon, as determined by the relative positions of the genes and the locations of putative promoter and terminator regions.

The other gene family that occurred in close proximity to the *srb* genes, were those encoding members of COG1342. COG1342 consists of a small family of closely-related proteins that are described in the COG database as "predicted DNA-binding proteins". At the time of this study, no research has been published regarding any members of this protein family. Genes encoding members of COG1342 were found adjacent to *srbCac*, *srbAfu2* and, *srbMth2*. Coincidentally, for *srbMth2*, the gene was the first member of the putative operon described above for *M. thermautotrophicus*. Genes encoding members of COG1342 are also found in *P. horikoshii*, *P. abyssi*, *P furiosus* and *M. jannaschii*. The fact that the *C. acetobutylicum* is the only bacterial genome sequenced to date to contain a gene encoding a member of COG1342, strongly suggests that it was transferred there by horizontal gene transfer, and moreover, given its close proximity to *srbCac*, it would seem likely that the two genes were transferred in tandem.

Part B: Computational Analysis of the *src* Family

2.3.6 *src* occurs exclusively amongst most archaeal genomes

Homologues of SRCTzi were identified using a combination of the COGnitor program and BLASTP searches at GenBank (NCBI). Initially, the theoretical amino acid sequence of the predicted SRCTzi homologue, PHS053 (named SRCPho) from *P. horikoshii*, was used to query the COG database at NCBI. SRCPho was found to belong to COG1581, which contained a total of ten genes as of May 2002. Using BLASTP searches, a further eight

genes encoding *src* homologues were identified in *P. furiosus*, *S. acidocaldarius*, *S. tokodaii* and *Pyrobaculum aerophilum*. TBLASTN searches did not reveal any further *src* homologues. A complete list of all identified members of the SRC family is shown in Table 2-3. A total of 18 SRC homologues were identified from 13 different species of archaea.

The ORFs identified in GenBank were examined to verify that the most likely start position had been correctly identified (Figure 2-8). Based on homology with other SRC sequences and the positions of the putative ribosome binding site, the predicted start positions for SRCMth and SRCSso1 were modified from those in GenBank. Moreover, the examination of the upstream regions also revealed that all of the *src* genes possessed probable TATA boxes at the expected position of approximately 50 bp upstream of the start codon, or 25-30 bp upstream of a potential initiator element (Figure 2-8).

The phyletic distribution of the *src* genes was examined, and the results illustrated by highlighting genera that possessed at least one homologue of *src* on a universal phylogenetic tree (Figure 2-9). Of the completely sequenced genomes available at the time of this study, 12 out of 14 archaeal genomes possessed at least one copy of *src*. In addition,

Protein Name	*Originating Organism*	*Original Gene ID*	*Size (aa)*	*MW (Da)*	*pI*	*Genome Reference*
SRCAfu1	*Archaeoglobus fulgidus*	AF1956	89	9894	6.6	(Klenk *et al.* 1997)
SRCAfu2		AF1067	89	9676	8.0	
SRCApe1	*Aeropyrum pernix*	APE1823	102	11380	9.1	(Kawarabayasi *et al.* 1999)
SRCApe2		APES060	98	10785	9.4	
SRCMja	*Methanococcus jannaschii*	MJ0212	87	9675	9.6	(Bult *et al.* 1996)
SRCMth	*Methanothermobacter thermautotrophicus*	MTH1483	91	9976	4.6	(Smith *et al.* 1997)
SRCPab	*Pyrococcus abyssi*	PAB7094	93	10352	8.0	(Unpublished)
SRCPae	*Pyrobaculum aerophilum*	PAE2226	92	9827	10.2	(Fitz-Gibbon *et al.* 2002)
SRCPfu	*Pyrococcus furiosus*	PF1881	93	10385	8.0	(Maeder *et al.* 1999)
SRCPho	*Pyrococcus horikoshii*	PHS053	93	10382	8.0	(Kawarabayasi *et al.* 1998)
SRCSsh1	*Sulfolobus shibatae*	T29101	97	10586	10.5	(Jaxel *et al.* 1996)
SRCSsh2		T29102	89	10265	9.4	
SRCSso1	*Sulfolobus solfataricus*	SSO0962	97	10586	10.5	(She *et al.* 2001)
SRCSso2		SSO6877	89	10239	9.4	
SRCSto1	*Sulfolobus tokodaii*	STS141	99	10772	10.5	(Kawarabayasi *et al.* 2001)
SRCSto2		STS142	90	10437	9.2	
SRCTac	*Thermoplasma acidophilum*	Ta0912	89	9938	9.0	(Ruepp *et al.* 2000)
SRCTvo	*Thermoplasma volcanium*	TVN1058	89	9966	9.1	(Kawashima *et al.* 2000)

Table 2-3. Summary information for SRC homologues, including originating organism, original gene name in the annotated genome data, predicted size in amino acids, theoretical molecular weight (MW), and theoretical pI. Theoretical MW and pI were determined using the ProtParam tool (ExPASy).

an *src* homologue was also identified in a cloned DNA fragment from *Sulfolobus shibatae*, which had been isolated for the sequencing of the reverse gyrase gene (Jaxel *et al.* 1996). BLASTP searches did not reveal significant homology with any eucaryal or bacterial sequences. A search of three other almost complete archaeal genome sequences, revealed that the *F. acidarmanus* genome contained one *src* homologue, but the genomes of *Methanosarcina acetivorans* or *M. mazei* did not possess sequences of significant similarity to *src*. A previous report on the Sac10b family, which is equivalent to the SRC family, stated that this family represented the first known, small, basic DNA-binding proteins that were both ubiquitous in archaea and specific to the domain (Forterre *et al.* 1999). The analysis of the genome data now available indicates that this finding has been partly disqualified: the SRC proteins remain specific to members of the Archaea, but their absence in *Methanopyrus kandleri*, both species of *Methanosarcina*, and *Halobacterium* sp. NRC-1, demonstrates that the gene is not ubiquitous in this domain. Although it remains a possibility that the gene has simply diverged to such an extent in these organisms that it cannot be detected by conventional BLAST searches, this would seem very unlikely, given that the sequences of all known members of this family are very highly conserved.

To ascertain if any other genes shared the phyletic distribution of *src*, a phylogenetic pattern search was performed using COGnitor. Although COGnitor returned five COGs that followed the phylogenetic pattern of the *src* family, further manual analysis of those genomes not covered in the COG database, eliminated all of these. The closest match was COG1888, for which the phyletic distribution was equivalent to that of the *src* family, with the exception of its presence in the *M. kandleri* genome. No members of COG1888 have to

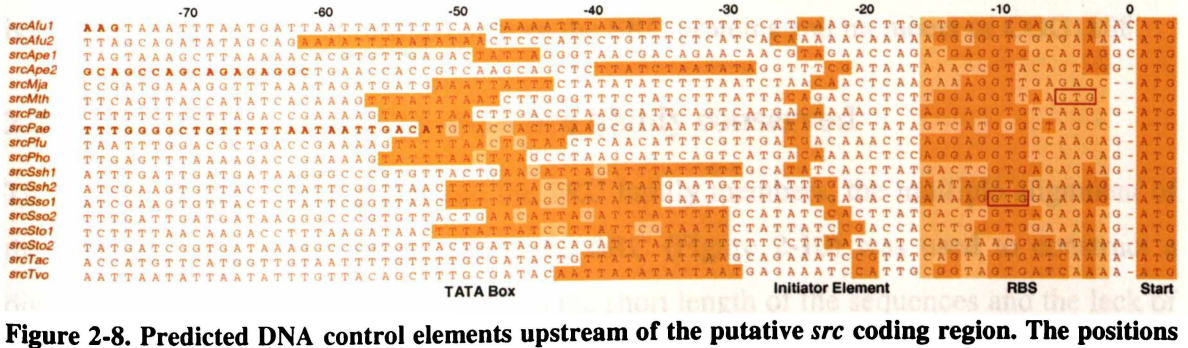


Figure 2-8. Predicted DNA control elements upstream of the putative *src* coding region. The positions of the TATA box, initiator element, ribosome binding site (RBS), and start codon (Start) are indicated. The regions are indicated by shading, with dark shading indicating identity with the consensus sequences for the respective control elements. Previously identified start codons, as predicted in the annotated genome data in GenBank (NCBI), are outlined in red. Coding regions of neighbouring or overlapping genes are in bold print. See Table 2-3 for the associated organism names and genome references.

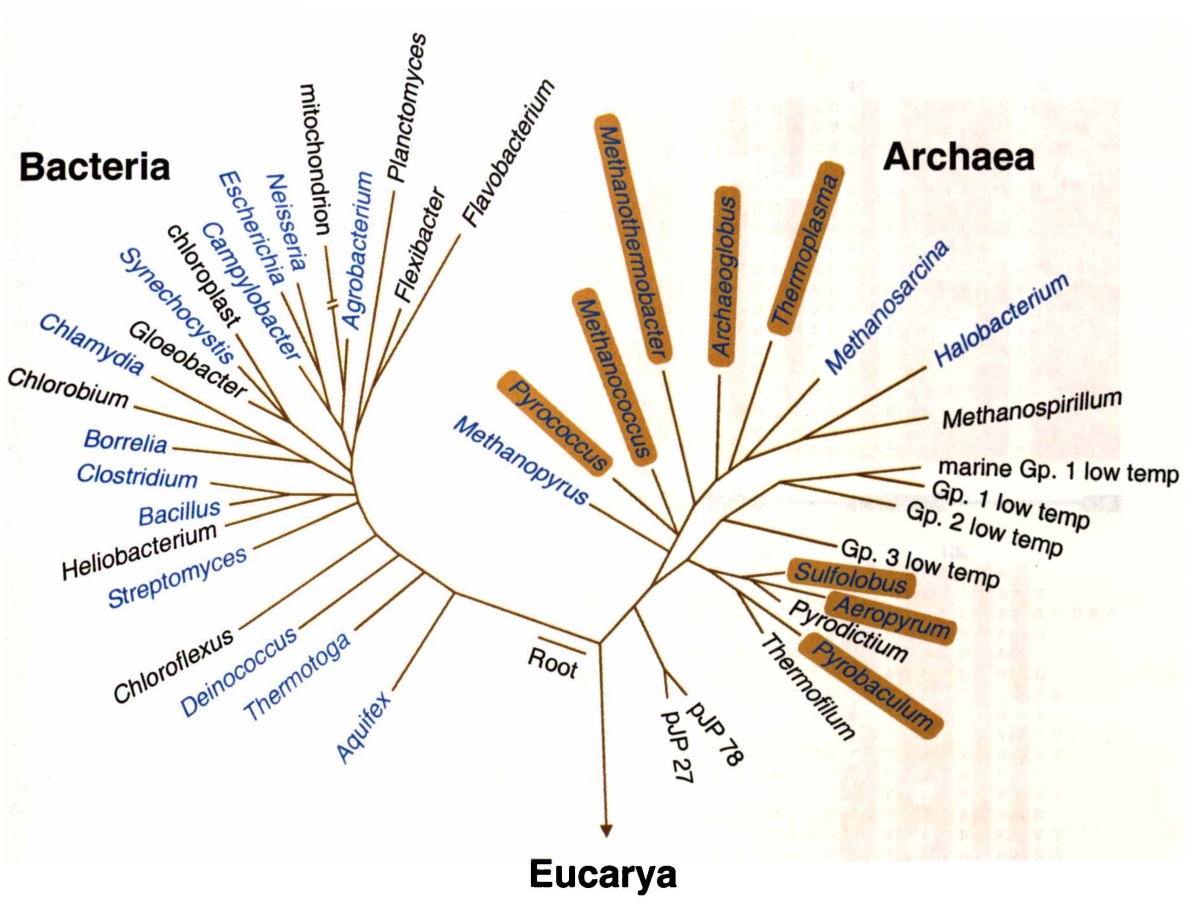


Figure 2-9. Phyletic distribution of the *src* gene amongst extant life. Representatives from the Archaea and the Bacteria are shown on the universal phylogenetic tree based on small subunit ribosomal RNA sequences, as constructed by Norman Pace and colleagues (Pace 1997). Genera in which the complete genome has been sequenced for one or more of its members, at the time of this study, are shown in blue lettering. Genera in which one or more of its members is known to possess a gene encoding an SRC homologue are highlighted.

date been examined. The COG1888 family members are highly conserved and are predicted to encode acidic proteins of approximately 10 kDa, with a secondary structure containing 39% helical, 24% beta sheet and 37% coiled regions. The functional characterisation of a member of COG1888 would need to be conducted to ascertain if there is indeed a functional relationship between the SRC proteins and members of COG1888.

2.3.7 SRC amino acid sequences are highly conserved

To examine the relationship between the SRC family members, the multiple alignment of the predicted amino acid sequences of the SRC (Figure 2-10) was used to construct a distance tree (Figure 2-11). Largely due to the short length of the sequences and the lack of considerable variation between the sequences themselves, statistically justified branching points could not be attained in many cases. Despite this, the tree does illustrate some useful relationships. First, sequences originating from members of the Crenarchaeaota and Euryarchaeota, are generally clustered into two individual groups. However, further data

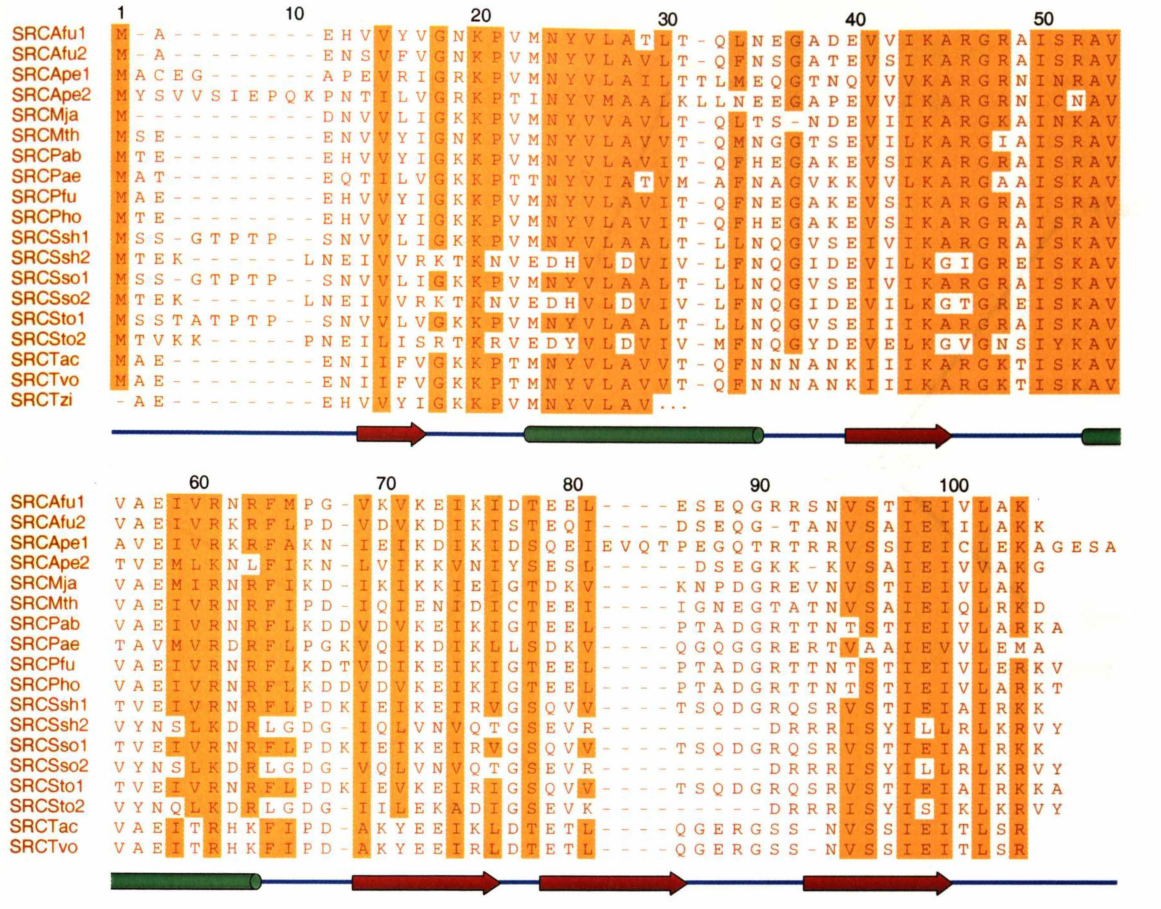


Figure 2-10. Multiple sequence alignment of predicted amino acid sequences of SRE and consensus predicted secondary structure. Similar residues are shown highlighted. Residues were considered similar if they were hydrophobic non-polar (A, V, I, L, P, M, F, W), polar non-charged (G, S, T, C, Y, N, Q), basic (H, R, K), or acidic (E, D) (Lehninger 1982; Karlin and Ghandour 1985). See Table 2-3 for the associated organism names and references. The consensus secondary structure, determined by Jnet (Cuff and Barton 1999), is aligned along the bottom of the sequences. α-helical residues, extended β-sheets and coiled regions are represented as green cylinders, red arrows and blue lines, respectively.

and perhaps more sophisticated analysis techniques would be required to ascertain the likelihood of whether or not the *src* genes have undergone any significant lateral transfer events. Second, the relative positions of the probable paralogues, which occur only in *Aeropyrum pernix*, *Archaeoglobus fulgidus* and the three *Sulfolobus* species, are evolutionarily informative. The observation that SRCApe1 and SRCApe2 group together, as do SRCAfu1 and SRCAfu2, suggests that the duplications that gave rise to these paralogues occurred independently. Conversely, the paralogues from the three Sulfolobus species form two distinct groups, suggesting that the duplication that gave rise to these genes occurred prior to the separation of these three species. Moreover, the considerable divergence of these two groups of *Sulfolobus* SRC sequences from each other, raises the possibility that these protein subgroups have also adopted distinctive functions.

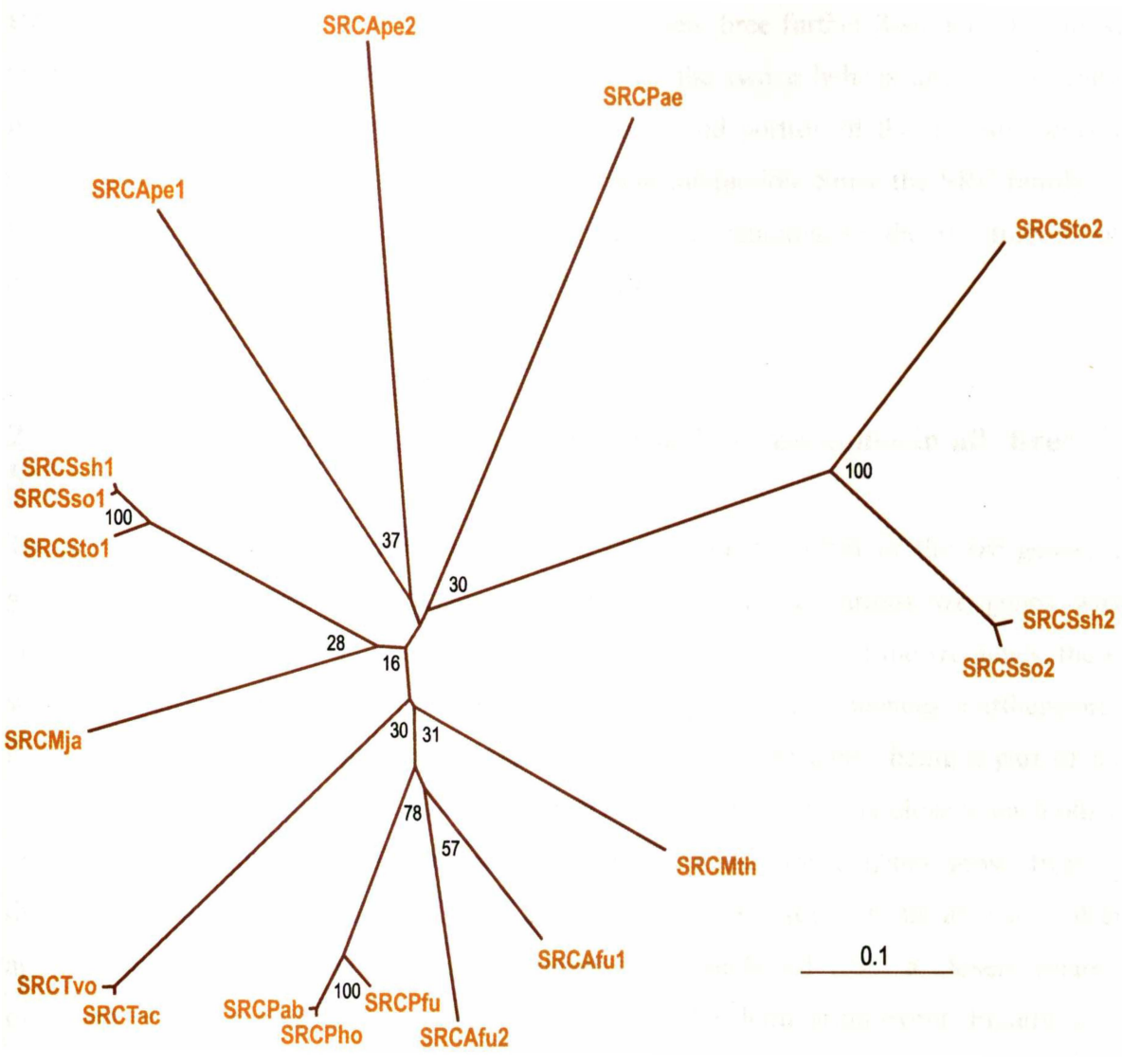


Figure 2-11. Distance tree of SRC protein sequences. The tree was inferred by neighbour-joining analysis. The percentage of 100 bootstrap resamplings that support each topological element is indicated at the branching points. The scale represents the expected number of amino acid changes per sequence position.

2.3.8 SRC proteins are predicted to fold into a mixture of α-helices and β-sheets

Since the SRC sequences did not significantly resemble those of any proteins for which the 3D structure is known, the structure of the SRC proteins was predicted using computational methods. Using the multiple alignment of SRC amino acid sequences, a consensus secondary structure for the SRC protein family was computed using the neural network prediction program, Jnet (Cuff and Barton 1999). The predicted secondary structure of the SRC proteins is shown together with the multiple alignment (Figure 2-10). The predicted structure essentially consists of a relatively long α-helix flanked by two

small β-strands, followed by a second α-helix, and then three further β-strands. The most conserved regions among the sequences align with the two α-helices and the β-strand regions between them. Presumably this highly conserved portion of the protein forms a critical interface for dimerisation and/or nucleic acid interaction. Since the SRC family of DNA-binding proteins represents a unique sequence-fold relationship, the 3D structure of one of its members would be of considerable interest.

2.3.9 *src* genes flank and partially overlap *topR* homologues in all three *Sulfolobus* sequences

To determine if there were any similarities in the genomic context of the *src* genes, a graphical map comparing the genomic neighbourhood of the various *src* genes was constructed (Figure 2-12). It was found that in the nearby proximity of the *src* genes, there was no intergenus conservation in the function of neighbouring sequences. Furthermore, none of the genomes examined indicated any evidence for *src* genes being a part of an operon. The SRC paralogues from *A. pernix* were found to be relatively close to each other in proximity, which is consistent with the suggestion that these genes arose from a duplication event. In contrast, SRCAfu1 and SRCAfu2, were found to be almost 1 Mbp apart, suggesting that either one of the genes was transferred from a closely related organism or that a gross translocation occurred after the duplication event. Finally, and perhaps most interestingly, the *src* paralogues in all three *Sulfolobus* genes was found to flank and partly overlap the gene for reverse gyrase, *topR*, and to precede the gene for an ATP-dependent helicase (COG1201). Reverse gyrase is a DNA topoisomerase that catalyses the formation of positively supercoiled DNA (Kikuchi and Asai 1984). Furthermore, due to the exclusive distribution of the *topR* gene amongst hyperthermophilic archaea and bacteria, it has been implicated as a critical adaptation for life at extreme temperatures (Bouthier de la Tour *et al.* 1991). Although no members of the COG1201 have been characterised to date, they include the DEAD helicase motif and on the basis of sequence homology have been placed in Superfamily II of the ATP-dependent helicases. The close positioning of these four genes, all of which are involved with the modulation of DNA topology, within three *Sulfolobus* genomes is intriguing. However, further research to understand the nature and role of these proteins in DNA topology would be required to appreciate if their related genomic context is significant, or merely coincidental.

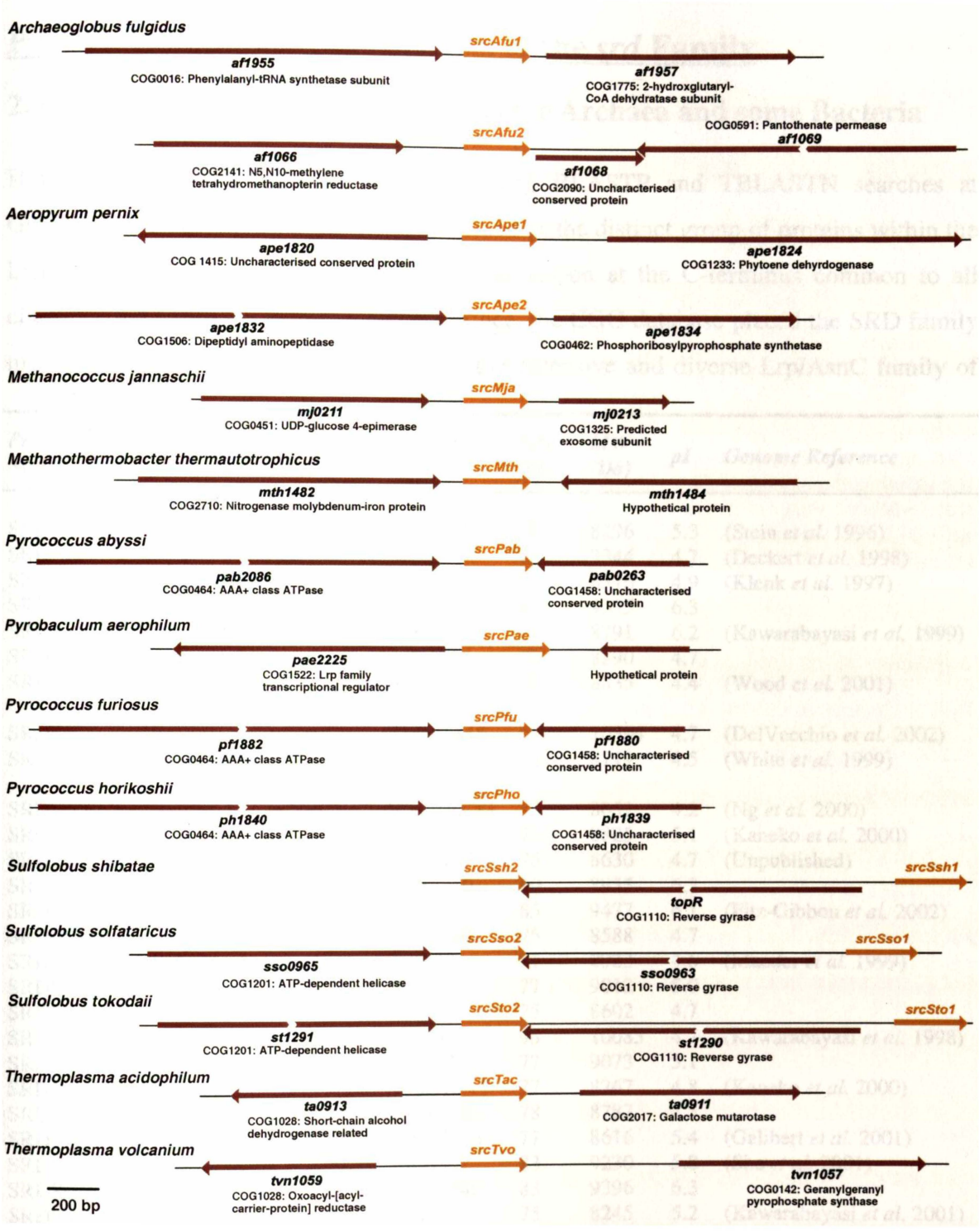


Figure 2-12. Comparison of relative positions, orientations and proposed functions of genes neighbouring or overlapping with *src* in various genomes. The arrows represent the relative size, position and orientation of *src* and the nearby genes. Overlapping genes are indicated by offset arrows. Functions for nearby genes were assigned using the Cognitor program at NCBI. Functions for genes that did not belong to any known COG were assigned according to the annotated data for the respective genomes from GenBank (NCBI). Genes are illustrated according to the scale shown, with the exception of those indicated with a break-point where the gene was resized to fit the figure.

Part C: Computational Analysis of the *srd* Family

2.3.10 *srd* occurs in all non-methanogenic Archaea and some Bacteria

Homologues of SRDTzi were identified using BLASTP and TBLASTN searches at GenBank (NCBI). The SRD family was defined as the distinct group of proteins within the Lrp/AsnC family that lacked the 60 to 80 aa region at the C-terminus common to all characterised Lrp/AsnC family members. Since, the COG database placed the SRD family members in the same group as the rest of the extensive and diverse Lrp/AsnC family of

Protein Name	*Originating Organism*	*Original Gene ID*	*Size (aa)*	*MW (Da)*	*pI*	*Genome Reference*
SRD4b7	Unknown isolate 4b7	AAK66787	76	8296	5.3	(Stein *et al.* 1996)
SRDAae	*Aquifex aeolicus*	N/A	84	9344	4.7	(Deckert *et al.* 1998)
SRDAful	*A. fulgidus*	AF0651	77	8576	4.9	(Klenk *et al.* 1997)
SRDAfu2		AF1404	83	9395	6.3	
SRDApe1	*A. pernix*	APE0659a	79	8791	6.2	(Kawarabayasi *et al.* 1999)
SRDApe2		APE1799a	78	8890	4.7	
SRDAtu	*Agrobacterium tumefaciens*	Atu2558	78	8835	4.4	(Wood *et al.* 2001)
SRDBme	*Brucella melitensis*	BMEI1845	93	10394	4.7	(DelVecchio *et al.* 2002)
SRDDra	*Deinococcus radiodurans*	DR2287	78	8719	4.5	(White *et al.* 1999)
SRDHNRC1	*Halobacterium* NRC1	VNG1922G	75	8052	4.2	(Ng *et al.* 2000)
SRDMlo	*Mesorhizobium loti*	msl0718	76	8825	5.1	(Kaneko *et al.* 2000)
SRDPab1	*P. abyssi*	PAB7115	75	8630	4.7	(Unpublished)
SRDPab2		PAB7359	80	8955	6.3	
SRDPae1	*Pyrobaculum*	PAE1507	85	9477	6.1	(Fitz-Gibbon *et al.* 2002)
SRDPae2	*aerophilum*	PAE3469	75	8588	4.7	
SRDPfu1	*P. furiosus*	PF1732	81	8923	4.6	(Maeder *et al.* 1999)
SRDPfu2		PF1022	77	9012	5.0	
SRDPfu3		PF1893	75	8602	4.7	
SRDPho1	*P. horikoshii*	PHs045	93	10083	4.7	(Kawarabayasi *et al.* 1998)
SRDPho2		PHs023	77	9073	5.1	
SRDSco1	*Streptomyces coelicolor*	SC6G10.13	77	8267	4.8	(Kaneko *et al.* 2000)
SRDSco2		SC7A1.05c	78	8787	4.5	
SRDSme	*Sinorhizobium meliloti*	SMc02235	77	8616	5.4	(Galibert *et al.* 2001)
SRDSso1	*S. solfataricus*	SSO5522	83	9230	5.8	(She *et al.* 2001)
SRDSso2		SSO10340	83	9396	6.3	
SRDSto1	*S. tokodaii*	STS042	75	8245	5.2	(Kawarabayasi *et al.* 2001)
SRDSto2		STS071	77	8586	5.2	
SRDTac	*T. acidophilum*	N/A	77	8586	5.2	(Ruepp *et al.* 2000)
SRDTvo	*T. volcanium*	TVN0295	78	8860	4.6	(Kawashima *et al.* 2000)

Table 2-4. Summary information for SRD homologues, including originating organism, original gene name in the annotated genome data, predicted size in amino acids, theoretical molecular weight (MW), and theoretical pI. Genes that had not originally been identified in the annotated genome data, are indicated with N/A (not applicable) in the Gene ID column. Theoretical MW and pI were determined using the ProtParam tool (ExPASy).

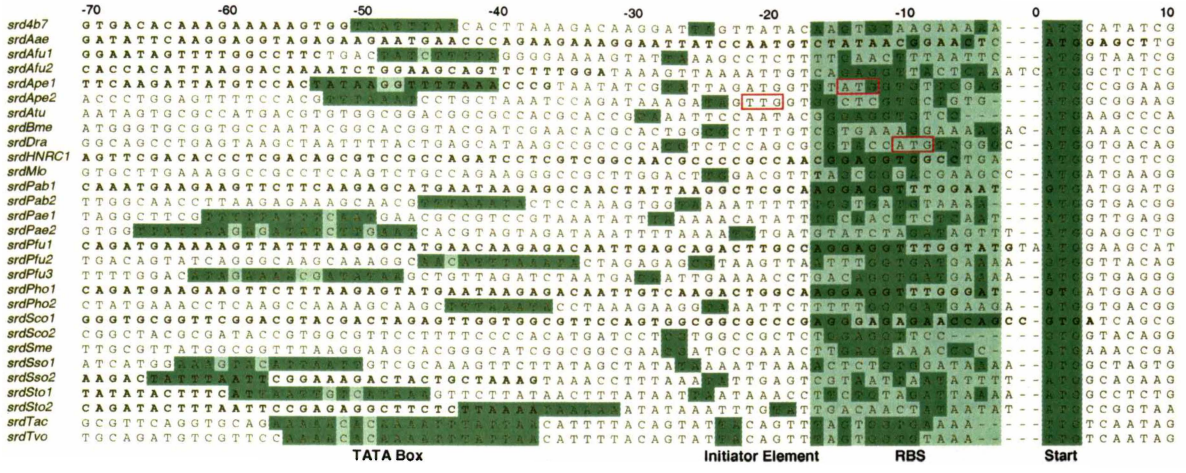


Figure 2-13. Predicted DNA control elements upstream of the putative *srd* coding region. The positions of the TATA box (in archaeal sequences), initiator element, ribosome binding site (RBS), and start codon (Start) are indicated. The regions are indicated by shading, with dark shading indicating identity with the consensus sequences for the respective control elements. Previously identified start codons, as predicted in the annotated genome data in GenBank (NCBI), are outlined in red. Coding regions of neighbouring or overlapping genes are in bold print. See Table 2-4 for the associated organism names and genome references.

transcriptional regulators (COG1522), the SRD family members had to be identified manually by recursive BLASTP searches. Initially, the theoretical amino acid sequence of the predicted SRDTzi homologue, PF1893 (named SRDPfu) from *P. furiosus*, was used to search the non-redundant database at GenBank (NCBI). Using the subsequently retrieved sequences of SRD family as further queries, the complete list of all known SRD homologues was collected. Finally, using TBLASTN searches, two further putative *srd* homologues, which had not been annotated in the genome data, were discovered in *Aquifex aeolicus* and *Thermoplasma acidophilum*. A complete list of all identified members of the SRD family is shown in Table 2-4. A total of 29 SRD homologues were identified in 11 different archaeal and eight different bacterial species.

The putative protein coding regions were examined to verify that the most probable start codon had been identified in the annotated genome data (Figure 2-13). Based on homology with other SRD sequences and the positions of the predicted ribosome binding site, the predicted start positions for SRDApe1, SRDApe2 and SRDDra were modified from those in the annotated genome data at GenBank. An examination of the upstream regions in the sequences, revealed that several of the *srd* genes, *srdAae*, *srdHNRC1*, *srdPab1*, *srdPfu1*, *srdPho1* and *srdSco1*, are likely to be part of an operon. Of the remaining sequences that were of archaeal origin, the majority possessed probable TATA boxes at the expected position of approximately 50 bp upstream of the start codon.

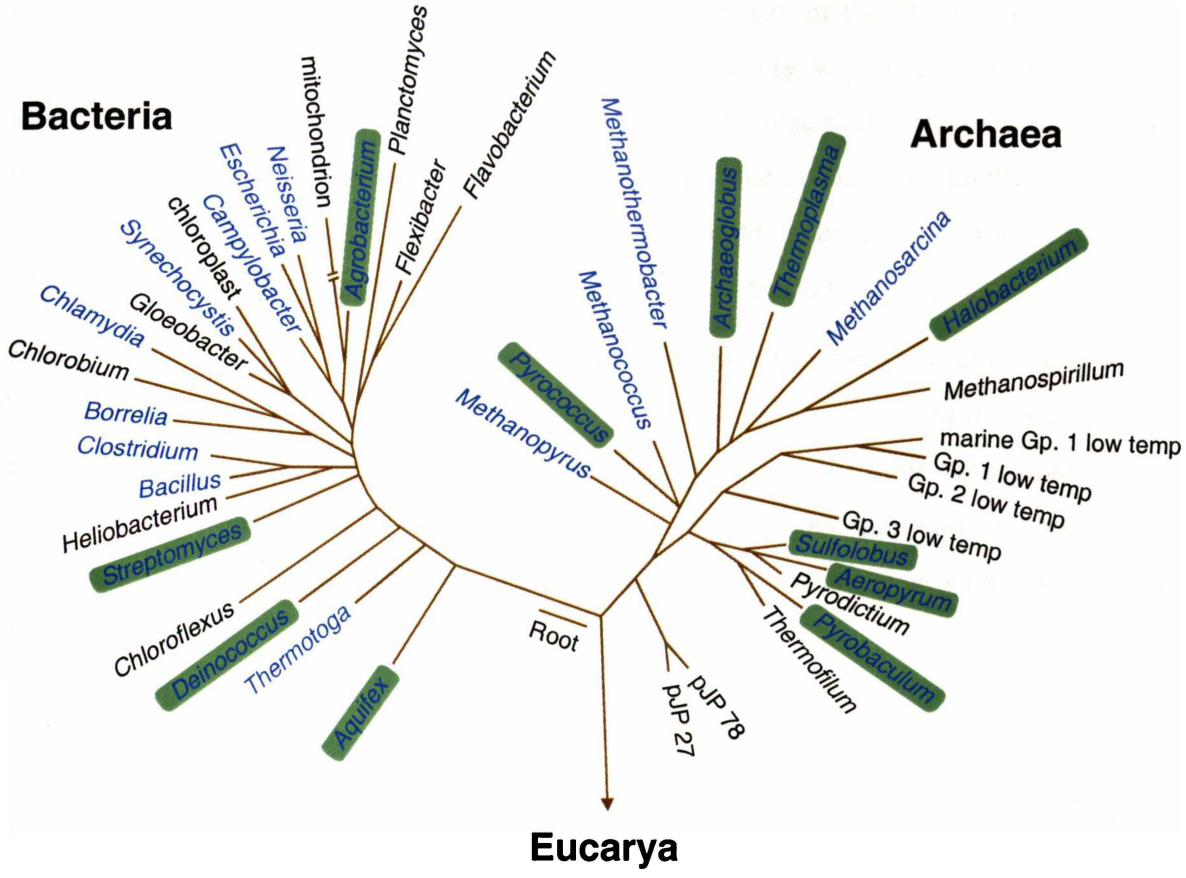


Figure 2-14. Phyletic distribution of the *srd* gene amongst extant life. Representatives from the Archaea and the Bacteria are shown on the universal phylogenetic tree based on small subunit ribosomal RNA sequences, as constructed by Norman Pace and colleagues (Pace 1997). Genera in which the complete genome has been sequenced for one or more of its members, at the time of this study, are shown in blue lettering. Genera in which one or more of its members is known to possess a gene encoding an SRD homologue are highlighted.

The phyletic distribution of the *srd* gene is shown in Figure 2-14. Of the completely sequenced genomes available at the time of this study, 10 out of 14 archaeal and 7 out of 60 bacterial genomes possessed at least one copy of *srd*. In addition, an *srd* homologue was also identified in the sequence of a genome fragment that was isolated from the environment and was suggested to originate from an archaeon (Stein *et al.* 1996). BLASTP searches did not reveal significant homology with any eucaryal sequences. A search of four other virtually complete archaeal genome sequences, revealed that the *F. acidarmanus*, *Methanosarcina acetivorans* and *M. mazei* genomes all contained at least one *srd* homologue. This data presents an interesting phyletic distribution of *srd* within the Archaea, in that *srd* homologues occur in every archaeal genome, except those of the methanogens (i.e. members of the genera, *Methanococcus*, *Methanothermobacter*, *Methanopyrus* and *Methanosarcina*). Conversely, amongst the genomes of members of the Bacteria, *srd* occurs only relatively sparsely, and, consequently, it is suggested that the *srd*

gene was most likely to have been transferred to members of this domain by horizontal transfer. Interestingly, *srd* homologues are present in every sequenced genome from the members of the Rhizobiacae family. The diversity in the metabolic activities and lifestyle of the bacteria in which *srd* occurs, suggests that the gene is likely to function in a fairly generic role, perhaps similar to that of the Lrp/AsnC proteins, as transcriptional regulators. The specific relationships between the *srd* genes is examined further in 2.3.12.

Finally, a phylogenetic pattern search using the COGnitor program, revealed that no other COG possesses an equivalent phyletic distribution to the *srd* gene family. This finding suggests that *srd* is not likely to be exclusively reliant on any other non-basal protein for its function. Furthermore, this lack of involvement with other non-core proteins is in accordance with the proposed ability of the *srd* gene to have transferred widely to different bacterial species.

2.3.11 SRD has sequence and structural homology with the C-terminal half of LrpA

As stated above, the SRD proteins possess significant sequence homology with the Lrp/AsnC family of DNA-binding proteins. Recently, the crystal structure of LrpA, a member of the Lrp/AsnC family found in *Pyrococcus furiosus*, was elucidated to 2.9 Å resolution (Leonard *et al.* 2001). The 3D structure data was used to compile a model of the protein (Figure 2-15), and the inferred secondary structure was compared to the consensus secondary structure predicted for the SRD protein family (Figure 2-16). The result vividly demonstrates that the βαββαβ folding topology of the C-terminal half of LrpA matches extremely well with that predicted for the SRD proteins. More intriguingly, the 3D structure of LrpA shows that this protein contains two very distinct domains, separated by a long coiled region, the second of which is wholly homologous with the SRD proteins. This observation further justifies the distinction of the SRD proteins from other members of the Lrp/AsnC family. Furthermore, this observation also raises the hypothesis that the Lrp/AsnC proteins originated from the fusion of genes encoding two separate proteins or protein domains. Although it is tempting to speculate that an ancestor of the genes encoding the SRD proteins was such a precursor, the possibility cannot be eliminated that the SRD proteins are merely a consequence of a truncation of an Lrp/AsnC gene. To

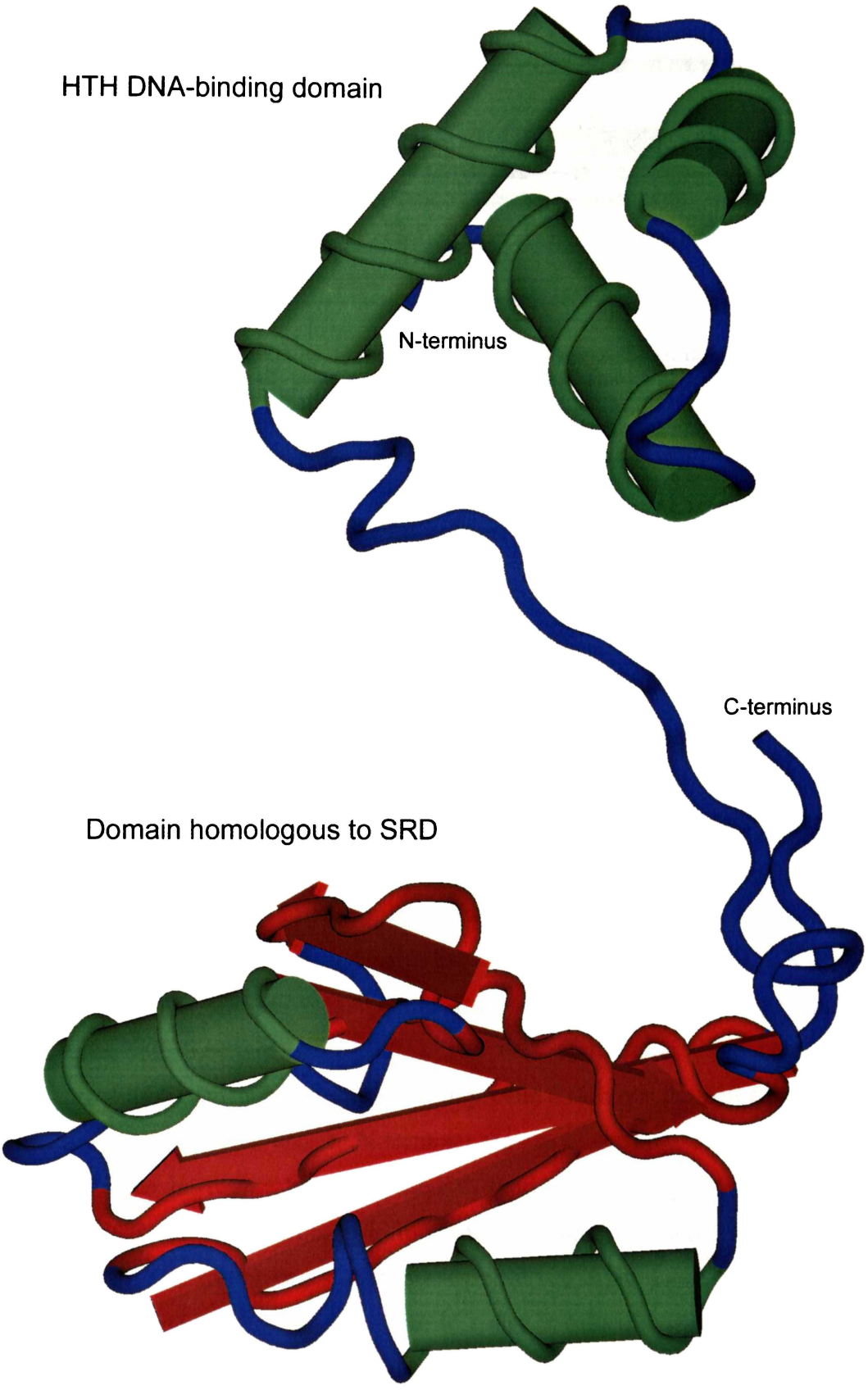


Figure 2-15. CN3D rendition of the crystal structure of LrpA from *Pyrococcus furiosus*. α-helical residues, extended β-sheets and coiled regions are represented as green cylinders, red arrows and blue curves, respectively. The helix-turn-helix (HTH) DNA-binding domain (top, N-terminal half) and the domain homologous with SRD (bottom, C-terminal half) are indicated. Structural data was acquired from the MMDB at the NCBI (PDB Id 1I1G) and visualised with CN3D 4.0 (NCBI). The data originated from a crystal structure determination at 2.9 Å resolution (Leonard *et al.* 2001).

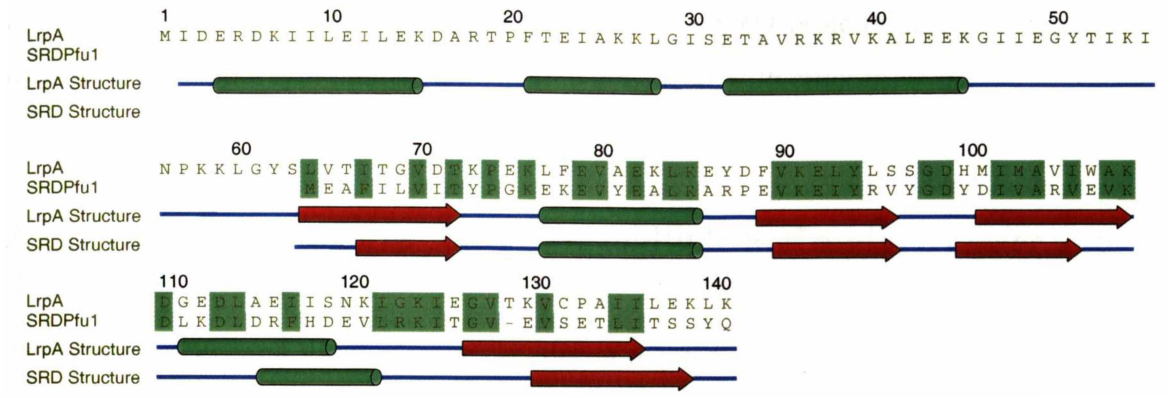


Figure 2-16. Alignment of the amino acid sequences of LrpA and SRDPfu1, with the secondary structures of LrpA and the SRD proteins. Similar residues are shown highlighted. Residues were considered similar if they were hydrophobic non-polar (A, V, I, L, P, M, F, W), polar non-charged (G, S, T, C, Y, N, Q), basic (H, R, K), or acidic (E, D) (Lehninger 1982; Karlin and Ghandour 1985). The LrpA secondary structure was derived from the x-ray crystal structure data of LrpA from *Pyrococcus furiosus* (Leonard *et al.* 2001). The SRD structure is the consensus secondary structure determined by the Jnet program (Cuff and Barton 1999), using the multiple alignment of SRD proteins (Figure 2-17). α-helical residues, extended β-sheets and coiled regions are represented as green cylinders, red arrows and blue lines, respectively.

attempt to distinguish between these possibilities, the relationship between the Lrp/AsnC and *srd* genes is explored further in 2.3.12.

The crystal structure of LrpA also provides some valuable insights into the possible function of the SRD proteins. Combined with mutational studies of LrpA and other Lrp/AsnC proteins, it has been shown that the helix-turn-helix motif at the N-terminus of LrpA forms the primary region for DNA-binding and the C-terminal domain is largely responsible for dimerisation and further quaternary associations (Platko and Calvo 1993; Leonard *et al.* 2001). Interestingly, although the structural comparison of LrpA with other structures revealed no matches over the entire structure, a number of hits were reported for the C-terminal half (Leonard *et al.* 2001). The best of these matches were with the ribosomal protein S6, which in conjunction with protein S18 binds the 16S ribosomal RNA subunit (Lindahl *et al.* 1994), and the N-terminal regions of the archaeal DNA polymerase II enzymes, which have also been proposed to bind to RNA (Zhao *et al.* 1999). The structural motif that these proteins have in common, a four-stranded antiparallel β-sheet with two α-helices packed on one side, has been observed in the DNA-binding domain of the bovine papillomavirus E2 protein, where it is also involved in dimerisation (Hegde *et al.* 1992). Clearly, this structural motif has remarkable versatility, and in all known instances functions in tandem with other proteins or protein domains. Consequently, the study of this domain as an individual protein is of considerable interest, not only from a molecular evolutionary standpoint, but also in understanding its role in the context of other

proteins. Moreover, it would seem plausible that the SRD proteins not only associate to form multimers, but also interact and operate in concert with other proteins.

2.3.12 *srd* is likely to have undergone multiple horizontal transfer and duplication events

To further understand the relationship between the *srd* genes, a multiple alignment of the amino acid sequences of the conceptual SRD proteins was constructed (Figure 2-17), and the resulting information used to generate a distance tree by neighbour-joining methods (Figure 2-18). Although in some instances, statistically significant branching could not be determined, the tree nonetheless highlights some important relationships. First, the SRD sequences from members of the Rhizobiacae family all cluster together. This observation, combined with the fact that *srd* occurs in every Rhizobiacae genome sequenced to date, suggests that *srd* was originally acquired by the cenancestor of the Rhizobiacae. Second, the clustering of all the bacterial sequences, except for one paralogue from *Streptomyces coelicolor*, is noteworthy. The simplest scenario that would result in this pattern, is that *srd* was horizontally transferred to the Bacteria from only one source, and then later transferred again, in at least two independent instances, to other members of the Bacteria. However, more data would be required to distinguish this from other possible, more complex, scenarios. Third, the branching order of the archaeal *srd* genes bears almost no resemblance to that of the corresponding tree of 16S rRNA sequences. However, due to the considerable variation in the relatively short SRD sequences, it is not possible to discern from this data whether this is a result of significant lateral transfer within the Archaea or merely a consequence of the high variability in the *srd* gene. More distinct, however, is the likelihood that the SRD genes have undergone duplication on several independent occasions. This is apparent from some of the homologues that occur within the same genome that are more closely related to one another than any other SRD sequence, such as SRDPae1 and SRDPae2, SRDAfu1 and SRDAfu2, and possibly the SRDPab1/SRDPho1/SRDPfu1 and SRDPab2/SRDPho2/SRDPfu2 homologues. In the other cases, it is not possible to distinguish whether the genes originated from a more distant duplication event and consequently became more considerably divergent or were acquired from another organism by lateral transfer.

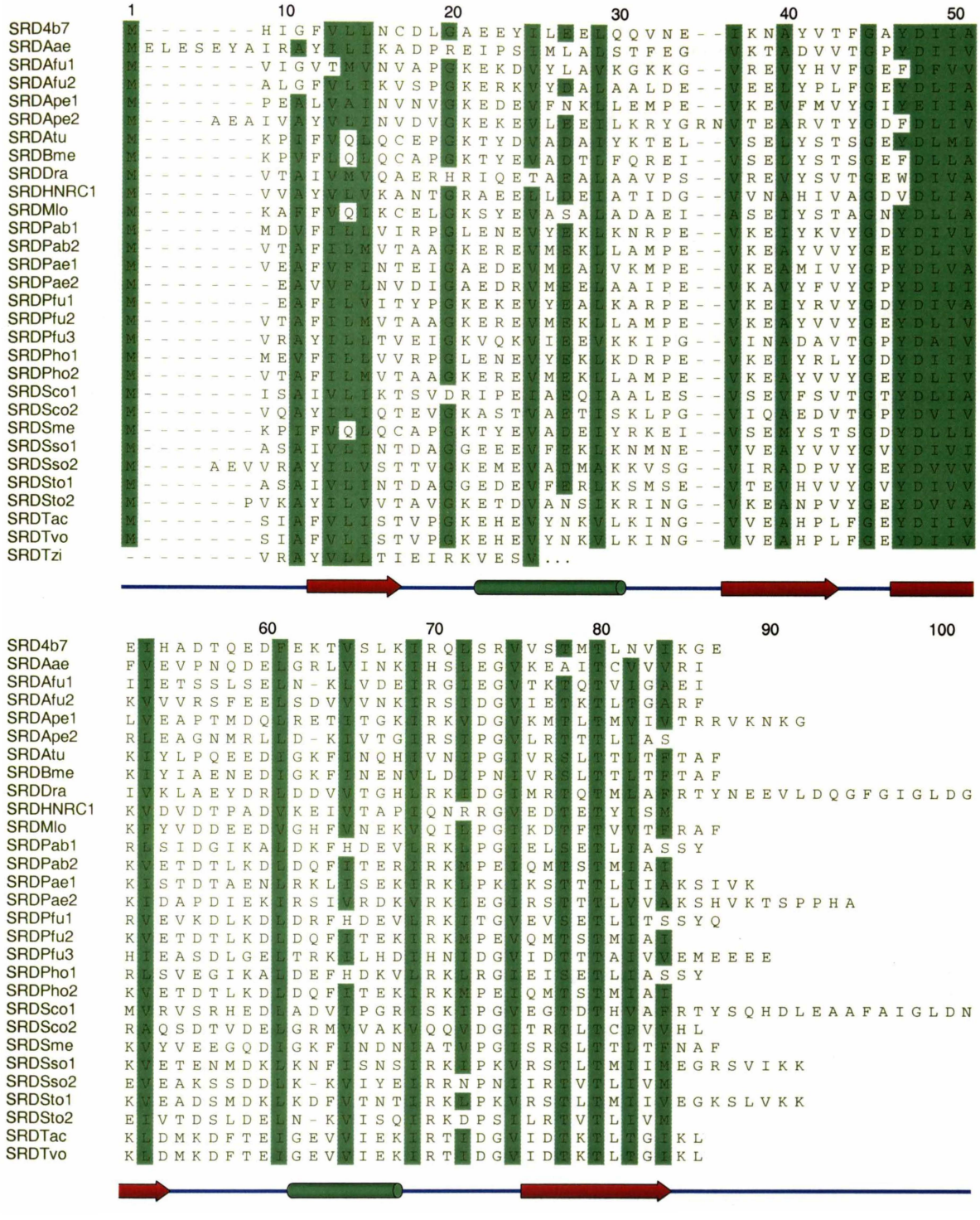


Figure 2-17. Multiple sequence alignment of predicted amino acid sequences of SRD and consensus predicted secondary structure. Similar residues are shown highlighted. Residues were considered similar if they were hydrophobic non-polar (A, V, I, L, P, M, F, W), polar non-charged (G, S, T, C, Y, N, Q), basic (H, R, K), or acidic (E, D) (Lehninger 1982; Karlin and Ghandour 1985). See Table 2-4 for the associated organism names and references. The consensus secondary structure, determined by Jnet (Cuff and Barton 1999), is aligned along the bottom of the sequences. α-helical residues, extended β-sheets and coiled regions are represented as green cylinders, red arrows and blue lines, respectively.

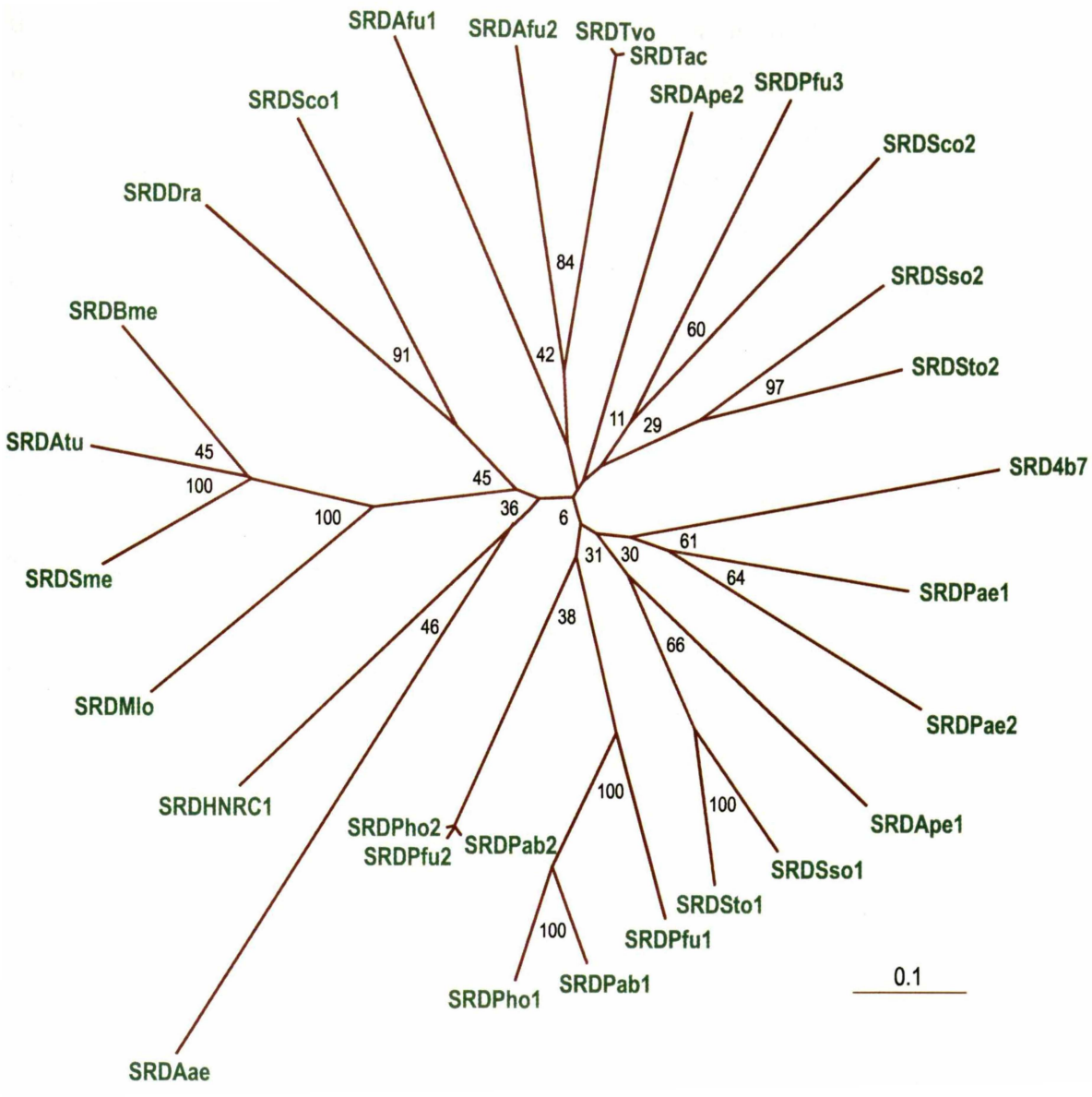


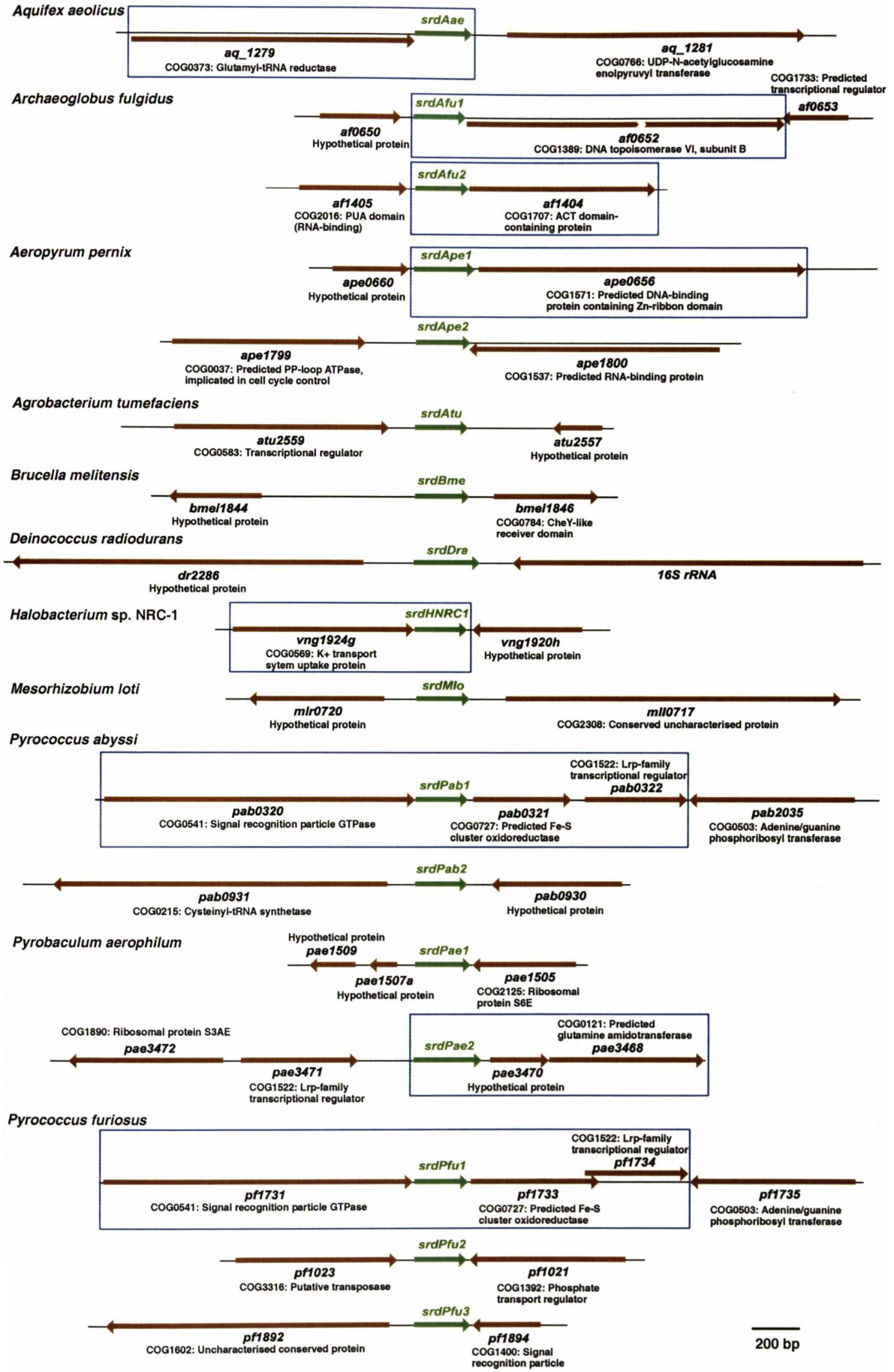
Figure 2-18. Distance tree of SRD protein sequences. The tree was inferred by neighbour-joining analysis. The percentage of 100 bootstrap resamplings that support each topological element is indicated at the branching points. The scale represents the expected number of amino acid changes per sequence position.

The understanding of the phylogenetic relationship between the *srd* and Lrp/AsnC gene families may form a critical aspect in the appreciation of the molecular evolution of the diverse and abundant group of transcriptional regulators that the latter family encodes. In an attempt to elucidate this relationship, approximately 100 representative amino acid sequences from members of the Lrp/AsnC family were aligned with the SRD sequences and the resulting alignment used to build a distance tree (results not shown). Due to the substantial variation in the sequences amongst this diverse family of proteins, statistically significant branching could not be determined in many cases. However, the results did indicate a tendency towards the SRD sequences being more closely related to one another

than to most of the Lrp/AsnC proteins. To more convincingly substantiate this suggestion, it is likely that further structural and functional data from SRD and Lrp/AsnC proteins will be necessary. If the *srd* genes do indeed form a phylogenetically distinct family from the Lrp/AsnC family, i.e. possess an SRD-like cenancestor, this would have important implications on the molecular evolution of the Lrp/AsnC family. That is, it would raise the possibility that an ancestor of the *srd* genes was a precursor for the formation of the Lrp/AsnC cenanestor, and reduce the likelihood that the *srd* family occurs merely as a result of truncations of Lrp/AsnC genes.

2.3.13 *srd* genes often occur in operon-like arrangements

To gather further insight into the possible role of SRD, the relative positions and functions of the genes neighbouring the *srd* genes were examined. A graphical map comparing the genomic neighbourhood of *srd* genes available at the time of this study, is shown in Figure 2-19. Although the data does not reveal any particular trend in terms of the types of genes associated with *srd*, it does appear that many of the genes, namely *srdAae*, *srdAfu2*, *srdApe1*, *srdHNRC1*, *srdPab1*, *srdPfu1*, *srdPho1*, *srdSco1*, *srdSso2*, *srdSto2*, *srdSso1*, *srdTac*, and *srdTvo*, occur in operon-like arrangements. That is, the genes that neighbour these members of the *srd* family, are oriented in the same direction and are either very close or overlap with the *srd* genes. Consistent with an operon structure, putative promoter and terminator regions were identified only at the 5'- and 3'-ends of the proposed operon regions, respectively (results not shown). In these putative operons, there appears to be no pattern in either the position of the *srd* gene relative to the other genes, or in the number of genes present, which ranges from two to four. The diverse, apparently unrelated functions of the putative *srd*-associated operons, suggests that the SRD proteins are unlikely to interact directly with the products of the other genes in the operon. One possibility, however, is that the SRD proteins play a somewhat more generic role, perhaps as part of an autoregulatory system for the operon itself. This notion is supported by the observation that the transcription of every Lrp/AsnC gene examined so far, is negatively autoregulated (Kolling and Lother 1985; Newman and Lin 1995; Napoli *et al.* 1999; Brinkman *et al.* 2000). Although it is not typical for an operon to encode a protein that regulates its own expression, instances of autoregulation by a member of an operon have been described in members of both the Archaea and the Bacteria (Chatterjee *et al.* 1996; Mayer *et al.* 1998; Xiong *et al.* 2000).



(continues on next page)

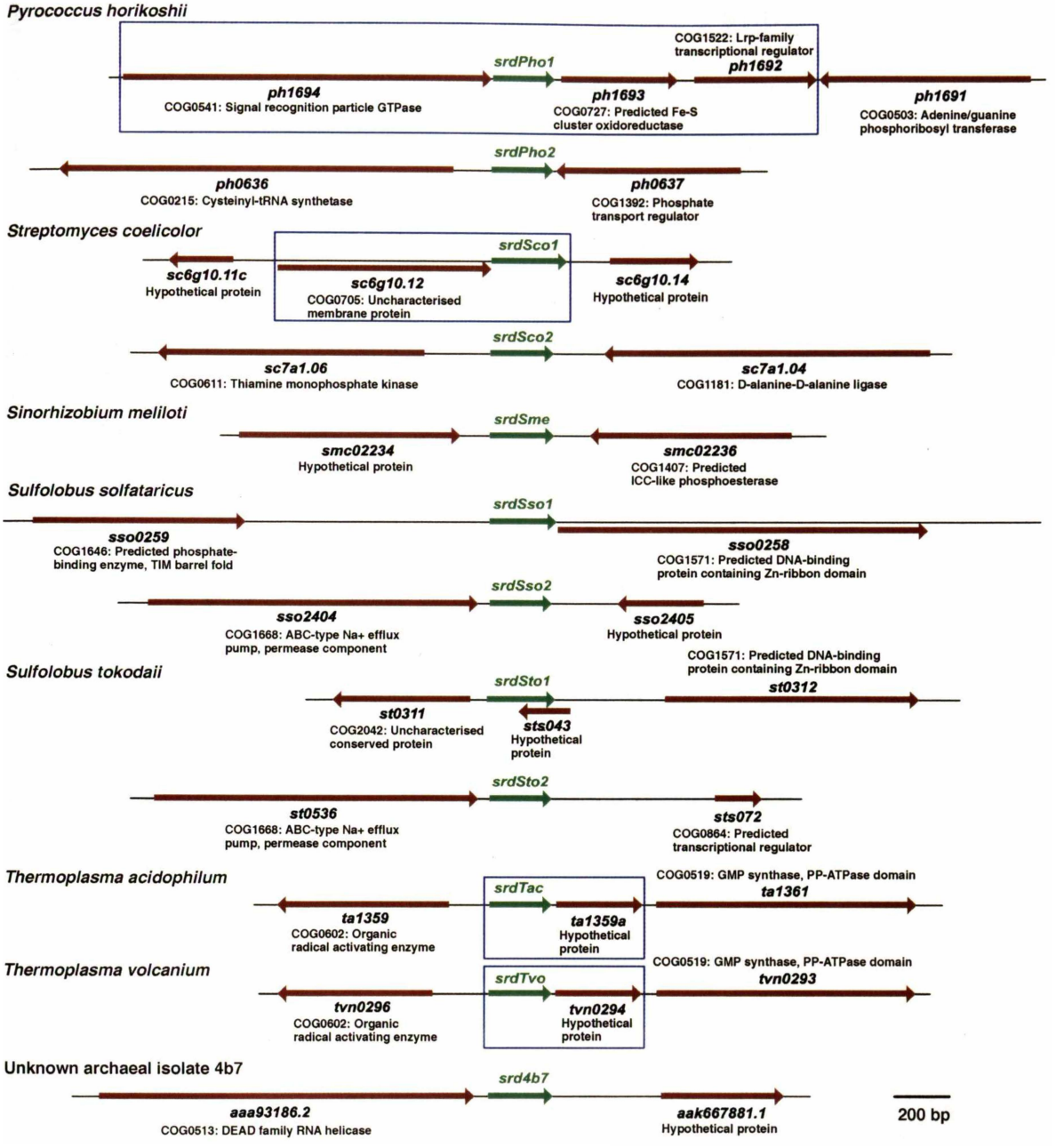


Figure 2-19. Comparison of relative positions, orientations and proposed functions of genes neighbouring or overlapping with *srd* in various genomes. The arrows represent the relative size, position and orientation of *srd* and the nearby genes. Overlapping genes are indicated by offset arrows. Functions for nearby genes were assigned using the Cognitor program at NCBI. Functions for genes that did not belong to any known COG were assigned according to the annotated data for the respective genomes from GenBank (NCBI). Genes are illustrated according to the scale shown, with the exception of those indicated with a break-point where the gene was resized to fit the figure. Boxed regions indicate groups of genes that are suspected to be part of an operon, as determined by the relative positions of the genes and the locations of putative promoter and terminator regions.

Part D: Computational Analysis of the *sre* Family

2.3.14 *sre* is found exclusively in most archaeal genomes

To identify potential homologues of SRETzi, the COGnitor program and BLASTP GenBank searches were performed at the NCBI. Initially, the conceptual amino acid sequence of the predicted SRETzi homologue, PHS059 (named SREPho) from *P. horikoshii*, was used to query the COG database at NCBI. SREPho was found to belong to COG1698, which contained a total of eight genes as of May 2002. Using BLASTP searches, a further three genes encoding *sre* homologues were identified in *P. furiosus*, *S. acidocaldarius*, and *S. tokodaii*. TBLASTN searches did not reveal any further *sre* homologues. A complete list of all identified members of the SRE family is shown in Table 2-5. A total of 12 SRE homologues were identified in 12 different archaeal genomes. Interestingly, no genome contained more than one homologue of *sre*.

The ORFs identified in GenBank were examined to verify that the most likely start position had been correctly identified (Figure 2-20). Based on homology with other SRE sequences and the positions of the putative ribosome binding site, the predicted start positions for SREApe, SREMka, SREMth, SREPab, SREPho, and SRETvo were modified from those in GenBank. Moreover, the examination of the upstream regions also revealed that all of the *sre* genes possessed probable TATA box elements (Figure 2-20).

Protein Name	*Originating Organism*	*Original Gene ID*	*Size (aa)*	*MW (Da)*	*pI*	*Genome Reference*
SREAfu	*A. fulgidus*	AF2370a	84	9484	4.9	(Klenk *et al.* 1997)
SREApe	*A. pernix*	APES069	89	9877	9.7	(Kawarabayasi *et al.* 1999)
SREMja	*M. jannaschii*	MJ1419	93	10278	5.0	(Bult *et al.* 1996)
SREMka	*M. kandleri*	MK1586	89	10074	4.5	(Slesarev *et al.* 2002)
SREMth	*M. thermautotrophicus*	MTH1407	84	9657	4.6	(Smith *et al.* 1997)
SREPab	*P. abyssi*	PAB1112	86	9770	4.7	(Unpublished)
SREPfu	*P. furiosus*	PF0239	86	9794	4.8	(Maeder *et al.* 1999)
SREPho	*P. horikoshii*	PHS059	86	9798	4.7	(Kawarabayasi *et al.* 1998)
SRESso	*S. solfataricus*	SSO6503	89	9820	6.6	(She *et al.* 2001)
SRESto	*S. tokodaii*	STS065	89	9874	6.6	(Kawarabayasi *et al.* 2001)
SRETac	*T. acidophilum*	Ta0600	85	9555	4.6	(Ruepp *et al.* 2000)
SRETvo	*T. volcanium*	TVN0626	85	9573	4.6	(Kawashima *et al.* 2000)

Table 2-5. Summary information for SRE homologues, including originating organism, original gene name in the annotated genome data, predicted size in amino acids, theoretical molecular weight (MW), and theoretical pI. Theoretical MW and pI were determined using the ProtParam tool (ExPASy).

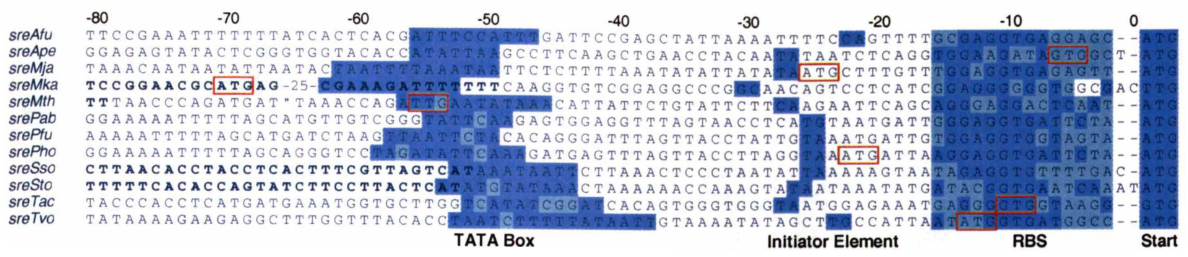


Figure 2-20. Predicted DNA control elements upstream of the putative *sre* coding region. The positions of the TATA box, initiator element, ribosome binding site (RBS), and start codon (Start) are indicated. The regions are indicated by shading, with dark shading indicating identity with the consensus sequences for the respective control elements. Previously identified start codons, as predicted in the annotated genome data in GenBank (NCBI), are outlined in red. Coding regions of neighbouring or overlapping genes are in bold print. See Table 2-5 for the associated organism names and genome references.

The phyletic distribution of the *sre* genes was examined, and the results illustrated on a universal phylogenetic tree, based on small subunit rRNA sequences (Figure 2-21). Of the completely sequenced genomes available at the time of this study, 12 out of 14 archaeal genomes possessed at least one copy of *sre*. BLASTP searches did not reveal significant homology with any eucaryal or bacterial sequences. A search of three other almost complete archaeal genome sequences, revealed that *F. acidarmanus*, *Methanosarcina acetivorans* and *M. mazei* each contained a gene with convincing similarity with other members of the *sre* family. These results demonstrate that the *sre* gene is extremely prevalent amongst the domain Archaea, with only the genomes of *Pyrobaculum aerophilum* and *Halobacterium* sp. NRC1, not possessing an *sre* homologue.

To determine whether any other genes shared the phyletic distribution of *sre*, a phylogenetic pattern search was performed using COGnitor. Although COGnitor returned five COGs that followed the phylogenetic pattern of the *sre* family, further manual searches of those genomes not currently included in the COG database, eliminated all of these. The closest match was COG1096, for which the phyletic distribution is the same as that of the *sre* family, except that it occurs in the *P. aerophilum* genome. Although no members of COG1096 have so far been examined, they are described in the COG database as RNA-binding proteins, since they contain a zinc-finger domain and a domain homologous to the S1 ribosomal protein. The sequences of the COG1096 genes are highly conserved and are predicted to encode proteins of approximately 20 kDa, with a secondary structure containing 40% helical, 15% beta sheet and 45% coiled regions. To ascertain if there is any functional relationship between the SRE proteins and members of COG1096, a more detailed understanding of the function of members of COG1096 will be required.

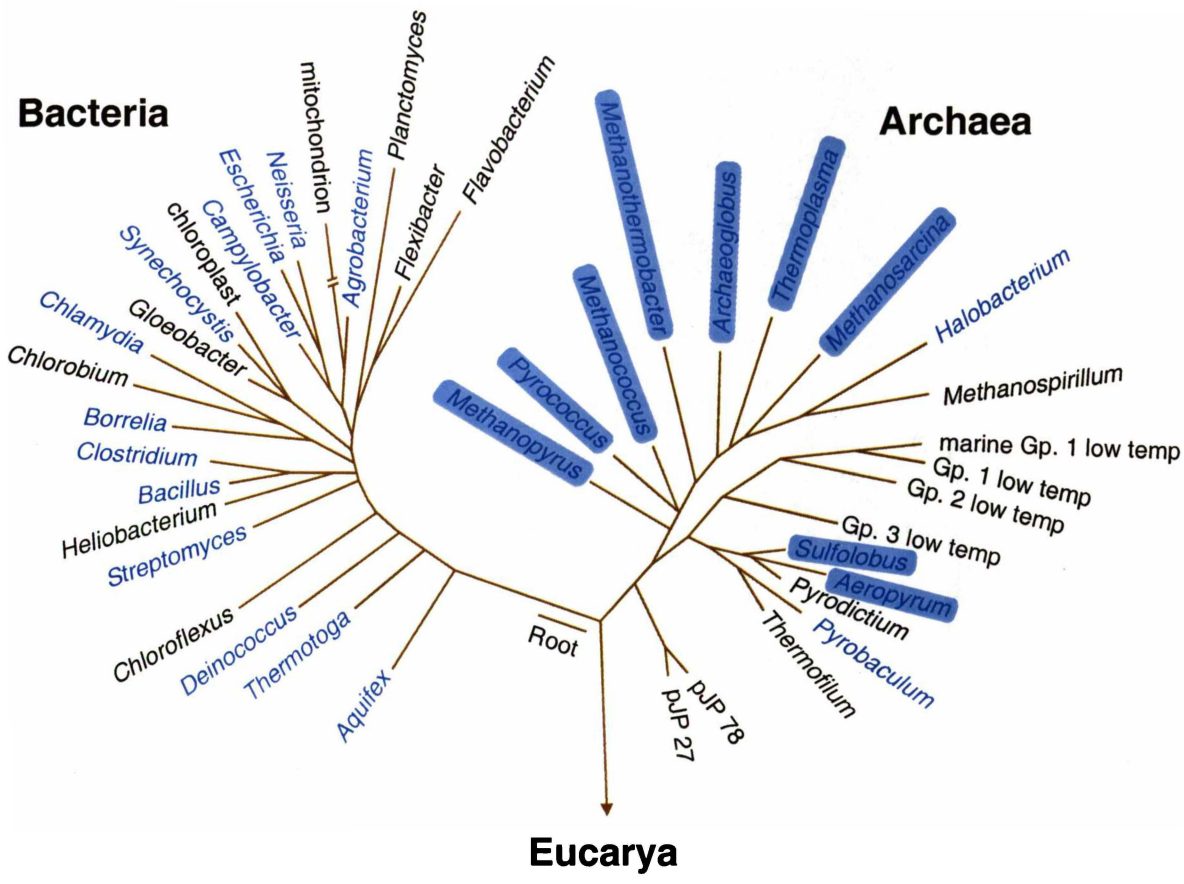


Figure 2-21. Phyletic distribution of the *sre* gene amongst extant life. Representatives from the Archaea and the Bacteria are shown on the universal phylogenetic tree based on small subunit ribosomal RNA sequences, as constructed by Norman Pace and colleagues (Pace 1997). Genera in which the complete genome has been sequenced for one or more of its members, at the time of this study, are shown in blue lettering. Genera in which one or more of its members is known to possess a gene encoding an SRE homologue are highlighted.

2.3.15 SRE sequences are very highly conserved

To further understand the relationship between the *sre* genes, the predicted amino acid sequences of the SRE proteins were aligned (Figure 2-22), and the resulting data used to construct a distance tree (Figure 2-23). As shown in the multiple alignment, the sequences of the SRE proteins are very highly conserved with more than 50% of the amino acid residues sharing similarity across all 12 sequences. The SRE proteins are all of very similar size, ranging from 84 to 93 amino acids in length, and with the exception of SREApe, are all acidic. A comparison of the SRE distance tree with a 16S rRNA tree with the same organisms, reveals that the topologies of the trees are relatively similar. This suggests that the *sre* genes are unlikely to have undergone any significant degree of lateral transfer. Consequently, it is also suggested that the cenancestor of the *sre* genes occurred in a very early archaeal lineage, prior to the bifurcation of the Crenarchaeota and the Euryarchaeota.

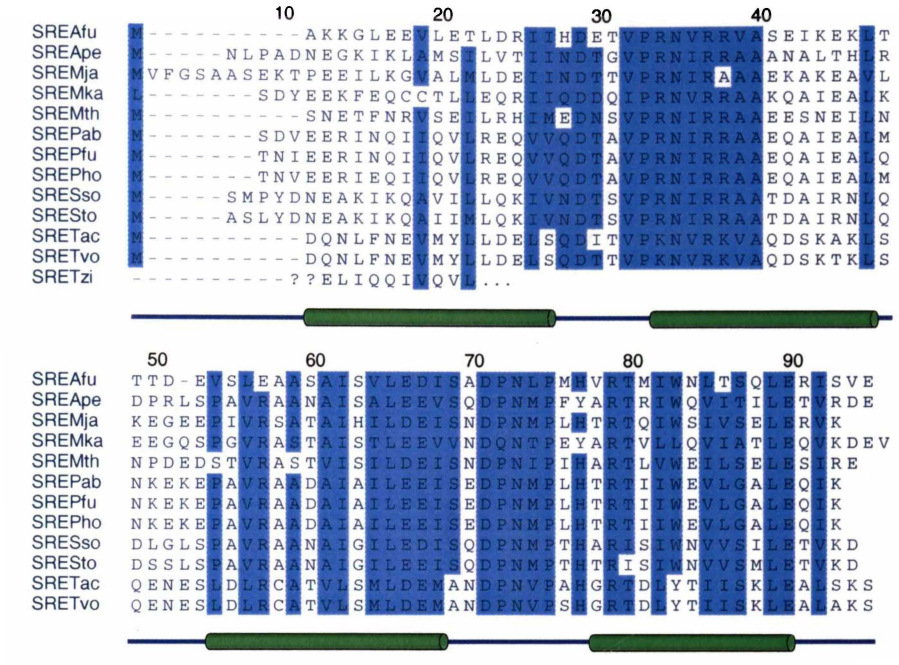


Figure 2-22. Multiple sequence alignment of predicted amino acid sequences of SRE and consensus predicted secondary structure. Similar residues are shown highlighted. Residues were considered similar if they were hydrophobic non-polar (A, V, I, L, P, M, F, W), polar non-charged (G, S, T, C, Y, N, Q), basic (H, R, K), or acidic (E, D) (Lehninger 1982; Karlin and Ghandour 1985). See Table 2-5 for the associated organism names and references. The consensus secondary structure, determined by Jnet (Cuff and Barton 1999) is aligned along the bottom of the sequences. α-helical residues are represented as green cylinders and coiled regions are represented blue lines.

2.3.16 SRE is predicted to consist of four helices interspaced with coiled regions

Using the Jnet algorithm with the multiple alignment of SRE amino acid sequences as the input data, the consensus secondary structure for the SRE proteins was computed (Figure 2-22). The predicted structure indicates the protein consists of four large α-helical regions, separated by short coiled regions. As shown in Figure 2-22, the second and fourth coiled regions and the beginning of the subsequent α-helical regions align with the most conserved residues found, suggesting that these structures play a critical role in the functioning and/or correcting folding of the protein. This is similar to the histone-fold motif, where the most conserved regions are also those involved in the coiled regions and start of the α-helices, since these are critical in determining the relative positions of the helical segments and in binding DNA (Arents and Moudrianakis 1995). The correct positioning of the archaeal histone helices is fundamental in quaternary structure formation and the non-specific binding to DNA (Arents and Moudrianakis 1995). In addition, the SRE sequences contain a high proportion of hydrophobic residues in the predicted helical

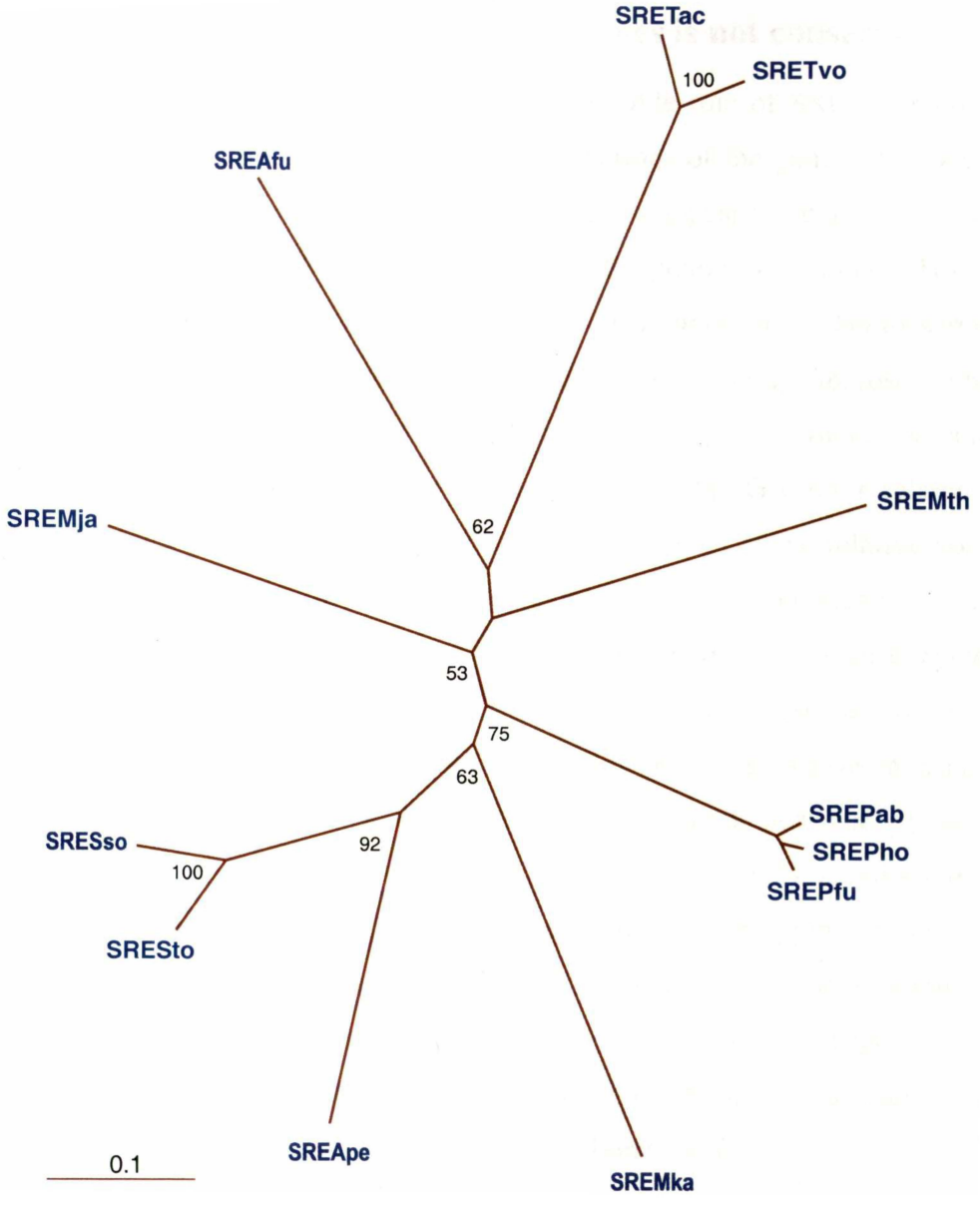


Figure 2-23. Distance tree of SRE protein sequences. The tree was inferred by neighbour-joining analysis. The percentage of 100 bootstrap resamplings that support each topological element is indicated at the branching points. The scale represents the expected number of amino acid changes per sequence position.

residues. This is also a characteristic of the histone-fold, where the hydrophobic residues in the central helix form the surface for dimerisation (Arents and Moudrianakis 1995). Consequently, it would seem feasible that the SRE proteins similarly form further quaternary associations and perhaps also bind to DNA.

Finally, since the SRE family represents a completely unique sequence-fold relationship, the structural determination of one of its members would be of considerable interest.

2.3.17 The position of *sre* relative to other genes is not conserved

In an attempt to gain more information about the possible role of SRE, a graphical map showing the positions, orientations and proposed functions of the genes neighbouring *sre* was constructed (Figure 2-24). Unfortunately, *sre* did not appear to be associated with any other genes in an operon-like arrangement, in any of the genomes examined. However, in each of the three methanogen genomes, *Methanococcus jannaschii*, *Methanothermobacter thermautotrophicus*, and *Methanopyrus kandleri*, the gene encoding ribulose 6-phosphate 4-epimerase occurs adjacent to the *sre* gene, though in the opposite orientation, and in the case of *M. kandleri* was actually found to overlap *sre* by 14 bp. Given the relatively large phylogenetic distance between these organisms, the persistence of the relative positioning of the genes seems surprising. Although a functional relationship between the products of these genes seems unlikely, since ribulose 6-phosphate 4-epimerase is an enzyme in the pentose phosphate pathway common to all known archaea and most bacteria, it may be possible that the molecular evolution of these two genes became somehow intermingled in an ancestor of the methanogens. Finally, of minor note, hypothetical coding regions are annotated in the genome that largely or entirely overlap *sreAfu*, *srePho*, and *sreSto*. Given the extent of the overlap and the fact that none of these overlapping genes have any identifiable homologues, it is most probable that these apparent ORFs occur only by coincidence and do not represent bona fide coding regions. However, it is also conceivable that regions produce transcripts employed for some form of antisense-mediated transcriptional regulation (Delihas 1995; Delihas and Forst 2001).

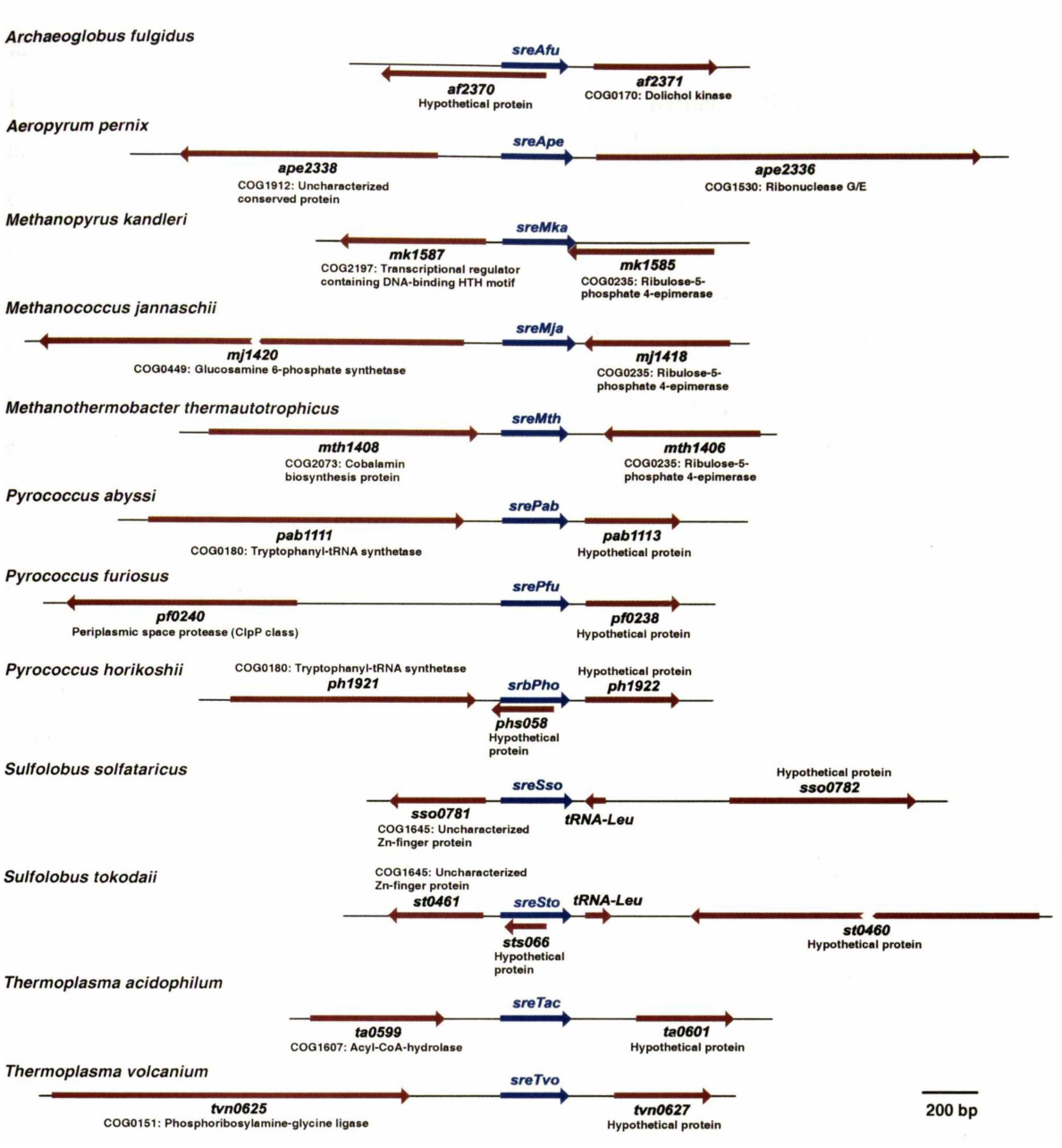


Figure 2-24. Comparison of relative positions, orientations and proposed functions of genes neighbouring or overlapping with *sre* in various genomes. The arrows represent the relative size, position and orientation of *sre* and the nearby genes. Overlapping genes are indicated by offset arrows. Functions for nearby genes were assigned using the Cognitor program at NCBI. Functions for genes that did not belong to any known COG were assigned according to the annotated data for the respective genomes from GenBank (NCBI). Genes are illustrated according to the scale shown, with the exception of those indicated with a break-point where the gene was resized to fit the figure.

2.4 Discussion

In the previous chapter, four stress-response proteins, named SRBTzi, SRCTzi, SRDTzi, and SRETzi, were identified in the thermophilic archaeon, *Thermococcus zilligii*. In this chapter, the computational analysis of the corresponding gene families, named *srb*, *src*, *srd* and *sre*, is described. Using the complete sequences of the closest identified homologues of the four different stress-response proteins as input, a combination of the COGnitor program and BLASTP and TBLASTN searches at GenBank were used to identify all current members of the four corresponding families. The large number of complete microbial genome sequences that are now freely available were utilised not only as a source for collecting sequence homologues of the four gene families of interest, but also to study the genomic context of the genes and their phyletic distribution amongst extant life. The relatively large number of homologous sequences available for each family also provided the opportunity to compare the relative differences between them and so begin building a picture of their evolutionary past. These analyses yielded some insight into which organismal lineages the genes first originated, at which stage or stages the genes were duplicated, and whether or not they are likely to have undergone horizontal gene transfer and if so, the most probable direction of such an event.

The members of the *srb* family originated mostly from the majority of known archaeal genomes, but were also found in the bacteria, *Thermotoga maritima* and *Clostridium acetobutylicum*. The amino acid sequences encoded by the *srb* genes were only relatively weakly conserved, with only a few positions sharing similarity amongst all members of the family. As a consequence of the diversity of the SRB sequences, the boundaries of the family are somewhat difficult to define and other factors such as the overall length of the sequence and similarity with multiple other members were used to validate *srb* family members. Without the knowledge of the structures or functions of the various SRB family members, it is difficult to ascertain whether the diversity of the sequences is a reflection of different structural adaptations to fulfil different functions, or rather an indication of a loose structural constraint to fulfil a particular function. The comparison of the 3D structure of SRBMth1, the only elucidated structure of an SRB protein, with the predicted secondary structures of other SRB proteins, suggested that the other proteins did exhibit a considerable degree of, at least, secondary structural homology. On the other hand, given that many of the *srb*-containing genomes examined contained more than one homologue of *srb*, it would seem reasonable to assume that these often substantially different proteins,

according to sequence similarity, do provide distinctive functions. Clearly, it will be necessary, ultimately, to study the function of a variety of different members of the SRB family, before the impact of the diversity of sequences on their function can be appreciated.

Among the archaeal representatives of the *srb* family, all belonged to the Euryarchaeota, with the exception of *Sulfolobus tokodaii*, which contained a single *srb* homologue. In light of the phyletic distribution of *srb*, it was suggested that the gene family originated in the euryarchaeal lineage, and was later transferred on several independent occasions to other organisms, including *S. tokodaii*, *Thermotoga maritima* and *Clostridium acetobutylicum*. Interestingly, *C. perfringens* contained a gene (*cpe1031*) that was apparently a fusion of an *srb* gene and a gene encoding a negative regulator of the arsenic resistance operon (*arsR*). Since this gene no longer appeared to be part of an operon of any sort, nor was neighboured by any genes associated with arsenic resistance, it was concluded that the *arsR* transcriptional regulator had somehow been recruited in a new role, in conjunction with *srb*. The examination of the role of CPE1031 may provide valuable insights into the function of SRB in general. Moreover, it would be interesting to investigate the possibility that SRB associates with transcriptional regulator proteins like ArsR in other organisms, since perhaps CPE1031 may represent an efficient adaptation of the sort of quaternary associations that SRB assumes in other organisms. Finally, the presence of both *cpe1031* in *C. perfringens* and *srbCac* in *C. acetobutylicum*, may prove to be extremely valuable in the study of the function of SRB, since, unlike the archaeal representatives of *srb*, both of these species have well-established recombinant systems available for their study, making mutational, knockout and over-expression studies of SRB very feasible.

In examining the genomic context of the *srb* genes, it was found that *srb* sometimes occurred as part of a putative operon with genes encoding MinD-like ATPases (tentatively named *minDfer*). The conservation of *minD* and *minD*-like genes in all three domains of life suggests that these proteins play a central role in cellular metabolism (Gerard *et al.* 1998). MinD ATPases have been characterised for their involvement in organising the correct placement of the site for cell division (de Boer *et al.* 1991). However, the MinD homologues observed in association with *srb* contain a ferredoxin domain, and to date have not been characterised. Furthermore, the absence of other cell-division related genes that typically occur together with *minD* in the *minB* operon (Cook *et al.* 1989; de Boer *et al.* 1990), suggests that MinDfer fulfils a different function to MinD. Given the myriad roles

that both ATPase and ferredoxin domains are involved in, it is very difficult to predict what the function of the MinDfer proteins may be, or how they are involved with SRB. However, the observation that in *A. fulgidus* and *M. thermautotrophicus*, the putative *srb*-containing operon also contains a gene encoding a protein predicted to be involved in chromosome partitioning, supports the notion that, at least in some instances, the operon is somehow associated with cell division. Upon confirmation that these genes are indeed expressed as an operon, the further study of the function of MinDfer and the other members of the operon will undoubtedly provide a critical insight into the role of SRB in these organisms.

The *src* family was found to occur exclusively within the Archaea, with a total of 18 members being identified in 13 different species. The SRC family is the same as the Sac10b family of DNA-binding proteins (Forterre *et al.* 1999). Although, Sac10b has been shown to bind DNA, it does not compact DNA, and has been suggested to play a role in recombination and repair processes (Lurz *et al.* 1986). The Sac10b family has previously been claimed to represent the first small, basic DNA-binding proteins to be both ubiquitous in archaea and specific to this domain (Forterre *et al.* 1999). However, with the completely sequenced archaeal genomes that have become available since this report, it is shown here that *src* genes are absent in the archaeal genomes of *Halobacterium* sp. NRC-1, *Methanopyrus kandleri*, *Methanosarcina mazei*, and *M. acetivorans*. Apart from their phylogenetic relationship, there are no other immediately apparent physiological or environmental characteristics that would seem to connect any of these four organisms, and indeed, no other genes match the phylogenetic pattern of the *src* family.

It is intriguing that among the Archaea, in addition to the SRC family, there are several other small, basic, highly abundant DNA-binding proteins, each with their own distinctive phylogenetic distributions. That is, using the currently available genomes, the archaeal histones (Sandman *et al.* 1990) occur exclusively within the Euryarchaeota and are absent only from members of the Thermoplasmatales; the Sac7d family (Dijk and Reinhardt 1986) occurs only within members of the Sulfolobales; and the MC1 proteins (Laine *et al.* 1991) are present only in *Halobacterium* and the Methanosarcinales. Considerable further research is required to understand the functional niches fulfilled by these and other as yet undiscovered proteins, and how or if these proteins complement or substitute one another's respective roles.

The *srd* family presented an unusual and unique phyletic distribution, in which horizontal transfer is likely to have played a significant role. Genes encoding SRD were found in the sequenced genomes of all non-methanogenic archaea, all four Rhizobiacae, *Streptomyces coelicolor* and *Deinococcus radiodurans*. The predominance of the genes occurring amongst members of the Archaea is a likely indication that the gene originated early in the archaeal lineage, and was later transferred in one or more horizontal transfer events to some members of the Bacteria. Given the apparent ubiquity of the gene in all other archaea the absence of the *srd* genes in the methanogens is intriguing. The most likely scenario to explain this, is that the cenancestor of the methanogens, which on the basis of whole genome analyses are indeed believed to be monophyletic (Slesarev *et al.* 2002), had lost its *srd* gene. It is also interesting that the Rhizobiacae all contain the *srd* genes. Presumably this was the consequence of the lateral transfer of an *srd* gene into the cenancestor of the Rhizobiacae. Naturally, this raises the issue as to how and why the methanogens became independent of SRD and the Rhizobiacae became, presumably, dependent on SRD. The characterisation of the function of the *srd* genes will be critical in understanding the unusual phyletic distribution of the *srd* genes. In addition, the presence of an *srd* gene in *Agrobacterium tumefaciens* may prove valuable in elucidating its function, since techniques to create gene mutants and knockouts are well-established in this organism.

As stated in the previous chapter, members of the SRD family are wholly homologous with the C-terminal half of the Lrp/AsnC proteins. The recent structural determination of LrpA, an Lrp homologue from *Pyrococcus furiosus* (Leonard *et al.* 2001), not only provides support for the predicted structures of the SRD proteins, but also vividly demonstrates that the C-terminal half of the LrpA forms a domain completely distinct to the N-terminal half of the protein. Leonard and co-workers also showed that the C-terminal region is primarily responsible for the formation of quaternary associations and suggested that the N-terminal domain, which contains a helix-turn-helix motif, was involved in DNA-binding. This structural and functional distinction of the N- and C-terminal halves of LrpA, further justifies the separation of the Lrp/AsnC and SRD protein families. Moreover, it provides some helpful insights into the possible behaviour of the SRD proteins. Beyond forming various multimeric species, LrpA has also been suggested to form macromolecular complexes during transcriptional regulation (Leonard *et al.* 2001). Consequently, it would seem feasible that the SRD proteins may also associate to form multimers, and perhaps also interact with the transcriptional machinery. Indeed, the C-terminal domain of

papillomavirus-1 E-2 protein, which shares the βαββαβ topology of the C-terminal half of LrpA and presumably the SRD proteins, has been indicated to interact with DNA-bound TATA binding protein and stimulate transcription (Steger *et al.* 1995). The observation that the *srd* gene is often included in potential operon-like arrangements, mostly encoding proteins of unrelated functions, prompts the possibility that SRD, in these instances, has a role involved in transcriptional autoregulation. i.e. SRD is produced to control the transcription of other members of the operon. To examine these intriguing possibilities, further *in vivo* and *in vitro* studies will be required. The examination of the activity and role of the SRD proteins may also aid in the further understanding of the mechanism and evolution of the Lrp/AsnC proteins.

The SRE family contains a distinct group of highly conserved exclusively archaeal proteins. One member of the *sre* family was found in every archaeal genome, with the exceptions of *Pyrobaculum aerophilum* and *Halobacterium* sp. NRC-1. The observation that the branching topology of the SRE distance tree was reasonably consistent with the corresponding 16S rRNA tree, suggests that the *sre* genes have not undergone any significant degree of horizontal transfer. Furthermore, the data indicates that it is likely that the cenancestor of the *sre* genes originated early in the archaeal lineage, and that relatively recent ancestors of organisms such as *P. aerophilum* and *Halobacterium* sp. NRC-1 lost the gene. The remarkably high degree of conservation among the SRE sequences leads to the suggestion that the protein may act in an intricate manner with a larger conserved piece of molecular machinery, such as the transcriptional or translational apparatus, or perhaps form complex higher-order associations with DNA, such as nucleosome-like structures. The notion of the SRE proteins binding to DNA is also supported by their predicted structure, which consists of two helix-turn-helix regions. The helix-turn-helix motif is one of the paradigms of DNA-protein interactions, and is found in a wide range of DNA-binding proteins, that includes proteins such as histones (Luger *et al.* 1997), the TFIIB family of transcriptional regulators (Nikolov *et al.* 1995), and the cro and CI repressor proteins of bacteriophage λ (Anderson *et al.* 1981; Pabo and Lewis 1982).

It is intriguing that members of the phylogenetically diverse and highly conserved *sre* family are absent from the two distantly related archaea, *P. aerophilum* and *Halobacterium* sp. NRC-1. No other genes were identified that matched the phylogenetic pattern of the *sre* family, nor are there any other apparent similarities between *P. aerophilum* and *Halobacterium* sp. NRC-1 that would set them apart from the other archaea. Unfortunately,

since the *P. aerophilum* genome data is not currently included in the COG database, it was not feasible to identify the genes that were exclusive to both *P. aerophilum* and *Halobacterium* sp. NRC-1. In the future, when the *P. aerophilum* and other archaeal genomes are added to the COG database, other characteristics that tie these two organisms together may aid in the understanding of the role of the SRE proteins.

To conclude, the computational analysis of the four gene families, *srb*, *src*, *srd* and *sre*, has proved to be a powerful tool in gaining an insight into the structures, possible functions and molecular histories of these genes. However, the results of the analysis also highlight many new questions and serve as a reminder to the considerable amount that is still unknown about not only the functions of many gene families, but also the minimal understanding of the relationship between protein structure and function. So, despite the importance and far-reaching implications of *in silico* research, the limits of its predictive power still remain relatively fixed by functional *in vivo* and *in vitro* studies of protein structure and function.

The following chapter describes the production and purification of a recombinant representative from each of the SRB, SRD, and SRE families, in *Escherichia coli*. Purified forms of each of the proteins can be utilised for future structural studies as well as allow the screening of their possible activities and biochemical properties.

CHAPTER THREE

Expression and Purification of Four Novel Recombinant Proteins from *Pyrococcus abyssi*, rSRBPab1, rSRBPab2, rSRDPab2 and rSREPab

Summary

In the previous chapter, it was found that the respective families of three of the four proteins identified in the stress-response of *Thermococcus zilligii*, namely SRB, SRD and SRE, are essentially undescribed in the literature. To learn about the role of these proteins, it seemed pertinent to examine the *in vitro* activities of some of the members of these novel families. To facilitate this, genes encoding members of the SRB, SRD and SRE families from *Pyrococcus abyssi* were cloned into an *Escherichia coli* expression vector, pET-19b. The four genes, *srbPab1*, *srbPab2*, *srdPab2* and *srePab*, were successfully expressed in *E. coli* in substantial amounts. Using the $(His)_{10}$ tag fused at the N-terminus of the resulting recombinant proteins, the proteins were purified to homogeneity by metal chelation chromatography. rSRBPab1, rSRBPab2, and rSREPab were purified in non-denaturing conditions, while rSRDPab2 was purified in denaturing conditions. The buffers that rSRBPab1 and rSREPab were purified in were changed by careful dialysis. Upon dialysing rSRBPab2, the protein became insoluble, however, it was found that the protein could be resolubilised by heating for 15 min at 85°C. After purification, rSRDPab2 required careful removal of the denaturant, and the presence of CHAPS buffer at 10 mM concentration was necessary for effective solubilisation.

3.1 Introduction

In the previous chapter, the sequences of all identifiable homologues of SRBTzi, SRCTzi, SRDTzi and SRETzi were extracted from the publicly available complete genome sequences. It was found that the SRB family was a diverse family of proteins that occurred primarily amongst the Euryarchaeota. Apart from the solution structure of an SRB protein from *Methanothermobacter thermautotrophicus*, the SRB proteins are to date undescribed and not substantially related to any characterised protein family. The SRC family was found to be the same as the Sac10b family of DNA-binding proteins, which have been relatively well described previously (Dijk and Reinhardt 1986; Forterre *et al.* 1999; Xue *et al.* 2000). The SRD family was found to form a distinct, hitherto uncharacterised subgroup of the Lrp/AsnC family of transcriptional regulators (see 2.3.10 and 2.3.11). The SRD proteins were distinguished from the Lrp/AsnC proteins (Newman and Lin 1995) in that they wholly consisted of only the C-terminal domain of these proteins, and hence lacked the characteristic N-terminal helix-turn-helix region of the Lrp/AsnC proteins that is thought to bind DNA (Leonard *et al.* 2001). Finally, the SRE proteins were found to represent a very highly conserved family of proteins that occurred exclusively in the archaeal domain. To date, no member of the SRE family has been studied.

To gather some insight into the possible role of the less well-described families, i.e. the SRB, SRD and SRE families, an examination of the *in vitro* activity of the individual members is pertinent. *In vitro* studies provide the opportunity to make a controlled examination of a protein's activity and biochemical properties, such as its capacity to form higher order structures with itself or other proteins, or its ability to bind and/or manipulate nucleic acids and so forth.

To obtain sufficient quantities and relatively homogeneous samples of a protein, for which the activity and biochemical characteristics are not known, the most elementary approach is to clone the respective gene into an *Escherichia coli* expression system that provides a fusion tag to allow simple purification of the target protein. Such expression systems obviate the need to optimise production of the protein in the native organism or develop potentially expensive, time-consuming and sophisticated purification strategies. Rather than attempting to identify and clone the corresponding genes of SRBTzi, SRDTzi and SRETzi, it was considered that studying the genes from an organism for which the complete genome sequence was known, would potentially prove more insightful and

relevant to future research. In addition, the availability of a complete genome sequence facilitates the PCR amplification of the target genes for cloning.

Amongst the complete genome sequences available at the time of this study, the closest relatives to *T. zilligii*, according to 16 rRNA sequence comparison, are the three members of the genus *Pyrococcus*, i.e. *P. abyssi*, *P. furiosus* and *P. horikoshii* (Ronimus *et al.* 1997). Moreover, the N-terminal sequences of SRBTzi, SRDTzi and SRETzi were also most similar to those from members of the genus *Pyrococcus* (see Figures 1-5, 1-6 and 1-8). Due to the ready availability of culture and experience in optimising its growth, *P. abyssi* was chosen as the target organism from which to clone the genes with closest homology to SRBTzi, SRDTzi and SRETzi.

In this chapter, the cloning of the *srbPab1*, *srbPab2*, *srdPab2*, and *srePab* genes from *P. abyssi*, into the pET-19b vector (Novagen) is described. The subsequent expression, purification, concentration and quantitation of the resulting proteins, rSRBPab1, rSRBPab2, rSRDPab2, and rSREPab, is also outlined. It is hoped that the availability of these recombinant proteins, and the established system for their production, will set a foundation from which the functions of these novel protein families can be examined.

Some of the results from this chapter were presented at the *Eleventh Annual Queenstown Molecular Biology* conference (Appendix F).

3.2 Materials and Methods

3.2.1 Culturing *Pyrococcus abyssi*

Pyrococcus abyssi was cultured essentially as described (Erauso *et al.* 1993). The mineral base used for the preparation of the media was prepared as a 5x stock solution that contained: 2.0 M (117 g/L) NaCl, 265 mM (54.0 g/L) $MgCl_2 \cdot 6H_2O$, 140 mM (20.0 g/L) Na_2SO_4, 271 mM (20.0 g/L) KCl, 13.5 mM (1.0 g/L) $NaHCO_3$, 7.0 mM (1.0 g/L) $CaCl_2 \cdot 2H_2O$, 4.0 mM (0.45 g/L) KBr, 0.47 nM (0.125 g/L) $SrCl_2 \cdot 6H_2O$, 2.0 nM (0.125 g/L) H_3BO_3, and 0.325 nM (0.015 g/L) NaF. The 5x mineral base solution was sterilised by autoclaving at 121°C for 20 min. To a 1 L Schott bottle, 170 mL 5x mineral base, 10.2 g cystine (as a sulfur source), 0.85 g Bactopeptone (Difco), 3.4 g Bacto yeast extract (Difco), 2.57 g piperazine-N,N'-bis-(ethanesulfonic acid) (PIPES; final concentration 10 mM) and 0.85 mL 0.1% (w/v) resazurin (as a redox indicator), was added. The volume was

made up to 830 mL with deionised water (>18 MΩ resistance). The pH of the medium was adjusted to 6.8 with KOH pellets. The bottle was sealed with a rubber stopper adapted with a rubber septum, and sterilised by autoclaving at 121°C for 20 min. The solution was left to cool to approximately 80°C, before adding 8.5 mL Zeikus' trace elements solution (Zeikus *et al.* 1979), and 10 mL of a solution containing 518 mM (70.0 g/L) KH_2PO_4 and 382 mM (140 g/L) NH_4Cl (sterilised by autoclaving at 121°C for 20 min). The headspace of the bottle was flushed with 0.2 μm filtered nitrogen gas for approximately five minutes or until the resazurin indicator turned fully yellow. Final anaerobiosis was achieved by the addition of 2 mL 5% (w/v) $Na_2S{\cdot}9H_2O$ (neutral and filter sterilised), using a syringe. The bottle was then placed in a 90°C incubator for approximately one hour, before it was inoculated.

The medium was inoculated by addition of 5 mL of *P. abyssi* strain Orsay (a generous gift from Patrick Forterre, Université Paris-Sud XI, Orsay-Cedex, France). The culture was grown overnight at 90°C, to a final optical density at 600 nm (OD_{600}) of approximately 0.3. To harvest the cells, the culture was transferred into two 450 mL centrifuge bottles and centrifuged at 11,500 *g* (8,000 rpm in a JA10 rotor) at 15°C for 15 min. The supernatant was virtually all poured off, and the cell pellet resuspended in the remaining volume. The concentrated solution was transferred to two Beckman JA20 (25 mL) centrifuge tubes and phosphate-buffered saline (PBS; 100 mM K_2HPO_4, 0.05% (w/v) KCl, 0.8% (w/v) NaCl) was added to almost fill the tubes. The tubes were centrifuged at 17,400 *g* (12,000 rpm in a JA20 rotor) at 15°C for 15 min. The supernatant was removed and the cells transferred to 1.5 mL tubes, and stored at -70°C.

3.2.2 Isolation of *Pyrococcus abyssi* DNA

P. abyssi genomic DNA was isolated from stored cells, using an organic solvent extraction method based on established methods as described (Sambrook *et al.* 1989). In a 1.5 mL Eppendorf tube, 100 mg of *P. abyssi* cells were resuspended in 450 μL of lysis solution (20 mM Tris-HCl pH 7.6, 100 mM EDTA pH 8.0, 0.2 mg/mL Proteinase K (Sigma-Aldrich), and 1% (w/v) SDS), and incubated at 55°C for 4 h, with occasional gentle mixing. To extract proteins and cellular debris, an equal volume of 25:24:1 (v/v) phenol/chloroform/isoamyl alcohol was added and mixed in thoroughly by shaking. The organic and aqueous phases were separated by centrifugation at 18,000 *g* at 4°C for 10

min. The top (aqueous) layer was transferred to a new 1.5 mL tube and an equal volume of 24:1 (v/v) chloroform/isoamyl alcohol was added and mixed. The phases were separated as before, and the top layer was again transferred to a new tube. Nucleic acids were precipitated by addition of 0.04 volumes of 5 M NaCl and 2.2 volumes of 100% ethanol, mixing the resulting solution by inversion, then incubating the sample at -20°C for 30 min. The nucleic acid was pelleted by centrifugation at 12,700 *g* at 4°C for 10 min. The pellet was washed with 1.2 mL of 70% ethanol, then vacuum dried with centrifugation (Savant SpeedVac SC110 with Savant Gel Pump GP110) for 5 min. The nucleic acid was resuspended in 400 µL TE buffer (10 mM Tris-HCl pH 7.6, 1 mM EDTA pH 8.0) with 50 µg/mL RNAase A and incubated at 37°C for 2 h. The protein was extracted with phenol/chloroform/isoamyl alcohol, and the DNA precipitated with NaCl and ethanol, as described above. Finally, the pellet was washed with 70% ethanol, dried under vacuum centrifugation, and resuspended in 100 µL of TE buffer.

The concentration of the DNA sample was determined by spectrophotometry, as described (Sambrook *et al.* 1989). The DNA sample was diluted 1 in 150 in deionised water (to a final volume of 750 µL) in an 800 µL quartz cuvette and mixed. The absorbance was measured using a Bio-Rad SmartSpec™ 3000 spectrophotometer, with deionised water in a matching cuvette as the reference. The purity of the DNA sample was determined by calculating A_{280}/A_{260} (a value between 1.6 and 1.8 indicating moderate quality), and the quantity estimated using A_{260} (a value of 1.0 being equivalent to approximately 50 µg/mL).

Genomic DNA was stored at -20°C.

3.2.3 Amplification and purification of target DNA

The genes targeted for cloning were amplified from *P. abyssi* genomic DNA using the polymerase chain reaction (PCR). In a 0.2 mL tube, 40 µL PCR master mix (3 mM $MgCl_2$, 12 mM Tris-HCl pH 8.3, 60 mM KCl, 0.25 mM dNTPs), 5-10 ng genomic DNA, 1 µL each of the forward primers and reverse primers (10 µM; Sigma-Aldrich), and 1 µL 1 U/µL AmpliTaq Gold DNA polymerase (Perkin Elmer) were mixed and the volume made to 50 µL with deionised water (>18 MΩ resistance). The forward and reverse primers that were used to amplify the target genes, and the expected size of the resulting amplicon, are shown in Table 3-1. *Nde*I and *Bam*HI restriction sites were incorporated into the forward and reverse primers, respectively, to enable the subsequent directional cloning of the

Primer name	Primer sequence (5' → 3')	Amplicon length (bp)
srbPab1 Fwd	GTGATG**CATATG**AGGATCGCGGTTCCAACTAA	678
srbPab1 Rev	TTCCTT**GGATCC**CATCCGGGGCCTCTTC	
srbPab2 Fwd	GGTGGG**CATATG**AGGTTCATAGTGGCAAC	442
srbPab2 Rev	ACTATG**GGATCC**TTAACTACCATCCTCC	
srdPab2 Fwd	GATGTA**CATATG**GTGACGGCTTTTATTTTGAT	284
srdPab2 Rev	CTCATA**GGATCC**GAAGAACAACCAGAAA	
srePab Fwd	GTGATT**CATATG**AGTGACGTAGAAGAGAGGAT	302
srePab Rev	GAGAAG**GGATCC**ATAAAATTAAGGTCTA	

Table 3-1. Forward (Fwd) and reverse (Rev) primers used for the PCR amplification of the *srbPab1*, *srbPab2*, *srdPab2* and *srePab* genes from *P. abyssi*. The introduced *Nde*I and *Bam*HI recognition sites used for the directional cloning into the pET-19b expression vector are shown in bold print. The expected lengths of the amplicons, determined from the genome sequence data of *P. abyssi* (GenBank accession NC_000868), is also shown. Primers were designed with the aid of Genamics Expression 1.1 (Genamics), using the appropriate screening parameters to optimise their efficiency.

amplicons. PCR amplifications were performed using a PTC-100™ Thermal Cycler (MJ Research, Inc.), programmed with an initial enzyme activation period of 10 min at 94°C, followed by 35 cycles of the following steps: 94°C for 30 s, 53°C for 30 s, and 72°C for 45 s.

After the PCR amplification was complete, the DNA was purified and concentrated using the Concert Rapid PCR Purification System (Life Technologies), according to the manufacturer's instructions. The size of the PCR product was checked by comparing its relative electrophoretic mobility through a 2% agarose (SeaKem LE, FMC Bioproducts) gel in 1x TBE (0.089 M Tris-HCl, 0.089 M boric acid, and 10 mM EDTA pH 8.0) to that of a standard 100 bp ladder (Life Technologies). DNA was electrophoresed in the presence of ethidium bromide to allow it to be visualised by fluorescence under UV light. The concentration of the PCR product could also be estimated by comparing its relative intensity to a similarly sized DNA fragment in the standard 100 bp ladder.

3.2.4 Cloning of target genes

The genes targeted for expression in *E. coli* were each cloned into a pET-19b expression vector (Novagen). The resulting open reading frame (ORF) constructed in the vector would encode a series of histidine residues and an enterokinase cleavage recognition site fused at

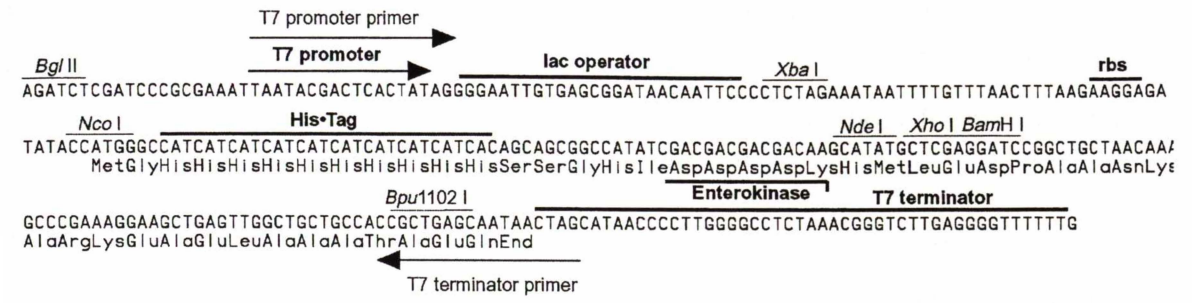


Figure 3-1. Sequence detail of the cloning and expression region of the pET-19b vector. The locations of the T7 promoter, lac operator, ribosome binding site (rbs), His·Tag, enterokinase cleavage site, and T7 terminator are indicated. Figure adapted from Novagen's technical literature.

the N-terminus of the protein to be synthesised. The pET-19b cloning/expression region is illustrated in Figure 3-1. All manipulations were carried out according to PC-2 guidelines and approved by ERMA New Zealand (Approval Number GMD000050).

In separate tubes, the pET-19b plasmid DNA and the DNA to be inserted were digested with the restriction enzymes, *Nde*I and *Bam*HI. To a 0.6 mL tube, either 200 ng pET-19b plasmid DNA or 100 ng insert DNA was added, together with 1 U *Nde*I (Roche Molecular Diagnostics), and 2 μL 10x suRE Cut Buffer B (10 mM Tris-HCl pH 8.0, 5 mM $MgCl_2$, 100 mM NaCl, 1 mM β-mercaptoethanol; Roche Molecular Diagnostics), and the volume made to 20 μL using deionised water (>18 MΩ resistance). After, the tubes were incubated at 37°C in a Compact Thermomixer (Eppendorf) for 1 h at 600 rpm, 1 U *Bam*HI (Roche Molecular Diagnostics) was added and the tubes incubated for a further 1 h.

After the incubation was complete, the samples were purified by agarose gel electrophoresis. The entire sample volumes were loaded on separate lanes of a 1% agarose gel in 1x TBE and electrophoresed at 10 V/cm for 30 min, in the presence of ethidium bromide. After the positions of the bands were determined under UV light, the gel regions containing the DNA were excised with sterile razors and placed into 1.5 mL tubes. The DNA was then extracted from the gel pieces using the QIAEX II Agarose Gel Extraction system (Qiagen), according to the manufacturer's instructions.

The gel-purified *Nde*I-*Bam*HI digested vector and insert were ligated according to established methods (Sambrook and Russell 2001). In a 0.6 mL tube, the following reagents were mixed: 2 μL 10x ligase buffer (200 mM Tris-HCl pH 7.6, 100 mM $MgCl_2$, 250 μg/mL acetylated bovine serum albumin (BSA); Roche Molecular Diagnostics), 2 μL 100 mM dithiothreitol (DTT), 1 μL 1 U/μL T4 DNA ligase (Roche Molecular Diagnostics), 1 μL 10 mM ATP, 50 ng (0.014 pmol) *Nde*I-*Bam*HI digested pET-19b

vector, 10 ng (~0.07 pmol) *Nde*I-*Bam*HI digested insert DNA, and deionised water to a final volume of 20 μL. The reaction was incubated overnight at 20°C. A control reaction was also performed, in which the insert DNA was omitted, to check for non-recombinant background.

The products of the ligation reactions were transformed by electroporation using conventional methods (Sambrook and Russell 2001). To maximise transformation efficiency and plasmid stability, *E. coli* DH5α (Invitrogen) was selected as the host strain, since it lacks T7 RNA polymerase and is *recA- endA-*. On ice, in pre-chilled 0.1 cm electrode cuvettes (Bio-Rad), 50 μL of electrocompetent *E. coli* DH5α cells (prepared in advance as described (Jessen 1999)) were added to 1 μL of each of the ligation reaction products. After mixing briefly with a pipette, the samples were electroporated using a Bio-Rad Gene Pulser™ electroporator, with voltage set at 1.8 V, resistance set at 200 Ω, and capacitance set at 250 μFD. Immediately after electroporation, 1 mL of Luria-Bertani medium (LB; 1% (w/v) Bacto tryptone, 0.5% (w/v) yeast extract, 1% (w/v) NaCl, sterilised by autoclaving at 121°C for 20 min) was added to each sample, and the contents transferred to 1.5 mL tubes. The tubes were incubated at 37°C in a Compact Thermomixer (Eppendorf) for 1 h at 600 rpm, before spreading 100 μL of the culture onto LB agar plates containing 75 μg/mL ampicillin. The plates were then incubated overnight in a 37°C incubator.

The resulting colonies were screened for successful transformants by PCR. Eight colonies from each plate were extracted using sterile pipette tips and suspended in 20 μL of LB medium in 0.6 mL tubes and stored at 4°C. In 0.2 μL tubes, 1 μL of cells was added to 40 μL PCR Master Mix (3 mM $MgCl_2$, 12 mM Tris-HCl pH 8.3, 60 mM KCl, 0.25 mM dNTPs), 1 μL of each of the appropriate forward and reverse primers (10 μM; Sigma-Aldrich), and 1 μL 1 U/μL *Taq* DNA polymerase (Roche Molecular Diagnostics) and deionised water (>18 MΩ resistance) to a final volume of 50 μL. PCR amplifications were performed using a PTC-100™ Thermal Cycler (MJ Research, Inc.), programmed with an initial denaturation period of 2 min at 94°C, followed by 35 cycles of the following steps: 94°C for 30 s, 53°C for 30 s, and 72°C for 45 s. The products of the PCR were analysed by agarose gel electrophoresis as described in 3.2.3. One or more of those colonies that yielded PCR products of the expected size were then used for a small scale preparation of plasmid DNA.

To provide a suitable quantity of plasmid DNA for further manipulations, cells from colonies identified to contain the target gene were used as inoculum to produce a culture of larger volume. Approximately 10 μL of the remaining cell suspension was used to inoculate 10 mL of LB medium, containing 100 μg/mL ampicillin, which was incubated overnight at 37°C with shaking at 250 rpm in an Innova™ 4300 Incubator Shaker (New Brunswick Scientific). The cells were pelleted by centrifugation at 6000 *g* for 10 min, at 4°C. The plasmid DNA was then isolated and purified using the Concert Rapid Plasmid Miniprep System (Invitrogen), according to the manufacturer's instructions. Purified plasmid DNA was stored in TE buffer (10 mM Tris-HCl pH 8.0, 1 mM EDTA pH 8.0) at -20°C.

To check whether the engineered plasmids had undergone any major rearrangements and that the target genes were in the expected positions in the vector, the plasmids were examined by restriction enzyme analysis. Restriction enzymes that provided a diagnostic fragment pattern were identified using Genamics Expression 1.1 (Genamics). Restriction enzyme digests were performed according to standard procedures, and the resulting fragments examined by agarose gel electrophoresis. The sizes of the fragments were estimated by comparing their relative electrophoretic mobility to 1 kb and 100 bp standard ladders.

Finally, to check for the introduction of any mutations in the target genes during PCR or cloning, the cloning/expression region of the recombinant pET-19b vectors was sequenced using the T7 promoter forward and T7 terminator reverse primers (Waikato DNA Sequencing Facility, The University of Waikato). The relative annealing positions of the primers in the pET-19b cloning/expression region are indicated in Figure 3-1.

3.2.5 Expression of recombinant genes

The genes cloned into the pET-19b vectors were expressed in *E. coli* BL21(DE3)pLysS (Novagen) according to the pET System Manual (Novagen). *E. coli* BL21(DE3)pLysS contains a copy of the gene for T7 RNA polymerase, which is required for transcription of the cloned genes. Recombinant pET-19b vectors were transformed into electrocompetent *E. coli* BL21(DE3)pLysS (prepared in advance, as described (Jessen 1999)) by electroporation as described in 3.2.4. Transformants for single colonies were streaked onto

LB agar plates containing 100 μg/mL ampicillin, and used to prepare glycerol stocks (stored at -70°C). The streaked plates were incubated overnight at 37°C and stored at 4°C.

Single colonies from freshly streaked plates were used to inoculate 120 mL volumes of LB containing 50 μg/mL ampicillin, in 500 mL Erlenmeyer flasks. The flasks were incubated at 37 °C with shaking at 250 rpm, in an Innova™ 4300 Incubator Shaker (New Brunswick Scientific). The optical density at 600 nm (OD_{600}) of the cultures was measured at regular intervals using a Bio-Rad SmartSpec™ 3000 spectrophotometer. When the OD_{600} of the cultures reached approximately 0.6, 20 mL of the cultures were transferred to 100 mL Erlenmeyer flasks as uninduced controls. To the remaining 100 mL of culture, isopropyl-β-D-thiogalactopyranoside (IPTG) was added to a final concentration of 1 mM from a 100 mM stock solution. Incubation of the induced and non-induced flasks was continued for a further 3 h.

When the cells were ready to be harvested, the flasks were put on ice for 5 min, then pelleted by centrifugation at 8000 *g* for 5 min at 4°C. After removing the supernatants, the pelleted cells were resuspended in 0.25 culture volumes of ice-cold 20 mM Tris-HCl pH 8.0, and centrifuged again at 8000 *g* for 5 min at 4°C. The supernatants were again removed, and the remaining cells stored as frozen pellets at -70°C.

To verify that the proteins had been produced in sufficient quantities, as well as ascertain whether the proteins separated into the soluble or insoluble cell extract fractions, crude extracts from the IPTG-induced and non-induced cultures were compared by SDS-PAGE (see 1.2.5).

3.2.6 Purification of recombinant proteins

Recombinant proteins were extracted and purified by metal chelation chromatography, using His·Bind Resin (Novagen), according to the manufacturer's instructions with some modifications.

(i) Protein extract preparation

In cases where the recombinant proteins were found to be soluble (rSRBPab1, rSRBPab2, and rSREPab), the cell pellets were resuspended in 4 mL of 1x Binding Buffer (diluted from 8x Binding Buffer, which contains 40 mM imidazole, 4 M NaCl, 160 mM Tris-HCl pH 7.9). Since the strain of *E. coli* used for expression carries pLysS, the cells lysed

completely on thawing. However, to reduce the viscosity of the solution due to the high molecular weight nucleic acids, the solution was sonicated for 15 s bursts using an Ultrasonic Processor Sonicator XL (Misonix, Inc.) at 40% power fitted with a microtip, until the samples were no longer viscous. The lysates were then centrifuged at 50,000 *g* (TLA100.3 rotor at 35,000 rpm) for 20 min at 4°C to remove cellular debris. Before loading onto the column, the supernatants were filtered through a 0.45 μm membrane (Sartorius).

For rSRDPab2, which was found to be insoluble, the cell pellet was resuspended in 40 mL of 1x Binding Buffer, sonicated briefly as above, then centrifuged at 5000 *g* for 15 min to collect the inclusion bodies and cellular debris. The supernatant was removed and the pellet resuspended in 20 mL of 1x Binding Buffer and centrifuged again at 5000 *g* for 15 min. After removing the supernatant, the pellet was resuspended in 5 mL of 1x Binding Buffer containing 6 M urea. To completely dissolve the protein, the solution was incubated on ice for 1 h. The remaining insoluble material was removed by centrifugation at 25,000 *g* (TLA100.3 rotor at 25,000 rpm) for 30 min at 4°C. Before loading onto the column, the supernatant was filtered through a 0.45 μm membrane (Sartorius).

(ii) Column chromatography

Chromatography of the prepared protein extracts for rSRBPab1, rSRBPab2, and rSREPab was performed in non-denaturing conditions, using a column format. The His·Bind Resin was gently mixed by inversion until completely suspended, then carefully added to a 1 cm diameter glass column until there was approximately 2.5 mL of settled resin. Once the level of storage buffer (20% ethanol) dropped below the top of the column bed, the column was washed with three column volumes of deionised water, followed by five volumes of 1x Charge Buffer (diluted from 8x Charge Buffer, which contains 400 mM $NiSO_4$) and then three volumes of 1x Binding Buffer. As the Binding Buffer drained to the top of the column bed, the prepared extract was loaded on. The flow-through from the column (unbound fraction) was retained for subsequent analysis. The column was then washed with ten volumes of 1x Binding Buffer, followed by six volumes of 1x Wash Buffer (diluted from 8x Wash Buffer, which contains 480 mM imidazole, 4 M NaCl, 160 mM Tris-HCl pH 7.9). Both of these wash fractions were also retained for subsequent analysis. The bound protein was eluted with six volumes of 1x Elute Buffer (diluted from 4x Elute Buffer, which contains 4 M imidazole, 2 M NaCl, 80 mM Tris-HCl pH 7.9). Finally any

remaining protein was removed by stripping the column of Ni^{2+}, with six volumes of 1x Strip Buffer (diluted from 4x Strip Buffer, which contains 400 mM EDTA, 2 M NaCl, 80 mM Tris-HCl pH 7.9). Once elution was complete, the resin was washed with five volumes of 20% ethanol and stored at 4°C.

For the purification of rSRDPab2, chromatography was performed under denaturing conditions. The procedure performed was identical to that described above, except that the 1x Binding, Wash and Elute Buffers were supplemented with 6 M urea, and the 1x Wash Buffer contained only 20 mM imidazole.

(iii) Sample processing after elution

To allow further study of the biochemical characteristics of the purified recombinant proteins, the buffers that the proteins had been eluted in were changed. Three different approaches were required.

rSRBPab1 and rSRE were dialysed (3,500 MWCO) against two changes of 4 L of a solution containing 100 mM NaCl, 50 mM Tris-HCl pH 8.0, allowing at least 3 h between changes. To remove the precipitate that formed during the dialysis, the dialysed solution was centrifuged at 10,000 *g* for 10 min. The solution was then concentrated to the desired volume using a Centricon-10 centrifugal concentrator (Amicon), and stored at -20°C.

rSRBPab2 was dialysed (3,500 MWCO) against two changes of 4 L of a solution containing 500 mM NaCl, 50 mM Tris-HCl pH 8.0, allowing at least 3 h between changes. The dialysed solution was then heated to 85°C for approximately 15 min to solubilise the protein. After allowing the solution to cool slowly to room temperature, it was centrifuged at 10,000 *g* for 10 min to remove the remaining insoluble material. The solution was then concentrated to the desired volume using a Centricon-10 centrifugal concentrator (Amicon), and stored at 4°C.

rSRDPab2 was renatured using a procedure based on that used for recombinant HIV-integrase (Wingfield and Palmer 1996). The rSRDPab2 solution was added drop-wise to 35 volumes of a folding buffer containing 50 mM Tris pH 7.5, 500 mM NaCl, 10 mM 3-[(3-Cholamidopropyl) dimethylammonio]-1-propanesulfonate (CHAPS; Sigma-Aldrich), and 2 mM DTT. The solution was then concentrated to the desired volume using a Centricon-10 centrifugal concentrator (Amicon), and stored at 4°C.

The quantity of protein yielded at the end of the purification processes was determined using a modification of the Bradford assay (Bradford 1976; Bollag and Edelstein 1991). The quality of the preparations was analysed by SDS-PAGE (see 1.2.5).

3.3 Results

3.3.1 Cloning of *srbPab1*, *srbPab2*, *srdPab2*, and *srePab*

The genes encoding SRBPab1, SRBPab2, SRDPab2 and SREPab were amplified from *Pyrococcus abyssi* genomic DNA then cloned into the pET-19b expression vector (Novagen). Genomic DNA was isolated from a late-log culture (OD_{600} of approximately 0.3) of *P. abyssi* strain Orsay. From 100 mg (wet-weight) of cells , approximately 0.5 μg of genomic DNA was isolated. The isolated DNA was free of major contaminants, as indicated by an A_{260}/A_{280} value of 1.6, and examination by agarose gel electrophoresis showed that the DNA was only minimally degraded (results not shown).

Using the complete genomic sequence data of *P. abyssi* from GenBank (NCBI), the DNA sequences of the four genes of interest, *srbPab1*, *srbPab2*, *srdPab2*, and *srePab*, were identified. These sequences were then used to design sets of forward and reverse primers that would specifically amplify the target regions by PCR (see 3.2.3). The forward and reverse primers also contained *Nde*I and *Bam*HI recognition sequences to allow for the subsequent directional cloning into the pET-19b vector. The resulting PCR products for each of the target regions, when analysed by agarose gel electrophoresis, yielded single bands, indicating that no major secondary products had been amplified. The sizes of each of the products, as estimated from their relative electrophoretic mobility compared to a 100 bp standard ladder, agreed with the expected sizes determined from the *P. abyssi* genome sequence (see 3.2.3).

To clone the amplified genes into the pET-19b vector, the PCR products and pET-19b vector were digested with *Nde*I and *Bam*HI, then ligated with DNA ligase. The resulting ligation products were transformed into *E. coli* DH5α by electroporation and cultured overnight on ampicillin-resistant selective media. Since the pET-19b vector does not contain any marker to discriminate between insert and non-insert containing transformants, a random selection of the resulting colonies were screened by PCR for the target gene using the primers initially used to amplify them. Colonies that screened positively for each of the target genes were then cultured and purified plasmid DNA was isolated from them.

To check that the vector contained an insert of the correct size and that no major rearrangements of the vector DNA had taken place, an aliquot of each of the vector DNA samples was used for restriction enzyme analysis. Restriction enzymes that produced a diagnostic fragmentation pattern were chosen for each of the vectors: *Rsa*I should cut pET19b-srbPab1 (Figure 3-2) to yield six fragments sized 2263, 1567, 1426, 680, 395, and 35 bp; *Hinc*II should cut pET19b-srbPab2 (Figure 3-3) to yield four fragments sized 3532, 1486, 1085 and 27 bp; *Rsa*I should cut pET19b-srdPab2 (Figure 3-4) to yield five fragments sized 2263, 1567, 1134, 680 and 328 bp; and *Rsa*I + *Xho*I should cut pET19b-srePab (Figure 3-5) to yield five fragments sized 2263, 1567, 1284, 680 and 199 bp. Each of the vectors was indeed found to produce the expected restriction fragmentation pattern, suggesting that no significant rearrangements had occurred within the plasmids and that the correct construct had been isolated (results not shown).

Finally, using automated sequencing, the cloning/expression of the four recombinant vectors were sequenced using the T7 forward and reverse primers. Comparison of the sequenced regions with the corresponding sequences from the *P. abyssi* genome indicated that no mutations had arisen in the process of cloning the genes (results not shown).

3.3.2 Expression of *srbPab1*, *srbPab2*, *srdPab2*, and *srePab*

To produce the four recombinant proteins, rSRBPab1 (Figure 3-2B), rSRBPab2 (Figure 3-3B), rSRDPab2 (Figure 3-4B), and rSREPab (Figure 3-5B), the corresponding vectors were individually transformed into *E. coli* BL21(DE3)pLysS cells. These four new strains were then separately cultured in liquid medium to an OD_{600} of approximately 0.6, before addition of IPTG to induce expression of the cloned genes. Non-induced cultures were also grown in parallel for each of the strains. To verify that the induction was successful and that a protein of the appropriate size was being produced in the induced cultures, whole cell extracts from the harvested cells of both the induced and non-induced cultures were isolated and analysed by SDS-PAGE. In each case, substantial production of a protein of the anticipated molecular weight was observed in the extracts from the induced cultures (Figure 3-2C, Figure 3-3C, Figure 3-4C, and Figure 3-5C). However, it was found for cell extracts isolated from cultures expressing *srdPab2*, that the induced protein could only be detected in the insoluble fraction. Consequently, rSRDPab2 was solubilised in a denaturant and purified in denaturing conditions.

3.3.3 Purification of rSRBPab1, rSRBPab2, rSRDPab2, and rSREPab

The recombinant proteins were purified from the cell lysates using metal chelation chromatography, and were then put into buffers appropriate for their further study.

Since the recombinant proteins were synthesised using the pET-19b expression system, they contained an N-terminal fusion that contains a string of ten histidine residues (His·Tag; see Figure 3-2B, Figure 3-3B, Figure 3-4B and Figure 3-5B). This distinctive region in the recombinant proteins was then exploited to separate them from the other proteins in the cell lysate. To achieve this, the filtered soluble protein extracts were loaded onto a column containing a resin with immobilised Ni^{2+} ions, to which the His•Tag-containing proteins specifically bind. The rest of the proteins, which do not bind to the column, were then washed from the column. Finally, the remaining bound proteins were removed from the column using a high salt buffer that contains imidazole (imidazole competes against the recombinant proteins for binding to the column). This procedure was successfully carried out to purify each of the recombinant proteins, rSRBPab1 (Figure 3-2D), rSRBPab2 (Figure 3-3D), rSRDPab2 (Figure 3-4D), and rSREPab (Figure 3-5D), to near homogeneity as judged by SDS-PAGE.

Since the chromatography procedure employed to purify the recombinant proteins resulted in them being in buffers not suitable for further experiments, schemes needed to be devised to put the proteins into more standard buffer conditions. For rSRBPab1 and rSREPab, which had been purified under non-denaturing conditions, this was relatively elementary and was achieved by a slow dialysis process followed by centrifugation to remove the resulting precipitate. Following dialysis, the purified rSRBPab1 and rSREPab solutions were each concentrated to approximately 1 mg/mL. The final yield of rSRBPab1 and rSREPab attained from 100 mL of induced culture attained, was approximately 2 and 3 mg, respectively.

Interestingly, for rSRBPab2, which had also been purified under non-denaturing conditions, the dialysis resulted in the complete precipitation of the protein. Attempts to dialyse the NaCl and imidazole in more gradual, progressive steps had the same result, as did attempts at changing only the imidazole concentration. Purely by serendipity, it was found that heating the precipitated rSRBPab2 to 85°C for approximately 15 minutes, almost completely resolubilised the protein. If the solution was then allowed to cool slowly to room temperature, the protein remained soluble. However, freezing or rapidly chilling

the solution, resulted in precipitation of the protein. A possible explanation for this behaviour is that the highly charged imidazole was somehow associating with the protein allowing it to remain in solution. Upon removal of the imidazole the protein may have folded incorrectly or aggregated, resulting in it becoming insoluble and precipitating. Heating the protein to a higher temperature, approximately equivalent to the growth temperature of *P. abyssi*, presumably allowed the protein to renature into its correct conformation. The final yield of rSRBPab2 obtained was approximately 1 mg, from 100 mL of induced culture.

Since rSRDPab2 was purified under denaturing conditions, the protein needed to be renatured. The conventional procedure recommended in the pET System Manual (Novagen), which involves a series of dialysis steps against decreasing urea concentrations, consistently resulted in precipitation of the protein. Ultimately, the successful strategy was found to be that adapted from the procedure used to renature recombinant HIV integrase (Wingfield and Palmer 1996). Like rSRDPab2, recombinant HIV integrase is insoluble when expressed in *E. coli* and is purified under denaturing conditions (Wingfield and Palmer 1996). One of the critical parameters in this protocol is the use of CHAPS buffer. CHAPS is a zwitterionic detergent that has been well characterised for its use in protein solubilisation (Perdew *et al.* 1983; Wingfield and Palmer 1996) and preventing aggregation of proteins at high concentrations (Anglister *et al.* 1993). The presence of CHAPS at the relatively low concentration used here (10 mM), is unlikely to significantly affect the activity of the protein (Perdew *et al.* 1983; Anglister *et al.* 1993). The renatured rSRDPab2 solution was concentrated to 1 mg/mL. The final yield of rSRDPab2 was approximately 2 mg from 100 mL of induced culture.

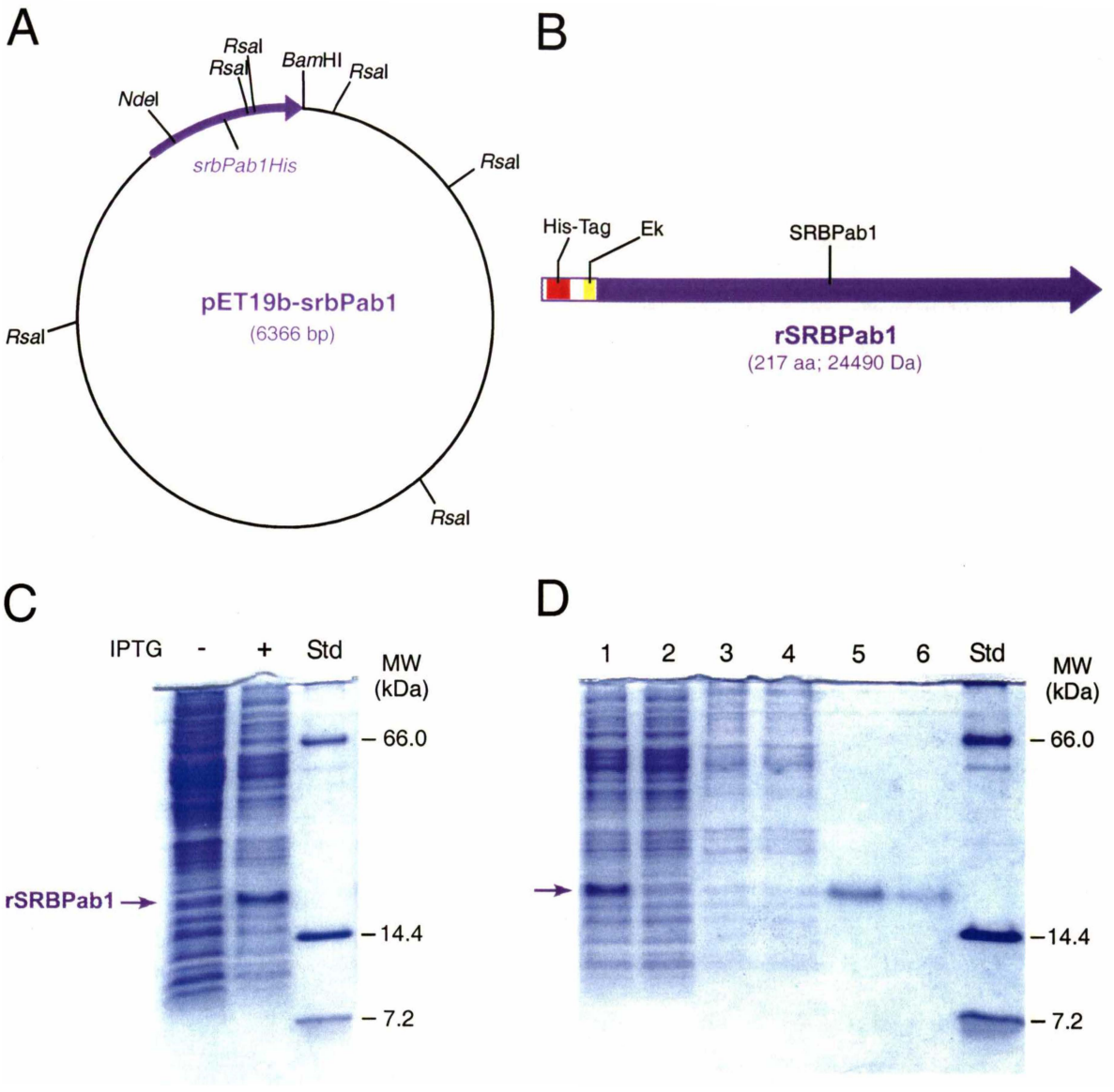


Figure 3-2. Production and purification of rSRBPab1. (A) Schematic of the pET19b-srbPab1 expression vector. The location of the gene encoding rSRBPab1 with His•Tag fusion is indicated (*srbPab1His*). The restriction sites used in the directional cloning of *srbPab1* (*Nde*I and *Bam*HI) and the diagnostic restriction sites used to verify the integrity of the plasmid are also shown (*Rsa*I). (B) Detail of the recombinant protein (rSRBPab1) encoded by *srbPab1His*. The relative sizes and positions of the His•Tag and enterokinase cleavage recognition region (Ek) are indicated. (C) Induction of rSRBPab1 expression in *E. coli* BL21(DE3)pLysS carrying the pET19b-srbPab1 plasmid. The total proteins from IPTG-induced cells (+) and non-induced cells (-) were isolated and resolved by SDS-PAGE and visualised by Coomassie staining. The position of rSRBPab1 is indicated. (D) Purification of rSRBPab1 using His•Tag resin in non-denaturing conditions. Lanes 1-6 were loaded as follows: (1) total soluble proteins isolated from induced *E. coli* BL21(DE3)pLysS carrying the pET19b-srbPab1 plasmid; (2) unbound fraction; (3) first wash fraction; (4) second wash fraction; (5) eluant; and (6) column strip fraction. Lane Std contains molecular weight (MW) markers. Samples were resolved by SDS-PAGE and visualised by Coomassie staining. The position of rSRBPab1 is indicated by the arrow.

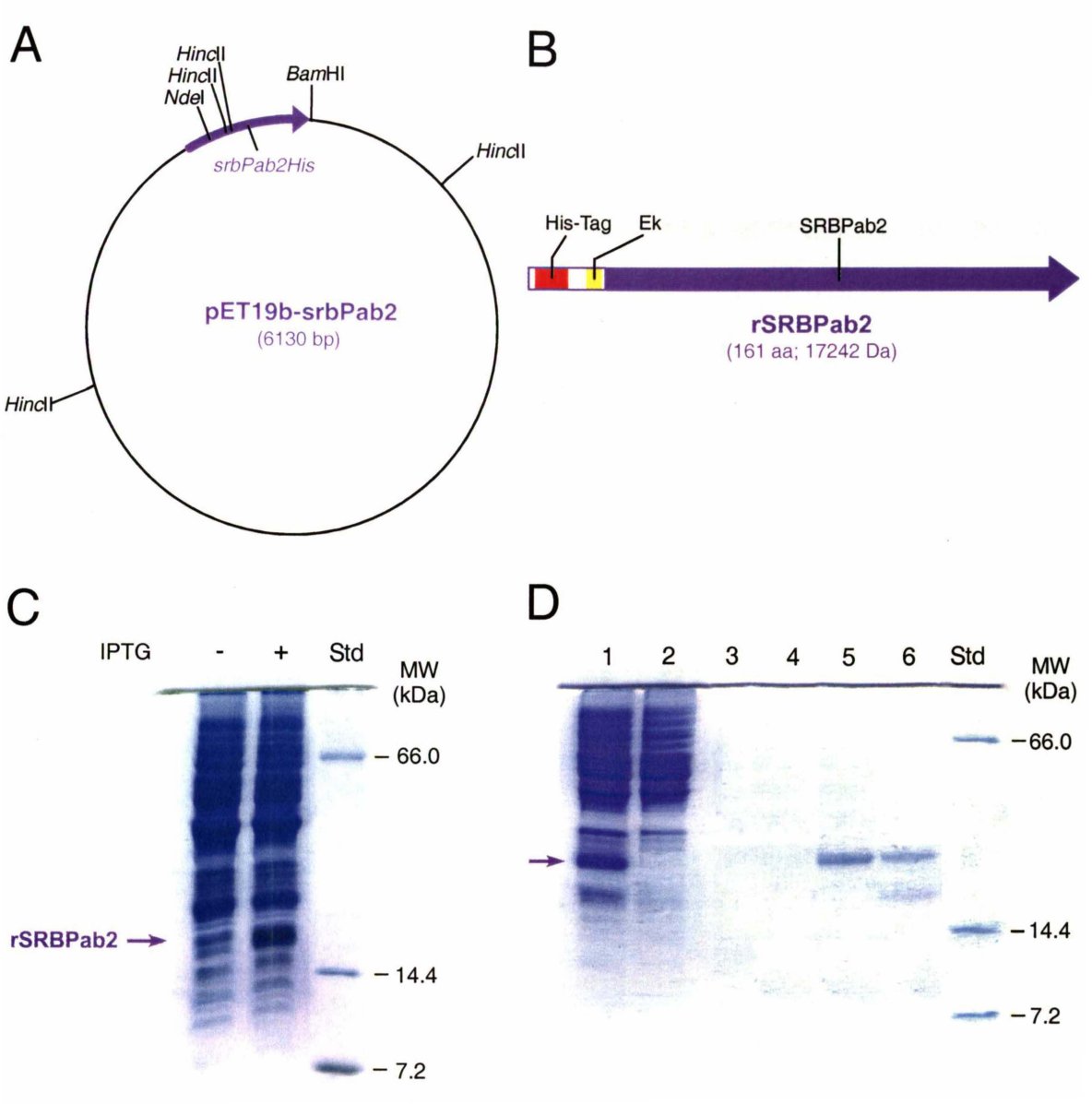


Figure 3-3. Production and purification of rSRBPab2. (A) Schematic of the pET19b-srbPab2 expression vector. The location of the gene encoding rSRBPab2 with His•Tag fusion is indicated (*srbPab2His*). The restriction sites used for the directional cloning of *srbPab2* (*Nde*I and *Bam*HI) and the diagnostic restriction sites used to verify the integrity of the plasmid are also shown (*Hinc*II). (B) Detail of the recombinant protein (rSRBPab2) encoded by *srbPab2His*. The relative sizes and positions of the His•Tag and enterokinase cleavage recognition region (Ek) are indicated. (C) Induction of rSRBPab2 expression in *E. coli* BL21(DE3)pLysS carrying the pET19b-srbPab2 plasmid. The total proteins from IPTG-induced cells (+) and non-induced cells (-) were isolated and resolved by SDS-PAGE and visualised by Coomassie staining. The position of rSRBPab2 is indicated. (D) Purification of rSRBPab2 using His•Tag resin in non-denaturing conditions. Lanes 1-6 were loaded as follows: (1) total soluble proteins isolated from induced *E. coli* BL21(DE3)pLysS carrying the pET19b-srbPab2 plasmid; (2) unbound fraction; (3) first wash fraction; (4) second wash fraction; (5) eluant; and (6) column strip fraction. Lane Std contains molecular weight (MW) markers. Samples were resolved by SDS-PAGE and visualised by Coomassie staining. The position of rSRBPab2 is indicated by the arrow.

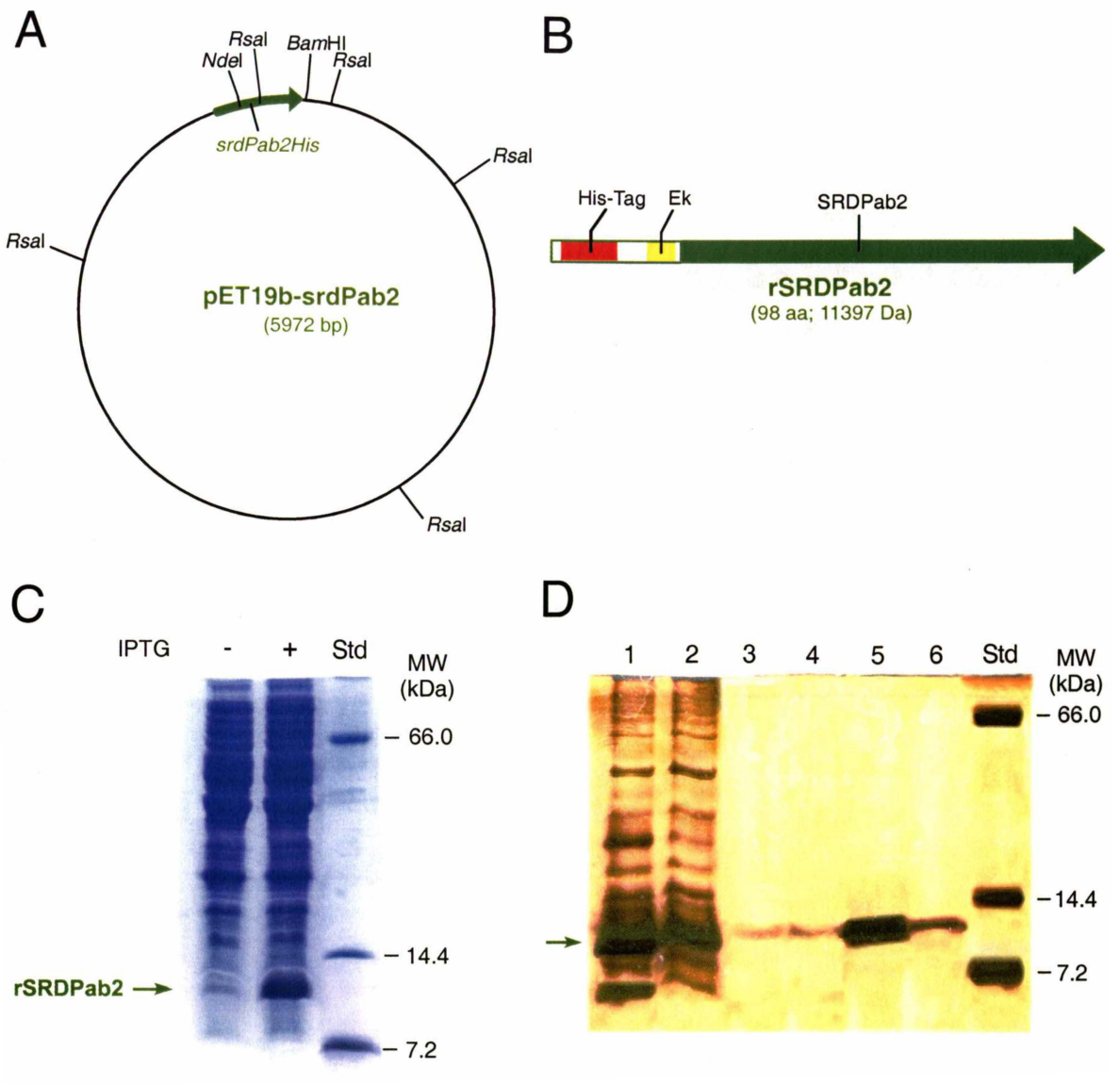


Figure 3-4. Production and purification of rSRDPab2. (A) Schematic of the pET19b-srdPab2 expression vector. The location of the gene encoding rSRDPab2 with His•Tag fusion is indicated (*srdPab2His*). The restriction sites used in the directional cloning of *srdPab2* (*Nde*I and *Bam*HI) and the diagnostic restriction sites used to verify the integrity of the plasmid are also shown (*Rsa*I). (B) Detail of the recombinant protein (rSRDPab2) encoded by *srdPab2His*. The relative sizes and positions of the His•Tag and enterokinase cleavage recognition region (Ek) are indicated. (C) Induction of rSRDPab2 expression in *E. coli* BL21(DE3)pLysS carrying the pET19b-srdPab2 plasmid. The total proteins from IPTG-induced cells (+) and non-induced cells (-) were isolated and resolved by SDS-PAGE and visualised by Coomassie staining. The position of rSRDPab2 is indicated. (D) Purification of rSRDPab2 using His•Tag resin in denaturing conditions. Lanes 1-6 were loaded as follows: (1) total proteins isolated from induced *E. coli* BL21(DE3)pLysS carrying the pET19b-srdPab2 plasmid; (2) unbound fraction; (3) first wash fraction; (4) second wash fraction; (5) eluant; and (6) column strip fraction. Lane Std contains molecular weight (MW) markers. Samples were resolved by SDS-PAGE and visualised by silver staining. The position of rSRDPab2 is indicated by the arrow.

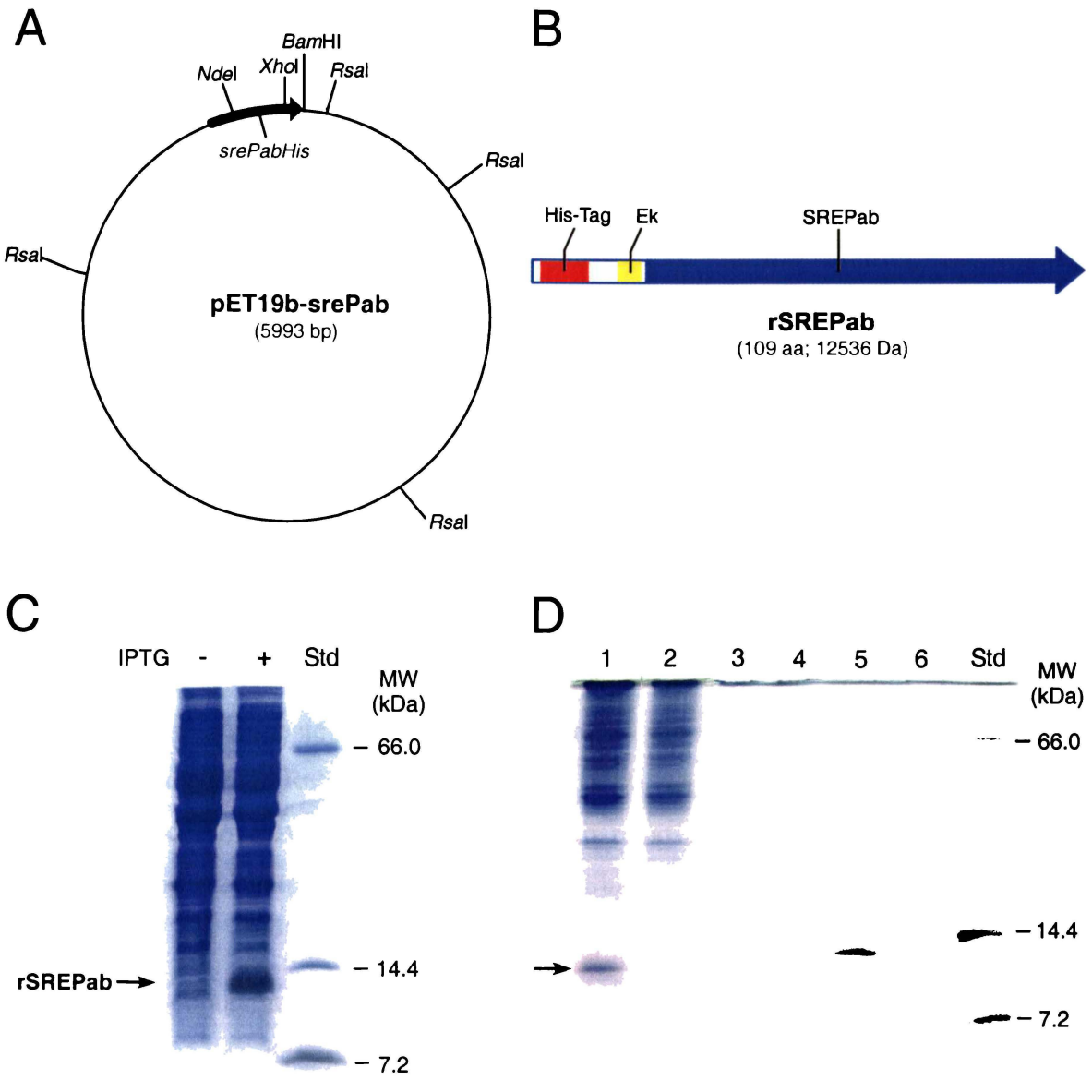


Figure 3-5. Production and purification of rSREPab. (A) Schematic of the pET19b-srePab expression vector. The location of the gene encoding rSREPab with His•Tag fusion is indicated (*srePabHis*). The restriction sites used in the directional cloning of *srePab* (*Nde*I and *Bam*HI) and the diagnostic restriction sites used to verify the integrity of the plasmid are also shown (*Rsa*I and *Xho*I). (B) Detail of the recombinant protein (rSREPab) encoded by *srePabHis*. The relative sizes and positions of the His•Tag and enterokinase cleavage recognition region (Ek) are indicated. (C) Induction of rSREPab expression in *E. coli* BL21(DE3)pLysS carrying the pET19b-srePab plasmid. The total proteins from IPTG-induced cells (+) and non-induced cells (-) were isolated and resolved by SDS-PAGE and visualised by Coomassie staining. The position of rSREPab is indicated. (D) Purification of rSREPab using His•Tag resin in non-denaturing conditions. Lanes 1-6 were loaded as follows: (1) total soluble proteins isolated from induced *E. coli* BL21(DE3)pLysS carrying the pET19b-srePab plasmid; (2) unbound fraction; (3) first wash fraction; (4) second wash fraction; (5) eluant; and (6) column strip fraction. Lane Std contains molecular weight (MW) markers. Samples were resolved by SDS-PAGE and visualised by Coomassie staining. The position of rSREPab is indicated by the arrow.

3.4 Discussion

In this chapter, the cloning and expression of four hitherto uncharacterised genes from *Pyrococcus abyssi*, *srbPab1*, *srbPab2*, *srdPab2*, and *srePab*, is described. These genes encode homologues of the stress-response proteins identified in Chapter One, i.e. SRBTzi, SRDTzi and SRETzi. Using the pET-19b expression system, the resulting recombinant proteins, rSRBPab1, rSRBPab2, rSRDPab2 and rSREPab, were readily purified by metal chelation chromatography.

It was found that a variety of approaches were required in order to successfully change the buffers that the different recombinant proteins were in, to something more appropriate for further biochemical studies. Careful dialysis proved effective to change the buffers of rSRBPab1 and rSREPab. However, rSRBPab2 precipitated when the imidazole was removed and could only be resolubilised again by heating. The use of heat to renature proteins has also been used with other proteins of thermophilic origins, such as D-glyceraldehyde-3-phosphate dehydrogenase from the hyperthermophilic bacterium *Thermotoga maritima* (Rehaber and Jaenicke 1992). Dialysis was also ineffective for rSRDPab2, and the only successful approach to remove the urea from the rSRDPab2 without causing precipitation, was to carefully dilute it into a buffer containing 10 mM CHAPS.

Difficulties in obtaining purified recombinant proteins, particularly those where the biophysical characteristics are essentially unknown, are relatively common and have been identified as a significant hurdle in structural proteomics (Christendat *et al.* 2000; Braun *et al.* 2002). Indeed, in a recent attempt to examine the structural proteome of *M. thermautotrophicus*, of the 424 genes predicted to encode non-membrane proteins that were targeted for cloning and expression, only 170 were successfully purified (Christendat *et al.* 2000). Although the purification strategies for different recombinant proteins varies, it is likely that the methods described here will at least form a useful basis from which to purify other recombinant proteins produced from members of the *srb*, *srd* and *sre* families.

The biochemical and biophysical studies that are now possible with large and readily renewable quantities of purified rSRBPab1, rSRBPab2, rSRDPab2 and rSREPab, should form a valuable foundation in furthering the understanding of the function of these proteins and their associated families. The purified protein samples provide the opportunity to analyse the activities and properties of these proteins in controlled conditions.

Furthermore, the purified protein samples are also likely to be useful for other future research applications, such as structural studies and the production of antibodies.

In the following chapter, a survey of some of the biochemical characteristics of rSRBPab1, rSRBPab2, rSRDPab2 and rSREPab is presented.

CHAPTER FOUR

Preliminary Examination of the Quaternary Associations and DNA-binding Capacity of rSRBPab1, rSRBPab2, rSRDPab2 and rSREPab

---

Summary

In the previous chapter, the production and purification of four novel proteins, rSRBPab1, rSRBPab2, rSRDPab2 and rSREPab, was described. The genes encoding these proteins were identified in the genome of *Pyrococcus abyssi*, and each belong to novel protein families that have, to date, not been functionally characterised. In order to gain a basic insight into the behaviour of these protein families, as well as establish some groundwork for future research, a preliminary examination of some biochemical characteristics of rSRBPab1, rSRBPab2, rSRDPab2 and rSREPab was performed.

To gather an understanding of the native state of the proteins, the potential for the formation of larger quaternary structures was investigated using glutaraldehyde crosslinking assays. The results suggested that rSRBPab1, rSRDPab2 and rSREPab all formed multimeric structures. rSRBPab2 did not form multimeric structures with itself or with rSRBPab1 in the conditions tested. The estimated molecular weight of the crosslinked product from the rSRBPab1 assay, which occurred only when the reaction was carried out at 80°C, was most consistent with that of rSRBPab1 dimers. The rSRDPab2 crosslinking assay also produced a dimer-sized product. Interestingly, the rSRDPab2 crosslinked product occurred preferentially at 80°C and when the reaction was performed in the presence of DNA. The rSREPab crosslinking reactions,

regardless of temperature or the presence of DNA, resulted in the formation of two high molecular weight products, the molecular weights of which were consistent with those predicted for rSREPab dimers and tetramers.

With the hypothesis in mind that these proteins may bind to DNA, three different DNA-binding assays were performed in a wide variety of conditions, with each of rSRBPab1, rSRDPab2 and rSREPab. The three DNA-binding assays were: (i) an agarose gel electrophoretic mobility shift assay using pUC19 DNA; (ii) a micrococcal nuclease protection assay using linearised pUC19 DNA; and (iii) a polyacrylamide gel electrophoretic mobility shift assay using DNA encompassing each of the promoters for the genes encoding SRBPab1, SRDPab2 and SREPab. Although the positive control protein, rHPa1, convincingly demonstrated DNA interaction in each of the assays, none of the recombinant proteins of interest revealed any indication of binding to the DNA in the conditions examined.

4.1 Introduction

In Chapter One, four distinct low-molecular weight stress-responsive proteins were identified from the total cytoplasmic protein complements of *Thermococcus zilligii*, and named SRBTzi, SRCTzi, SRDTzi and SRETzi. It was hypothesised that these low-molecular weight stress-response proteins were likely candidates as regulators of gene expression or as chromosomal organisers. Moreover, the profound decrease with increasing culture age in the levels of the archaeal histone, HTz, further supported the possibility for the presence of a replacement chromosomal organising protein. The levels of SRBTzi, SRDTzi and SRETzi had been observed to increase in level with increasing culture age, while the level of SRCTzi had been observed to decrease. By determining the N-terminal sequences of the stress-responsive proteins, homologues of the proteins were able to be identified in the genomes from members of the closely related genus, *Pyrococcus*. In Chapter Two, the conceptual protein sequences of the identified homologues from *Pyrococcus* were used to identify other homologous genes from the GenBank sequence database. It was found that SRBTzi, SRDTzi and SRETzi belonged to three different, essentially undescribed protein families, which were named SRB, SRD and SRE. SRCTzi was found to be a member of the Sac10b family of DNA-binding proteins, which has been somewhat characterised (see 1.1). In Chapter Three, the cloning of the genes encoding the *P. abyssi* homologues of SRBTzi, SRDTzi and SRETzi, namely

srbPab1, *srbPab2*, *srdPab2*, *srePab*, into an *E. coli* expression vector, was described. The subsequent production and purification procedure yielded purified preparations of the four proteins, which were named rSRBPab1, rSRBPab2, rSRDPab2, and rSREPab.

An understanding of the characteristics of rSRBPab1, rSRBPab2, rSRDPab2, and rSREPab *in vitro* will hopefully provide some understanding of the role of their respective protein families. Although there are numerous possible studies that can be performed using these purified proteins, two basic characteristics are of particular interest here. First, is to obtain some understanding of the quaternary structure that these proteins form in solution. The knowledge of the quaternary structure of a protein is not only an instrumental aspect in developing a model for how a protein functions and interacts with its environment, but is also an important consideration in the design of future experiments. Second, is to determine whether or not these proteins interact with DNA. Since one of the areas of interest for this thesis was to identify new regulatory or chromosomal organising proteins, testing the capacity of the recombinant proteins to bind DNA is pertinent. Moreover, as discussed in Chapter Two, the predicted structural data for each of the SRB, SRD and SRE families of proteins, indicated that associations with nucleic acids were reasonably conceivable: the structure of an SRB family member, SRBMth1, has been reported to resemble members of the RNase H superfamily, all of which can bind DNA or DNA/RNA hybrid duplexes (Cort *et al.* 2000); the structure of the C-terminal half of LrpA, which is homologous with the SRD proteins, resembles the DNA-binding domain of the bovine papillomavirus E2 protein (Hegde *et al.* 1992); the consensus predicted secondary structure of the SRE proteins was found to consist of four helical regions interspaced by coiled regions (see 2.3.16), which resembles the commonly observed helix-turn-helix DNA-binding motif (Chavali *et al.* 2001). In addition, SRBTzi, SRDTzi and SRETzi were found to be significantly enriched in the acid-wash extract of total cellular proteins, which is sometimes indicative of nucleoid-associated proteins (see 1.3.2).

In this chapter, an initial examination of the quaternary structures and DNA binding capacities of rSRBPab1, rSRBPab2, rSRDPab2, and rSREPab is presented. The results presented here are considered only preliminary, since a substantial research effort will be required to fully evaluate the characteristics under examination. Ideas for future research using rSRBPab1, rSRBPab2, rSRDPab2, and rSREPab are discussed.

Some of the results from this chapter were presented at the *Eleventh Annual Queenstown Molecular Biology* conference (Appendix F).

4.2 Materials and Methods

4.2.1 Glutaraldehyde crosslinking reactions

To investigate the possibility that the recombinant stress proteins formed quaternary structures, the proteins were subjected to glutaraldehyde crosslinking under several different conditions. Glutaraldehyde crosslinking assays were based on established methods (Silman *et al.* 1966; Grayling *et al.* 1995). To 0.2 mL tubes, 1 μg target protein, 5 μL reaction buffer (400 mM KCl, 80 mM Tris-HCl pH 7.8), 1 μL 0.2% (v/v) glutaraldehyde and deionised water (>18 MΩ resistance) to a final volume of 20 μL were added and mixed. To test the effect of DNA on quaternary structure formation, additional reactions were also performed in the presence of 500 ng of supercoiled pUC19 DNA (isolated by alkaline lysis as described (Sambrook *et al.* 1989)). In the case of reactions performed at high temperature, the reaction mixture was pre-equilibrated at 80°C for 10 min using a PTC-100™ Thermal Cycler (MJ Research, Inc.), before adding the glutaraldehyde. Reactions were allowed to proceed for 10 min at either room temperature (RT) or 80°C, as required. After the reactions were completed, 10 μL 2x gel loading buffer (8% (w/v) SDS, 24% (w/v) glycerol, 100 mM Tris-HCl pH 6.8, 0.02% (w/v) Coomassie Brilliant Blue G-250) was added to each tube. The samples were then incubated at 100°C for 10 min in a PTC-100™ Thermal Cycler (MJ Research, Inc.), centrifuged briefly, and loaded onto a 16.5% T/3% C polyacrylamide gel (prepared as described in 1.2.5). Finally, the samples were electrophoresed and stained with Coomassie Blue as described in 1.2.5.

4.2.2 Agarose gel electrophoretic mobility shift assays

To test the ability of the recombinant proteins to bind to plasmid DNA, agarose gel electrophoretic mobility shift assays (EMSAs) were conducted using modifications of established methods (Kerr 1995; Ronimus and Musgrave 1996). As the template DNA, a pUC19 plasmid DNA (2686 bp in size) preparation was used, that contained a mixture of nicked open circular molecules and supercoiled molecules. The pUC19 was isolated by alkaline lysis as described (Sambrook *et al.* 1989). As a positive control, a purified recombinant preparation of an archaeal histone from *P. abyssi*, rHPa1, was used. In 0.6 mL tubes, 100 ng pUC19 DNA, 200 ng or 400 ng recombinant protein, 5 μL 4x reaction buffer and deionised water to a final volume of 20 μL were added and mixed. The various 4x

reaction buffers contained 100 mM Tris-HCl pH 8.0, and either potassium glutamate at concentrations of 400 mM, 1 M, 1.5 M, 2 M, 3 M, or 4 M; or KCl at concentrations of 500 mM, 1 M, 2 M, or 3 M. The reactions were allowed to equilibrate at room temperature or 80°C for 15 min, before addition of 5 μL of 5x gel loading buffer (0.25% (w/v) bromophenol blue, 0.25% (w/v) xylene cyanol FF, and 15% (w/v) Ficoll). The tubes were centrifuged briefly, and the samples loaded onto agarose gels.

Agarose gel electrophoresis was conducted using a Horizon 11·14 gel electrophoresis system (BRL, Life Technologies, Inc.). The gels contained 2% (w/v) agarose (SeaKem LE, FMC Bioproducts) and 2x TBE buffer (0.178 Tris-HCl pH 8.3, 0.178 boric acid, and 10 mM EDTA), with wells formed using a 20 tooth comb (3.8 x 2 mm). To maximise any potential differences in the relative migration rates of the different topological states of the DNA examined, the gels were electrophoresed in the absence of ethidium bromide. The gels were electrophoresed in 2x TBE buffer at 1.25 V/cm for 15 h, using an ATTO Crosspower 1000 (Model AE-8400) power supply. Following electrophoresis, the gels were stained in a 0.5 μg/mL solution of ethidium bromide for approximately 30 min on a rocking platform, then destained in tap water for a further 45 min. The DNA was then visualised under UV light.

4.2.3 Micrococcal nuclease protection assays

To investigate the possibility that the recombinant proteins were able to protect regions of DNA from micrococcal nuclease digestion, micrococcal nuclease assays were performed. The assays were conducted according to a described method (Musgrave *et al.* 2000), with some modifications. As a positive control, a purified recombinant preparation of an archaeal histone from *P. abyssi*, rHPa1, was used. In 0.2 mL tubes, 700 ng *Eco*RI linearised pUC19 DNA, 700 ng recombinant protein, 5 μL 4x reaction buffer (80 mM Tris-HCl pH 8.0, 12 mM $CaCl_2$ and either 400 mM or 1.2 M potassium glutamate) and deionised water to a final volume of 20 uL, were added and mixed. The tubes were pre-incubated for 10 min at 70°C, before adding approximately 80 mU micrococcal nuclease (Sigma-Aldrich) in 0.1% (w/v) BSA (Sigma-Aldrich). Reactions were allowed to proceed for 10 min, before stopping them by addition of 6 μL 4x stop buffer (5% (w/v) SDS and 250 mM EDTA pH 8.0) and a further 10 min incubation at 70°C and 10 min on ice. In reactions performed at a final concentration of 300 mM potassium glutamate, a salt pellet

would form on cooling, which was removed by centrifuging the samples for 1 min at 15,000 *g*. Before loading the samples onto the gel, 6 μL of 5x gel loading buffer (0.25% (w/v) bromophenol blue, 0.25% (w/v) xylene cyanol FF, and 15% (w/v) Ficoll) was added. The results of the reactions were analysed by agarose gel electrophoresis. The samples were loaded onto 2% agarose gels in 1x TBE in the presence of 0.1 μg/mL ethidium bromide. The samples were electrophoresed at 4 V/cm for 4 h, then visualised under UV light.

4.2.4 Promoter binding assays

To examine the potential for the recombinant proteins to bind to promoter sequences, polyacrylamide gel EMSAs were performed. This section describes the preparation of the target DNA, the reaction conditions used for the assay, and the EMSA procedure used to analyse potential protein-DNA interactions.

(i) Preparation of promoter DNA regions

The promoter sequences used as targets for binding, were those belonging to the genes of the recombinant proteins being examined. To produce the target DNA for the assays, the regions encompassing the promoter for *srbPab1*, *srdPab2* and *srePab* were amplified by PCR. In a 0.2 mL tube, 40 μL PCR Master Mix (3 mM $MgCl_2$, 12 mM Tris-HCl pH 8.3, 60 mM KCl, 0.25 mM dNTPs), 5-10 ng genomic DNA, 1 μL each of forward primers and reverse primers (10 μM; Sigma-Aldrich), and 1 μL 1 U/μL AmpliTaq Gold DNA polymerase (Perkin Elmer) were mixed and the volume made to 50 μL with deionised water (>18 MΩ resistance). The forward and reverse primers that were used to amplify the promoter sequences, and the region that they span relative to the start codon, are shown in Table 4-1. PCR amplifications were performed using a PTC-100™ Thermal Cycler (MJ Research, Inc.), programmed with an initial enzyme activation period of 10 min at 94°C, followed by 35 cycles of the following steps: 94°C for 30 s, 53°C for 30 s, and 72°C for 45 s. The size of the PCR product was checked by comparing its relative electrophoretic mobility through a 2% agarose (SeaKem LE, FMC Bioproducts) gel in 1x TBE to that of a standard 100 bp ladder (Life Technologies). The DNA was purified and concentrated using the Concert Rapid PCR Purification System (Life Technologies), according to the manufacturer's instructions. The sequence of the PCR products was verified by sequencing

Primer name	Primer sequence (5' → 3')	Region spanned by amplicon
srbPab1 Promoter Fwd	GATTGCGCTTGTTAATCACT	-127 ... 0
srbPab1 Promoter Rev	TACCCATCACCTCAGAATTT	
srdPab2 Promoter Fwd	CTCTTATTAGACTGTCGG	-88 ... +5
srdPab2 Promoter Rev	GGTGATGTAAAAATGGTG	
srePab Promoter Fwd	TGTTCAAGGAATGCGAAATG	-99 ... +2
srePab Promoter Rev	CATTAGAATCACCTCCCAAT	

Table 4-1. Forward (Fwd) and reverse (Rev) primers used for the PCR amplification of the *srbPab1*, *srdPab2* and *srePab* promoter regions from *P. abyssi*. The regions that the amplicons span relative to the start codon of the respective genes, as inferred from the genome sequence data of *P. abyssi* (GenBank accession NC_000868), are indicated. The primers were designed with the aid of Genamics Expression 1.1 (Genamics), using the appropriate screening parameters to optimise their efficiency.

using their respective forward primers (Waikato DNA Sequencing Facility, The University of Waikato; results not shown).

(ii) Protein-DNA binding reactions

The promoter binding assays were performed between each combination of the *srbPab1*, *srdPab2*, and *srePab* promoter regions; the rSRBPab1, rSRDPab2 and rSREPab proteins; and three different buffer environments. As a positive control, a purified recombinant preparation of an archaeal histone from *P. abyssi*, rHPa1, was also assayed with each combination of promoter region and buffer environment. rHPa1 binds non-specifically to target DNA larger than 80 bp (Dinger, unpublished results).

The three different buffers used were selected to provide a diverse range of conditions to maximise the likelihood of identifying a favourable environment for potential DNA-protein interaction. Binding Buffer I, a relatively low salt buffer yielding final reactions concentrations of 12.5% (v/v) glycerol, 20 mM Tris-HCl pH 8.0, 0.4 mM EDTA pH 8.0, 0.1 mM DTT, 1 mM $MgCl_2$ and 50 mM NaCl, had been used in forming a protein-DNA complex between Lrp from the archaeon *Sulfolobus acidocaldarius* and the *sa-Lrp* control region (Enoru-Eta *et al.* 2000). Binding Buffer II, a relatively complex buffer yielding final reaction concentrations of 40 mM N-2-hydroxyethylpiperazine-N'-2-ethanesulfonic acid (HEPES; pH adjusted to 7.3 using NaOH), 200 mM KCl, 2.5 mM MgCl2, 2 mM DTT, 1 mM $CaCl_2$, 100 mM EDTA, and 10% (v/v) glycerol, had been used in forming a protein-DNA complex between an Lrp-like transcriptional regulator from the archaeon *Pyrococcus furiosus* and the *LrpA* promoter fragment (Brinkman *et al.* 2000). Binding Buffer III, a relatively high salt buffer resulting in final reaction concentrations of 300 mM potassium

glutamate, 50 mM Tris-HCl pH 7.8, and 10% (v/v) glycerol, was based on the conditions used to promote interactions between DNA and the archaeal histones from *Methanopyrus kandleri*, *Methanothermus fervidus* and *Thermococcus zilligii* (Musgrave *et al.* 2000).

The protein-DNA binding reactions were conducted using a procedure similar to those that the buffers described above were originally used in. To 0.2 mL tubes, 10 ng promoter DNA, 10 μL 3x Binding Buffer I, II or III, 1 μg recombinant protein, and deionised water to a final volume of 30 μL were added and mixed. The tubes were incubated at 80°C for 10 min to allow protein-DNA complexes to form. After incubation was complete, the samples were analysed by polyacrylamide gel electrophoresis.

(iii) Polyacrylamide gel electrophoresis

In order to resolve the relatively small molecular mass difference between bound and unbound DNA, the products of the protein-DNA binding assays were separated by continuous polyacrylamide gel electrophoresis using a procedure based on established methods (Carey 1991). Non-denaturing 0.75 mm thick 10% polyacrylamide gels were prepared by combining 5 mL 30% 29:1 acrylamide/bis-acrylamide (Bio-Rad), 3 mL 5x TBE, 7 mL deionised water (>18 MΩ resistance), 75 μL 10% APS (prepared fresh) and 7.5 μL TEMED. Gels were cast using a Joey™ Model JGC-2 Gel Casting System (Owl Separation Systems), and allowed to polymerise for at least 1 h before use. Before loading, 5 μL of 6x gel loading buffer was added to each sample and the tubes centrifuged briefly. Gels were electrophoresed at 100 V for 2 h.

To visualise the DNA, the gels were stained with silver essentially as described (Sambrook and Russell 2001). Following electrophoresis, gels were fixed in 100 mL 10% ethanol for 10 min. The ethanol was then removed by suction and replaced with 50 mL 0.7% (v/v) nitric acid and incubated with agitation for 5 min. The gels and the staining container were then rinsed twice in deionised water (> 18 MΩ resistance), before placing the gels in 100 mL 0.2% (w/v) $AgNO_3$ for 30 min with agitation. After the gels were rinsed a further three times with deionised water, they were developed in 100 mL of a solution containing 3% (w/v) $NaCO_3$ and 0.037% (v/v) formaldehyde. Once gels were developed to the desired extent, staining was stopped by replacing the developing solution with 100 mL 3% (v/v) acetic acid. After staining, gels were scanned (Microtek ScanMaker V300) and stored in hermetically sealed plastic bags.

4.3 Results

4.3.1 rSRBPab1 forms potential dimers at high temperature

To investigate the potential for rSRBPab1 to form quaternary structures, samples of rSRBPab1 were subjected to glutaraldehyde crosslinking reactions, then examined by SDS-PAGE (Figure 4-1). Since the native SRBPab1 protein originated from a hyperthermophilic organism, the crosslinking assay was performed at both room temperature and at 80°C. In addition, the effect of DNA on the formation of any quaternary structure was also examined.

The results indicate that at room temperature, in the presence or absence of DNA, no multimeric structures were formed. However, at high temperature, a small distinct population of high molecular weight molecules were formed. Although an accurate estimate of the size of the high molecular weight band could not be made from the gel, its relative electrophoretic mobility would be consistent with that of an rSRBPab1 dimer, which would have a theoretical molecular weight of approximately 49 kDa; any larger multimer would be larger than the 66 kDa MW marker. To confirm that this band represents rSRBPab1 dimers, a method to more accurately determine its molecular weight

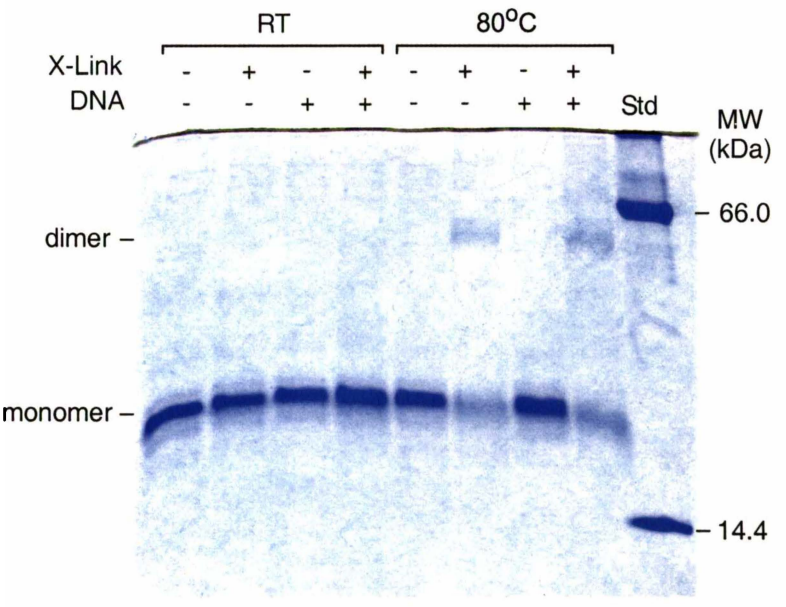


Figure 4-1. Glutaraldehyde crosslinking assay for rSRBPab1. rSRBPab1 was incubated at room temperature (RT) and high temperature (80°C) in the presence (+) and absence (-) of crosslinking agent (X-Link) and pUC19 DNA (added at a 1:2 mass ratio to the protein). Lane Std contains molecular weight (MW) markers of the sizes indicated. The samples were resolved by SDS-PAGE and Coomassie-stained. The positions of the rSRBPab1 monomer, and the putative dimer quaternary structures are indicated.

would be required, such as size-exclusion chromatography, mass spectrometry, or large format SDS-PAGE optimised for resolution in this molecular weight range with appropriate molecular weight markers.

4.3.2 rSRBPab2 does not form multimers with itself or with rSRBPab1

To examine the possibility of rSRBPab2 forming quaternary structures, glutaraldehyde crosslinking assays were performed in the presence and absence of DNA and at room temperature and 80°C (Figure 4-2). Under each of the conditions tested, no high molecular weight molecules were detected in the presence of the crosslinking agent.

Since most *srb*-containing genomes possessed at least two homologues of the gene, it was hypothesised that the different SRB homologues may associate in some way. To test this hypothesis, glutaraldehyde crosslinking assays were carried out as above, only in the presence of both rSRBPab1 and rSRBPab2. The results did not indicate the presence of any additional high molecular weight species (results not shown); the only multimers observed were those that were directly attributed to rSRBPab1 (see previous section).

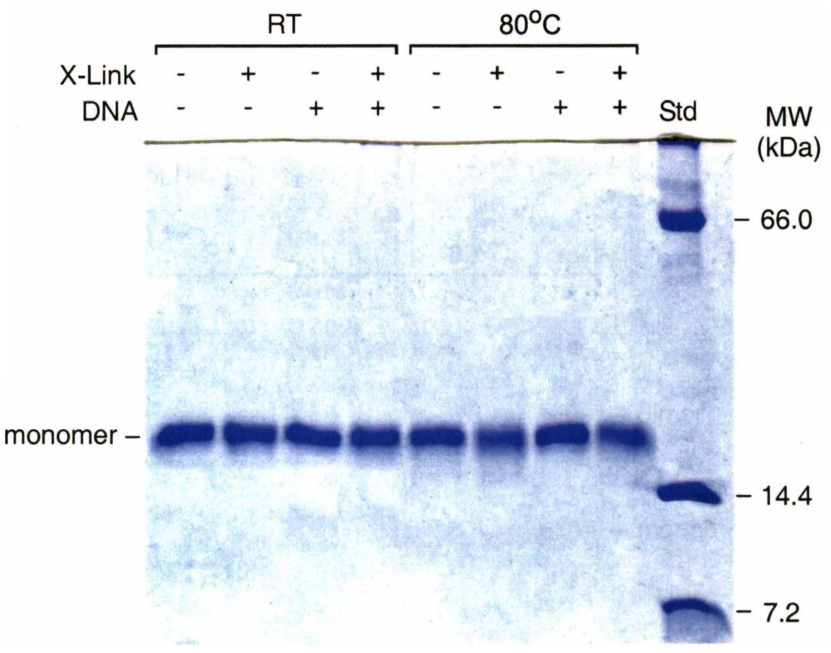


Figure 4-2. Glutaraldehyde crosslinking assay for rSRBPab2. rSRBPab2 was incubated at room temperature (RT) and high temperature (80°C) in the presence (+) and absence (-) of crosslinking agent (X-Link) and pUC19 DNA (added at a 1:2 mass ratio to the protein). Lane Std contains molecular weight (MW) markers of the sizes indicated. The samples were resolved by SDS-PAGE and Coomassie-stained. The position of the rSRBPab2 monomer units is indicated.

4.3.3 rSRDPab2 forms putative dimers preferentially at high temperature in the presence of DNA

To explore the prospect of rSRDPab2 forming higher order structures, samples of rSRDPab2 were subjected to glutaraldehyde crosslinking reactions in various conditions (Figure 4-3). Due to the sequence similarity between the SRD proteins and the LRP/AsnC family of transcriptional regulators (see 2.3.11), it seemed especially pertinent to examine the effect of DNA on the crosslinking assays. Indeed, the results indicated that a high molecular weight band was formed in the crosslinking reactions at both room temperature and 80°C when DNA was present. Moreover, this high molecular-weight product was formed in considerably increased quantities at high temperature. Although an accurate molecular weight determination of the high molecular weight product was not carried out, a separate gel with molecular weight markers (not shown) indicated that the product migrated at a molecular weight consistent to that of an rSRDPab2 dimer, which has a theoretical molecular mass of 22.8 kDa.

To ascertain whether the increased amount of product was purely a consequence of the faster reaction rates at high temperature, crosslinking assays were carried out for a prolonged duration at room temperature. The results indicated that the relative proportion of the dimer-sized product did not increase significantly with longer incubations (results not shown). Consequently, it is suggested that the increased frequency of the dimer-sized molecules at 80°C are a result of an increased stability of the quaternary association at the higher temperature.

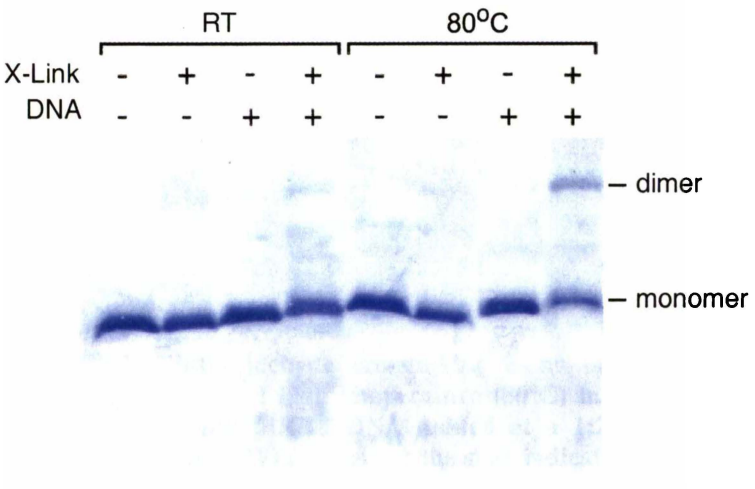


Figure 4-3. Glutaraldehyde crosslinking assay for rSRDPab2. rSRDPab2 was incubated at room temperature (RT) and high temperature (80°C) in the presence (+) and absence (-) of crosslinking agent (X-Link) and pUC19 DNA (added at a 1:2 mass ratio to the protein). The samples were resolved by SDS-PAGE and Coomassie-stained. The positions of the rSRDPab2 monomer and putative rSRDPab2 dimer and quaternary structures are indicated.

4.3.4 rSREPab forms putative dimers and tetramers

To examine the potential for rSREPab to form quaternary structures, glutaraldehyde crosslinking assays were performed at room temperature and 80°C and in the presence and absence of DNA (Figure 4-4). The results showed that two high molecular weight products were formed in the presence of glutaraldehyde in each of the conditions tested. Performing the assay at high temperature or in the presence of DNA did not have a discernible affect on the levels of the high molecular weight bands. The comparison of the two bands formed in the crosslinking assays with the molecular weight standard, indicated that their sizes were consistent with those of rSREPab dimers and tetramers, which have theoretical molecular masses of 25 and 50 kDa, respectively. To verify the suggestion that these high molecular weight bands do indeed represent dimers and tetramers of rSREPab, a more accurate determination of the molecular weights of these proteins would need to be performed, such as size-exclusion chromatography, mass spectrometry, or resolution by large format SDS-PAGE in conjunction with appropriately sized molecular weight standards.

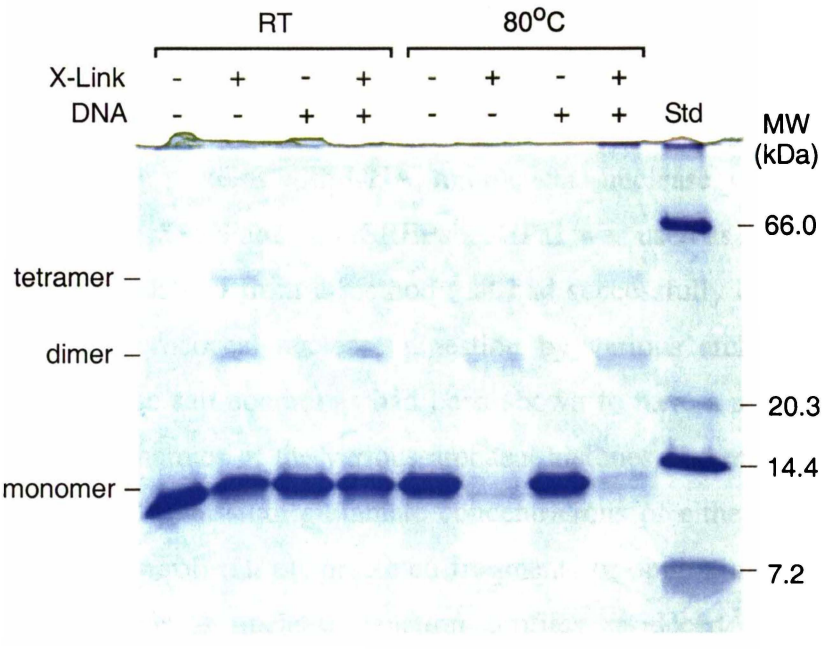


Figure 4-4. Glutaraldehyde crosslinking assay for rSREPab. rSREPab was incubated at room temperature (RT) and high temperature (80°C) in the presence (+) and absence (-) of crosslinking agent (X-Link) and pUC19 DNA (added at a 1:2 mass ratio to the protein). Lane Std contains molecular weight (MW) markers of the sizes indicated. The samples were resolved by SDS-PAGE and Coomassie-stained. The positions of the rSREPab monomer, and the putative dimer and tetramer quaternary structures are indicated.

4.3.5 rSRBPab1, rSRDPab2 and rSREPab do not affect the electrophoretic mobility of DNA

To ascertain the non-specific DNA-binding capacity of rSRBPab1, rSRDPab2 and rSREPab, electrophoretic mobility shift assays (EMSAs) were performed. rHPa1, an archaeal histone from *P. abyssi* known to bind DNA (Dinger, unpublished results), was used as a positive control. To maximise the opportunity for DNA-binding, the various recombinant proteins and DNA were equilibrated in a broad range of salt conditions, including KCl at concentrations ranging from 150 to 750 mM and potassium glutamate at concentrations ranging from 100 mM to 1 M. DNA-protein binding reactions were incubated at either room temperature or 80°C. The examination of the reaction products by agarose gel electrophoresis, indicated that the electrophoretic mobility of the DNA was not affected by either rSRBPab1, rSRDPab2 or rSREPab, in any of the reaction conditions tested (results not shown). As expected, the positive control, rHPa1, enhanced the mobility of the DNA molecules in all the reaction conditions, to varying extents.

4.3.6 rSRBPab1, rSRDPab2 and rSREPab do not protect DNA from micronuclease digestion

As an alternative approach to EMSAs to examine the potential for interaction of the recombinant proteins with DNA, micrococcal nuclease assays were performed for each of rSRBPab1, rSRDPab2 and rSREPab. rHPa1 was used as a positive control. The procedure used was adapted from a method that had successfully demonstrated protection of DNA against micrococcal nuclease digestion by various archaeal histones (Musgrave *et al.* 2000). Since salt conditions had been shown to have a significant influence on the DNA binding dynamics of the various archaeal histones, the micronuclease assays were carried out in final potassium glutamate concentrations of either 100 or 300 mM. Although the positive control, rHPa1, protected fragments of approximately 60 bp in size as expected, the micrococcal nuclease reaction profiles produced in the presence of rSRBPab1, rSRDPab2 or rSREPab could not be distinguished from the negative controls, which contained no protein (results not shown).

4.3.7 rSRBPab1, rSRDPab2 and rSREPab do not bind to their promoters

To investigate the possibility that rSRBPab1, rSRDPab2, and rSREPab bound specifically to promoter sequences, every combination of the recombinant proteins and their respective promoter regions were incubated in a variety of different buffer conditions, and the products analysed by polyacrylamide gel electrophoresis. The promoter DNA was produced by PCR from *P. abyssi* genomic DNA, using primers flanking an approximately 100 bp region upstream of the start codons of *srbPab1*, *srbPab2* and *srePab*. The three buffers chosen for equilibration of DNA-protein binding represent a diverse range of salt and buffer conditions, and were derived from established DNA-protein binding procedures (Brinkman *et al.* 2000; Enoru-Eta *et al.* 2000; Musgrave *et al.* 2000). The resulting polyacrylamide gels, which had been optimised to resolve subtle differences in the electrophoretic mobility of the target DNA, did not suggest the presence of DNA-protein interaction by any of the recombinant proteins with any of the target DNA tested (results not shown). However, the positive control, rHPa1, markedly retarded each of the promoter DNA samples, in all the buffers tested (results not shown).

4.4 Discussion

In this chapter, a preliminary examination of some of the biological characteristics of a member from each of the SRB, SRD and SRE families is reported. Using recombinant proteins produced in *E. coli*, which were encoded by the *srbPab1*, *srbPab2*, *srdPab2* and *srePab2* genes from *P. abyssi*, an investigation of the quaternary associations and DNA-binding capacities of the proteins was performed. It was found that each of the recombinant proteins, with the exception of rSRBPab2, formed higher order structures in the glutaraldehyde crosslinking assays. However, the various DNA-binding assays attempted with rSRBPab1, rSRDPab2 and rSREPab, did not indicate any evidence for DNA-protein interaction.

The rSRBPab1 crosslinking assays yielded a product that was consistent with the expected molecular weight of an rSRBPab1 dimer. Interestingly, however, this product was only formed when the reaction was carried out at 80°C. Since the gene encoding rSRBPab1 originated from *P. abyssi,* which has an optimal growth temperature of approximately 96°C (Erauso *et al.* 1993), the most likely explanation for this observation is that the high incubation temperature causes a conformational change or adjustment in the structure of

rSRBPab1, which in turn affects the dimerisation interface in such a way that the proteins are able to associate. Such temperature dependent structural changes have also been observed in other proteins originating from thermophilic organisms, such as glutamate dehydrogenase from *Sulfolobus solfataricus* (Facchiano *et al.* 1995), the archaeal histones from *Methanothermus fervidus* (Grayling *et al.* 1995), and propylamine transferase from *S. solfataricus* (Facchiano *et al.* 1992). Providing these structural changes are not too subtle, the hypothesis that temperature influences the structural conformation of rSRBPab2, could be supported by conducting further biophysical analyses of the protein, such as circular dichroism or infra-red spectroscopy. Using such methods, it should be possible to not only detect the putative structural transition of the protein with changing temperature, but also over what temperature range the transition occurs. The suggestion that the native conformation of rSRBPab1 is assumed only at high temperature, may cast some doubt on the validity of the solution structure for SRBMth1, which was performed at 25°C (Cort *et al.* 2000). Although the native growth temperature of *Methanothermobacter thermautotrophicus* is only 65°C (Zeikus and Wolfe 1972), it would nonetheless be pertinent to determine whether or not SRBMth1 maintains the same structure with changing temperature.

In contrast to the crosslinking assay results for rSRBPab1, rSRBPab2 did not reveal any indication of a crosslinked product. Furthermore, there was also no evidence for any crosslinking between rSRBPab1 and rSRBPab2. From a sequence perspective, the most distinguishing feature between rSRBPab1 and rSRBPab2 is that rSRBPab1 has a long C-terminal extension, which is rich in glycine and tryptophan residues (see Figure 2-3). However, it is difficult to envisage how such an extension could play any involvement in dimerisation. From a biochemical perspective, the purification procedures for rSRBPab1 and rSRBPab2 differed significantly; rSRBPab2 precipitated on removal of imidazole and required heating to 80°C to allow for resolubilisation of the protein. Given these differences, it may be pertinent to ascertain whether or not rSRBPab2 has indeed refolded into its native conformation. Using basic biophysical techniques, such as infra-red spectroscopy or circular dichroism, it should be possible to at least determine whether or not the ratios of secondary structural elements are in line with expectations, which would provide some insight into the folded state of rSRBPab2. It may also be worthwhile to examine whether or not different buffers affect the conformational dynamics of rSRBPab2, and if other conditions do allow the formation of multimeric structures.

The crosslinking assays for rSRDPab2 yielded a product that was consistent in size with that expected for an rSRDPab2 dimer. Interestingly, the product formed preferentially at 80°C and in the presence of DNA. As was suggested for rSRBPab1, the preferential formation of quaternary associations at high temperature may be accounted for by variations in the dimerisation interface that are influenced by temperature. As stated above, such changes should be measurable using biophysical approaches. More intriguing, however, was the observation of the substantial relative increase in the formation of the crosslinked product in the presence of DNA, compared to that without, at either room temperature or 80°C. A possible explanation for this observation is that rSRDPab2 associates with DNA, and that the consequent higher, more localised concentrations of rSRDPab2 result in a higher proportion of crosslinked molecules. Another possibility is that the dimerisation process occurs cooperatively in the presence of DNA. That is, the binding of the rSRDPab2 monomer to DNA somehow enhances the dimerisation process, either by modifying or stabilising the conformation of the protein. A combination of these two possibilities would also be feasible. A similar phenomenon is observed in the crosslinking of the archaeal histones, HMfA and HMfB, from *M. fervidus*, where crosslinking in the presence of DNA yields substantially greater proportions of dimers and tetramers (Grayling *et al.* 1995; Grayling *et al.* 1996). The authors suggested that at sufficient concentrations of DNA-bound histone dimers, dimer-dimer interactions become increasingly favourable, and this resulted in the observed increase in levels of the tetramer structures.

As reported in Chapter Two, the SRD proteins are homologous with the C-terminal half of the Lrp/AsnC family of transcriptional regulators (see 2.3.11). The structure of LrpA and other mutational analysis studies of Lrp from *E. coli*, have indicated that the N-terminal half of the Lrp/AsnC proteins, which contains a helix-turn-helix motif, is responsible for DNA binding (Platko and Calvo 1993; Leonard *et al.* 2001). However, the folding topology of the C-terminal half of LrpA also resembles other known DNA-binding domains (Leonard *et al.* 2001), and the unusually large recognition region of the Lrp proteins (Brinkman *et al.* 2000) also retains the possibility that the C-terminal half has some involvement with DNA association. Consequently, it would seem quite feasible that rSRDPab2 has the capacity to bind to DNA. In addition, consistent with the proposed dimer formation of rSRDPab2, the Lrp/AsnC proteins occur as dimers in solution

(Newman and Lin 1995) and the structure of LrpA indicates that the C-terminal half is responsible for a major proportion of the dimerisation interface (Leonard *et al.* 2001).

In light of the positive affect that the presence of DNA had on the proposed dimer formation of rSRDPab2, as well as its homology with the DNA-binding Lrp proteins, it was somewhat surprising that none of the DNA binding assays attempted suggested any evidence of DNA-protein association. However, it remains a possibility that any rSRDPab2-DNA interactions were too weak or not sufficiently stable to be detected by the assays performed. That is, they disassociated before being able to be analysed by gel electrophoresis. Indeed the preference for the suggested DNA binding by rSRDPab2 at high temperature makes assaying the protein-DNA interactions technically difficult using conventional gel electrophoresis methods. Two possible approaches that would circumvent this technical difficulty are: (i) further developing footprinting methods for DNA-protein complexes that can be conducted at high temperature; or (ii) attempting to vary the buffer the conditions to optimise DNA-protein associations at lower temperatures. Finally, it also remains a possibility that the N-terminal fusion created in cloning the *srdPab2* gene, has a destabilising effect on the resulting protein. This hypothesis could readily be tested by examining an enterokinase-treated sample of rSRDPab2, which would lack the N-terminal fusion (see 3.2.4).

The crosslinking assays for rSREPab indicated the presence of the formation of two different quaternary structures The estimated molecular weights for these structures were consistent with that predicted for rSREPab dimers and tetramers. Unlike rSRBPab1 and rSRDPab2, neither temperature nor the presence of DNA had any affect on the levels of the multimeric structures relative to that of the monomers. In light of the predicted consensus secondary structure for the SRE proteins, which consists essentially of four helices interspaced by coiled regions, the observation that it appears to form dimers and tetramers is not unexpected; helix-loop-helix structures are a relatively common feature in dimerisation motifs (Garrell and Campuzano 1991; Chavali *et al.* 2001). Since helix-loop-helix motifs can also commonly form heterodimers, it may be worthwhile to investigate if SREPab also associates with other proteins. With the aid of an anti-SREPab antibody, it may be possible to immunoprecipitate SREPab together with any associated proteins from a *P. abyssi* total protein preparation. In addition, the elucidation of the three dimensional structure of rSREPab may be especially significant, since it would be likely to represent a novel helix-loop-helix dimerisation motif. Finally, although rSREPab did not bind DNA in

the assays attempted here, it may be possible that the N-terminal fusion created in the production of the recombinant protein is somehow interfering with any potential binding. Consequently, it may be worthwhile attempting further DNA binding assays with an enterokinase-treated sample of rSREPab.

To conclude, the results presented in this chapter set the scene for further research into the characteristics of the rSRBPab1, rSRBPab2, rSRDPab2 and rSREPab proteins, and the families from which they originated. Clearly, there are a vast number of potential future research opportunities available for these proteins. Hopefully, the further advancement of this research will eventually lead to an understanding of these novel protein families and ultimately a better appreciation of the stress-response in microorganisms and perhaps also a further insight into the origins of gene regulatory systems for life on Earth.

Growth phase-dependent expression and degradation of histones in the thermophilic archaeon *Thermococcus zilligii*

Marcel E. Dinger, Gregory J. Baillie and David R. Musgrave*
Department of Biological Sciences, The University of Waikato, Private Bag 3105, Hamilton, New Zealand.

Summary

HTz is a member of the archaeal histone family. The archaeal histones have primary sequences and structural similarity to the eukaryal histone fold domain, and are thought to resemble the archetypal ancestor of the eukaryal nucleosome core histones. The effects of growth phase on the total soluble proteins from *Thermococcus zilligii*, isolated after various stages of growth from mid-logarithmic to late stationary phase, were examined by denaturing polyacrylamide gel electrophoresis. On entry into stationary phase, at least 11 proteins were detected that changed considerably in level. One of these proteins was identified by Western hybridization as HTz. The level of HTz decreased dramatically as cells entered stationary phase, and it could not be detected by late stationary phase. Unexpectedly, the Western hybridization detected a second protein, with an estimated molecular mass of approximately 14 kDa, which paralleled the decrease in level of HTz. Native purified HTz was shown to retain complete activity after prolonged incubation at the growth temperature of the organism, suggesting that the decrease in HTz was a specific cell-regulated process. Analysis of native purified HTz by electrospray ionization mass spectrometry revealed the molecular masses of HTz1 and HTz2 to be 7204 ± 3 Da and 7016 ± 3 Da respectively. The only non-covalent species that was detected corresponded to the molecular mass of an HTz1–HTz2 heterodimer. Northern analyses of *T. zilligii* total RNA with an *htz1* gene probe indicated a rapid decrease in expression of *htz1* with progression of the growth phase, and complete repression of *htz1* transcript synthesis by late logarithmic phase. Three proteins that changed in level with growth phase were identified by N-terminal sequence analysis. The first was homologous to a hypothetical protein conserved in all Archaea sequenced to date, the second to the Sac10b family of archaeal DNA-binding proteins and the third to the C-terminal region of the leucine-responsive regulatory family of DNA-binding proteins (LRPs).

Received 29 March, 1999; revised 21 February, 2000; accepted 29 February, 2000. *For correspondence. E-mail musgrave@waikato.ac.nz; Tel. (+64) 7 856 2889; Fax (+64) 7 838 4324.

Introduction

Histones with primary sequences similar to the eukaryal histone fold domain have been identified in several members of the Euryarchaeota (Sandman *et al.*, 1990; Tabassum *et al.*, 1992; Sandman *et al.*, 1994a; Ronimus and Musgrave, 1996a). The histones, HTz1 and HTz2, isolated from the sulphur-metabolizing thermophilic archaeon *Thermococcus zilligii* (Ronimus *et al.*, 1997) have been shown to compact DNA and protect it from thermal denaturation (Ronimus and Musgrave, 1996b). Archaeal histones have been shown to wrap DNA toroidally into structures that resemble those formed by the histone $(H3 + H4)_2$ tetramer at the centre of the eukaryal nucleosome and, consequently, are thought to resemble the archetypal ancestor of the eukaryal nucleosome core histones (Musgrave *et al.*, 1991; Pereira *et al.*, 1997). In *Methanothermus fervidus*, the archaeal histones, HMfA and HMfB, have been shown to constitute approximately 4% of the total soluble protein, which was suggested to be sufficient to package almost the entire genome of the organism (Pereira *et al.*, 1997). Furthermore, the observation that HMf can affect transcription *in vitro* and that HMf does not associate with certain sequences *in vivo* has supported the suggestion that the histones may influence gene expression (Pereira *et al.*, 1997; Soares *et al.*, 1998). The finding that the relative levels of HMfA and HMfB vary with the growth phase of the culture and that HMfA and HMfB homodimers have different DNA-binding properties also suggested the involvement of histones in the regulation of gene expression (Sandman *et al.*, 1994b).

In various members of the Bacteria, the onset of stationary phase is coupled with changes in protein synthesis and turnover (Kolter *et al.*, 1993). Many of the proteins regulated are involved in the response of the cell to stress conditions and survival in the stationary phase (Hengge-Aronis, 1996). The induction and repression of

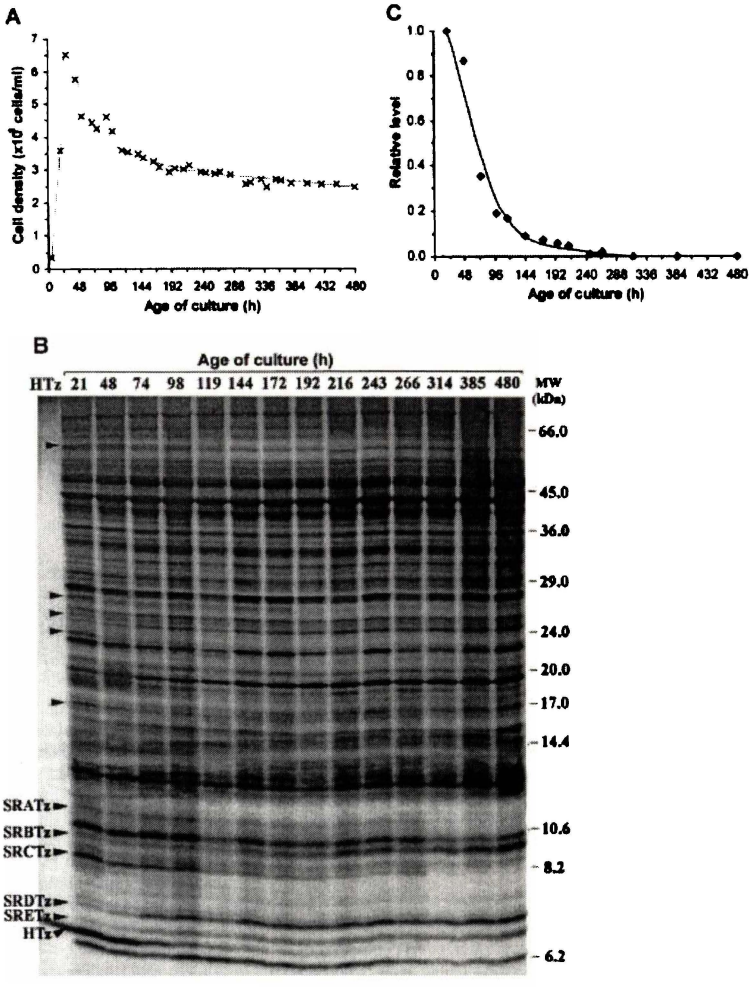


Fig. 1. Total cellular protein patterns of *Thermococcus zilligii* cultures after various periods of incubation from 1 to 20 days.
A. Cell density of a *T. zilligii* culture grown to late stationary phase in rich medium at 75°C. Cell densities were obtained by measuring the OD_{600} and converting the readings using conversion factors that had been determined by direct cell counts at various stages of the growth cycle. The unusual decrease in cell numbers after 48 h of growth is presumably a result of the stresses imposed by the limiting growth conditions, resulting in some cell lysis.
B. Total cellular protein samples from *T. zilligii* cells after each growth period resolved by Tris-tricine SDS–PAGE and stained with silver. Track 'HTz' contains 0.5 μg of purified native HTz. The remaining tracks each contain 10 μg of the total soluble proteins isolated after the incubation period indicated. Arrows indicate protein bands that change in level with growth phase.
C. Relative levels of the HTz band as a function of the age of the culture from which they were isolated.

transcription of certain genes with changing growth phase has been largely attributed to the effects of sigma factors and DNA- and RNA-binding proteins (Hengge-Aronis, 1996). In *Escherichia coli*, the abundant DNA-binding proteins IHF, FIS and H-NS all exhibit growth phase-dependent expression, and their involvement in both positive and negative regulatory effects on the production of a wide range of proteins has been demonstrated (Nyström, 1995; Osuna *et al.*, 1995; Laurent-Winter *et al.*, 1997). These DNA-binding proteins have been suggested to influence transcription via their effects on DNA topology (Pagel *et al.*, 1992; Tupper *et al.*, 1994; Travers and Muskhelishvili, 1998).

We describe here the growth phase-dependent expression of the *htz1* gene and degradation of the HTz1 and HTz2 proteins in *T. zilligii*. The results suggest that histones function primarily in the regulation of gene expression in the logarithmic phase and further support the suggestion that the archaeal histones have roles beyond that of DNA compaction and protection. We also report the identification of three other growth phase-dependent proteins that may be involved in the regulation of gene expression in *T. zilligii*.

Results

The level of HTz decreases with the onset of stationary phase

T. zilligii cell cultures were grown to various stages of their growth cycle, ranging from mid-log to late stationary phase (Fig. 1A). Total soluble proteins were extracted from the harvested cells, purified and resolved by Tris-tricine SDS–PAGE. As shown in Fig. 1B, although the overall protein patterns remained very similar over the period of the growth cycle examined, at least 11 proteins (indicated by arrows) can be identified that vary notably in level. Of particular interest is the putative HTz protein,

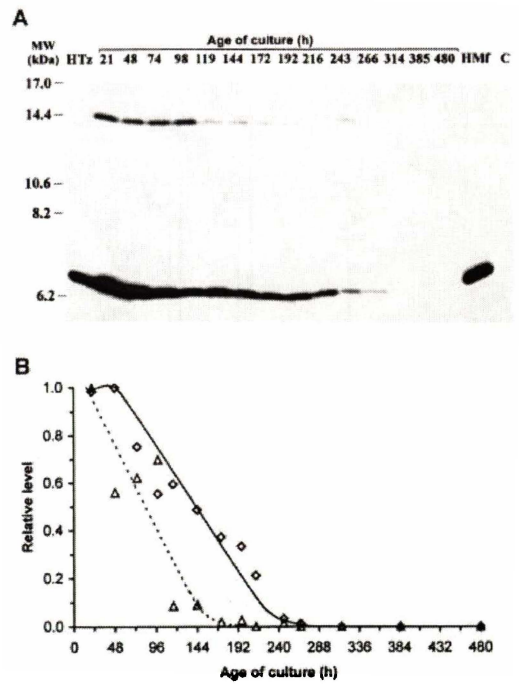


Fig. 2. Immunodetection of HTz.
A. Western blot of total cellular protein samples from *T. zilligii* cells after various periods of growth, resolved by SDS–PAGE using anti-HMf antibody. Tracks 'HMf' and 'HTz' contain 1 μg of purified native HTz and 100 ng of purified native HMf respectively. Track 'C' contains 50 μg of total soluble proteins isolated from *Thermotoga maritima* as a negative control. The remaining tracks each contain 50 μg of total soluble proteins isolated from *T. zilligii* after the incubation period indicated.
B. Relative levels of the two signals, HTz (◇) and 14 kDa species (△), detected in the total cellular protein samples in (A) as a function of the age of the culture from which they were isolated. The relative levels were normalized with respect to the logarithmic level (21 h) obtained for each protein. The curves were fitted to best represent the observed trends.

which occurs at a relatively high abundance in prestationary phase cells, then decreases rapidly in level as the culture ages (Fig. 1C).

The proposition that the protein corresponding in size to the HTz standard was indeed HTz was substantiated by Western blotting of the total soluble proteins, using a polyclonal antibody that had been raised against the archaeal histones, HMfA and HMfB (Fig. 2A). Previous investigations have shown that these antibodies bind with similarity affinity to HTz1 and HTz2 (D. R. Musgrave, unpublished results,). Figure 2B shows that the relative level of the HTz monomers drops off sharply after 2 days growth, at the onset of stationary phase, then decreases gradually until it is completely absent after approximately 11 days. Interestingly, a second signal, corresponding to a protein with a molecular mass of approximately 14 kDa, was also detected, which appeared to decrease in level parallel to the HTz signal (Fig. 2B). A protein of similar molecular mass, which decreased similarly in level, could be identified on an SDS–PAGE gel run with the acid-soluble fraction of the total soluble proteins (results not shown), suggesting that the protein may also be nucleoid associated. Although no further attempt has been made to extend this observation, the molecular mass and parallel degradation suggests that the band may correspond to non-denatured dimers of HTz1 and HTz2. As the native purified HTz standard that is run concurrently with the proteins does not reveal any additional band, the incomplete denaturation is suggested to result from the stabilizing effects of other proteins present in the sample. However, the possibility still remains that the 14 kDa species corresponds to a novel protein.

HTz is extremely stable at high temperature in vitro

To determine whether the observed decrease in level of HTz was a consequence of thermodegradation, the stability of the protein was examined *in vitro*. Native purified HTz was incubated at 90°C for 24, 48 or 72 h; then, its activity was determined by its ability to compact DNA, using electrophoretic mobility shift assays (EMSAs). No decrease in activity of any of the heat-treated HTz samples was observed relative to the non-heat-treated control (results not shown). A recent study on the stability of the recombinant archaeal histone, rHPyA1, has similarly shown the retention of full activity after 60 h incubation at 95°C (Soares *et al.*, 1998). The observation that the overall protein complement remains very similar throughout the growth period examined suggests that the decrease in level of HTz is a specific occurrence. Combined with the finding that HTz is extremely stable at high temperature, it is suggested that the decrease in level of HTz is associated with growth phase regulation in the stationary phase. Additionally, the increase in levels of other proteins (described above) provides further evidence of the continued regulation of protein turnover in stationary phase cells.

Native preparations of Archaeal histones, HTz1 and HTz2, form heterodimers in solution

To gain a better appreciation of the quaternary association and the relative abundance of HTz1 and HTz2, a native purified sample of the histones was analysed by electrospray ionization mass spectrometry (ESI-MS). The feasibility of ESI-MS for examining non-covalent association in proteins has been demonstrated in several investigations (Loo *et al.*, 1993; Light-Wahl *et al.*, 1994).

As shown in Fig. 3, three molecular species can be identified in the ESI-MS spectra of a purified solution of HTz1 and HTz2, with molecular masses of 7204 ± 3,

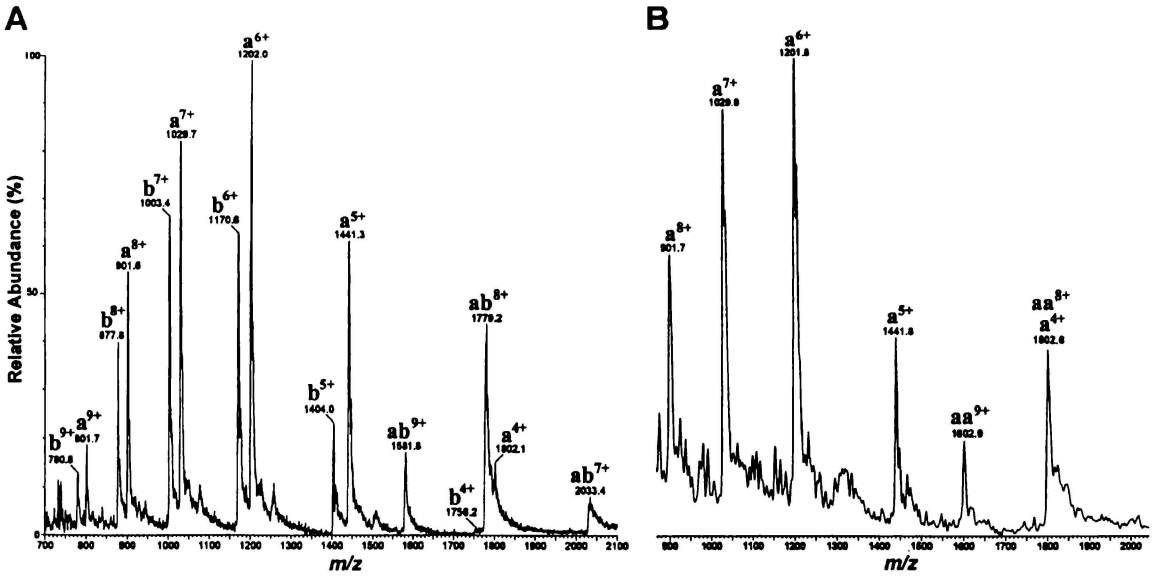


Fig. 3. Positive ion ESI-MS spectrum of purified native HTz (A) and rHTz1 (B). The peaks were assigned as follows: HTz1 monomer (a); HTz2 monomer (b); HTz1–HTz2 heterodimer (ab) and HTz1–HTz1 homodimer (aa). Samples were diluted in 1:1 (v/v) acetonitrile–water, and a 10 mM ammonium acetate solution in 5% (v/v) methanol was used as the mobile phase. The cone voltage was 40 V, and the inlet capillary temperature was 60°C.

7016 ± 3 and 14226 ± 5 Da. The molecular mass of the first molecule agrees with the theoretical molecular mass of HTz1, 7205 Da, as calculated from its translation sequence predicted from the nucleotide sequence of the *htz1* gene, excluding the N-terminal methionine residue (Ronimus and Musgrave, 1996b). This result indicates that the protein does not undergo any additional post-translational modifications. The absence of the N-terminal methionine residue of both subunits has been confirmed by sequencing of native purified HTz heterodimers (Ronimus and Musgrave, 1996b). The second molecule detected was predicted to correspond to HTz2, which migrates at approximately the same position as HTz1 in SDS–PAGE (Ronimus and Musgrave, 1996a), and would therefore be expected to have a similar molecular mass. The molecular mass of the third molecule coincides almost exactly with the combined mass of the HTz1 and HTz2 monomers (14220 ± 6), suggesting that this molecule represents an HTz1–HTz2 heterodimer. Consistent with this proposition is the narrow range of charge states and lower average charge per subunit observed in the proposed heterodimer species (7+, 8+ and 9+) compared with the monomer forms (4+, 5+, 6+, 7+, 8+, 9+), which has been suggested to be characteristic of quaternary association (Light-Wahl *et al.*, 1994). The partial disruption of non-covalent association is typical of the ionization process (Przybylski and Glocker, 1996) and, consequently, it is suggested that the monomeric species observed in the spectrum are caused predominantly by disruption of the heterodimer.

To examine the possibility that other quaternary associations of HTz1 and HTz2 were being completely disrupted in the ionization process, a variety of modifications to the ionization conditions were tested; however, no other multimeric forms were detected (results not shown). To demonstrate that the ionization conditions used could detect homodimers, a sample of purified recombinant HTz1 monomers was also analysed (Fig. 3B). The predicted molecular weight of the rHTz1 monomers agreed almost perfectly with their native counterparts, suggesting that the recombinant histones had not been modified. Although a clear series of peaks corresponding to a homodimer could not be identified, the distinct peak at 1603 would correspond to a +9 charged HTz1 homodimer with almost the exact calculated molecular weight of the homodimer, 14 418 Da. In addition, we suggest that the disproportionately large peak at 1802.6 is the result of a combination of both the homodimer (8+ charged) and the monomer (4+ charged). As no contaminants were detected in the sample by SDS–PAGE and the non-covalently bound molecules characteristically give only a narrow range of charge states (see above), we believe it is reasonable to conclude that these peaks demonstrate that HTz1 homodimers can be detected in the ionization conditions used.

The ESI-MS spectrum also reveals that the relative abundance of the two monomers, HTz1 and HTz2, in the native purified HTz extract is approximately 1.4:1. A similar inequality in the proportions of the archaeal

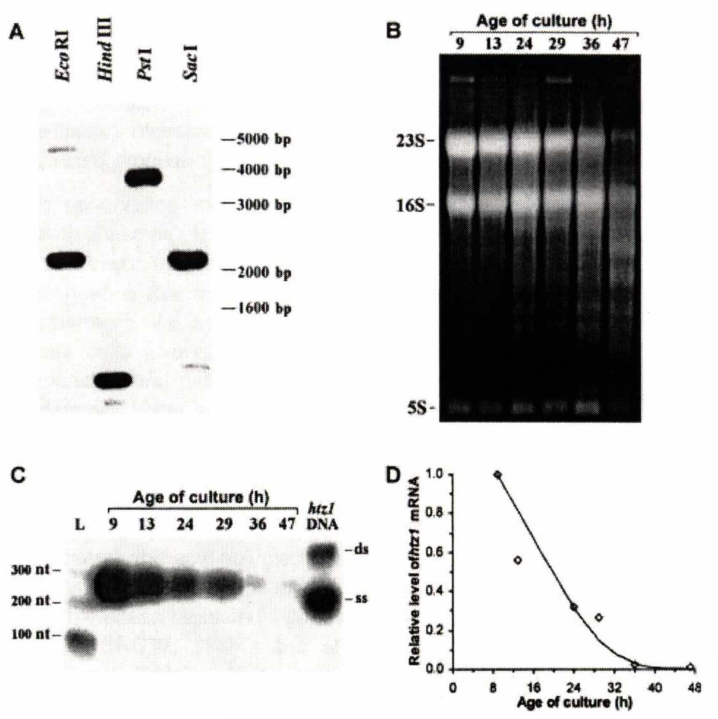


Fig. 4. Detection of *htz1* gene and transcript.
A. Southern hybridization of restriction endonuclease-digested *T. zilligii* genomic DNA with an *htz1* gene probe. Each track contains 5 μg of genomic DNA, treated by the restriction endonuclease indicated. The positions of the fragments of a 1 kb ladder are shown.
B. Total RNA samples extracted from *T. zilligii* cells at various early stages of growth resolved by formaldehyde agarose gel electrophoresis. Each lane contains 15 μg of total RNA isolated from cultures at the times indicated.
C. Northern hybridization, with an *htz1* gene. Track 'L' contains 1 μg of [γ-^{33}P]-5′ end-labelled 100 bp ladder, with sizes indicated. Track '*htz1* DNA' contains 500 pg of partially denatured *htz1* DNA as a positive control, with the single-(ss) and double-stranded (ds) bands indicated. The remaining tracks each contain 30 μg of total RNA isolated after the indicated incubation period.
D. Levels of *htz1* mRNA as a function of the age of the culture, relative to the early logarithmic level of *htz1* mRNA (9 h).

histones, HMfA and HMfB, has been reported in exponentially growing *M. fervidus* cells (Sandman *et al.*, 1994b). In addition, acid-urea PAGE analysis of purified native histones from *T. zilligii* revealed a discernible difference in the proportions of HTz1 and HTz2 (R. S. Ronimus and D. R. Musgrave, unpublished results).

Transcription of htz1 *occurs primarily during early log phase*

The temporal expression of the *htz1* gene was determined by Northern hybridization, using an *htz1* gene probe and total RNA isolated from *T. zilligii* cells after varying periods of growth.

To ascertain the extent of any cross-hybridization with any other potential *htz1*-like genes, Southern hybridization of *T. zilligii* genomic DNA was performed with an *htz1* gene probe (Fig. 4A). In each of the digested genomic DNA tracks, two signals were present. The predominant signal presumably corresponds to the *htz1* gene, while the second weaker signal is suggested to correspond to the *htz2* gene. This result suggests that the gene probe was almost entirely specific to *htz1* and therefore suitable for monitoring the level of *htz1* mRNA. The presence of only two histone genes has also been indicated in other members of the Thermococcales, such as *Pyrococcus abyssi* (P. Forterre, personal communication), *Pyrococcus horikoshii* (Kawarabayasi *et al.*, 1998) and *Pyrococcus* strain GB3a (Sandman *et al.*, 1994a).

Total RNA was isolated from *T. zilligii* cultures after various periods of growth from early log (9 h) to early stationary (2 days) phase, and resolved by formaldehyde agarose gel electrophoresis (Fig. 3B). The Northern hybridization revealed that the level of *htz1* transcription decreases rapidly as the culture ages, and diminishes almost entirely by the onset of stationary phase (Fig. 4C and D). *htz1* mRNA was not detected in the total RNA from cultures older than 2 days (results not shown). Clearly, the level of *htz1* transcription decreases more rapidly than the level of HTz protein. As the overall patterns of total RNA remain similar until early stationary phase and the lanes are loaded with equal quantities of RNA (Fig. 4B), we conclude that the rapid decrease in level of *htz1* transcript observed is the result of a specifically controlled process and not a representation of a general decrease in mRNA with culture age. Relative to the control DNA loaded, the maximal level of expression observed (at 9 h) corresponds to approximately 800 pg of *htz1* mRNA per 30 μg of total RNA, or 0.05% of the total mRNA (assuming 5% of the total RNA is mRNA). A densitometric analysis of the total soluble proteins separated by denaturing PAGE (Fig. 2A) reveals that HTz constitutes approximately 4% of the total protein in exponentially growing cells, a similar proportion to that estimated for HMf in exponentially growing *M. fervidus* cells (Pereira *et al.*, 1997). This apparent anomaly of the low quantity of *htz1* transcript compared with the level of protein observed might have implications for the efficiency

of translation of *htz1* mRNA and the turnover of HTz in exponentially growing cells.

Preliminary characterization of other growth phase-regulated proteins

The observation that the levels of HTz were indeed growth phase regulated suggested that the histones might be involved in more than just DNA compaction, and could also have a role in the regulation of gene expression. Furthermore, the absence of histones in the stationary phase cells also raises the question of how DNA is compacted and protected during stationary phase? To understand better what other regulatory factors might be involved in the transition to stationary phase, we attempted to identify some of the other growth phase-regulated proteins.

To enrich proteins that were likely to be nucleoid associated, the acid-soluble fraction of the total cellular proteins was prepared and analysed by SDS–PAGE. Five growth phase-regulated proteins, labelled SRATz, SRBTz, SRCTz, SRDTz and SRETz (Fig. 1), were all significantly enriched in the acid-soluble fraction (results not shown). Of these five, only SRETz, SRDTz and SRCTz were able to be concentrated and resolved sufficiently to allow their N-terminal sequences to be determined. SRETz was found to have homology with a family of highly conserved archaeal proteins. The N-terminal sequence of SRETz and its sequence analysis has been published elsewhere (Dinger and Musgrave, 1999). A BLAST search of the N-terminal sequence of SRDTz (VRAYVLLTIEIRKVESV) was performed against the NCBI protein sequence database. The results indicated that this protein shares significant identity with a group of hypothetical proteins identified in the complete genome sequences of *Pyrococcus horikoshii*, *Pyrococcus furiosus*, *Pyrococcus, abyssi*, *Archaeoglobus fulgidus*, *Streptomyces coelicolor* and *Aeropyrum pernix*. Interestingly, these proteins have sequence homology to the C-terminal half of the leucine-responsive regulatory proteins (LRPs) (Newman and Lin, 1995). The complete sequence analysis of SRDTz protein is to be published elsewhere. A BLAST search using the N-terminal sequence of SRCTz, AEEHVVYIGKKPVMNYVLAV, revealed that this protein belongs to the Sac10b family, a universally conserved DNA-binding protein in Archaea (Forterre *et al.*, 1999).

Discussion

In this study, we have examined the global protein patterns of *T. zilligii* and were able to detect that at least 11 proteins vary notably in level with changing growth phase. One of these proteins was identified as the archaeal histone, HTz, and the nature of this regulation was examined further.

The dramatic decrease in the level of HTz1 and HTz2 as the culture aged was in stark contrast to the majority of the other cellular proteins, which remained relatively constant in level. This implies that these histones may have a fundamental role in the chromosomal organization and, presumably, gene expression in the logarithmic phase. Considering the relatively high proportion of HTz in the cell and its role as a histone with the ability to bind and wrap DNA, the decline in HTz would be expected to have a considerable effect on the overall topology of genomic DNA. The superhelicity of DNA has been shown to vary with growth phase in various members of both Archaea and Bacteria (Balke and Gralla, 1987; López-García and Forterre, 1997). In *E. coli*, the growth phase-dependent DNA-binding protein H-NS can influence the degree of negative superhelicity and affect chromosome structure and, consequently, has been suggested as being involved in defining chromosomal domains and regulating the availability of genes for transcription (Owen-Hughes *et al.*, 1992; McGovern *et al.*, 1994). It is suggested that the regulation in the level of HTz would have considerable effects on gene expression, which may be brought about by binding DNA directly or through topologically induced changes.

The examination of the stability of HTz revealed that this protein is entirely resistant to thermodegradation for prolonged periods at its native growth temperature. This indicates that the observed decrease in level of HTz with changing growth phase is likely to be a consequence of a specific cell-mediated mechanism, which causes its breakdown. From the analysis of the total cellular protein profiles, the induced synthesis of several proteins was found to occur on entry into stationary phase. A stationary phase-induced protein would be a likely candidate to partake in the regulation of HTz levels, and would presumably function by either specifically degrading or destabilizing the histones.

Electrospray ionization mass spectrometry was used to determine accurately the molecular mass of the archaeal histones, HTz1 and HTz2, and to identify the quaternary structure formed by these subunits. The almost exact correspondence of the theoretical molecular mass of HTz1 from sequence data with the observed mass indicates that, aside from the removal of the N-terminal methionine residue, this histone does not undergo any further post-translational modification. The ESI-MS spectrum also revealed that the HTz subunits form heterodimers in solution. This observation is contrary to the quaternary structures formed by HMfA and HMfB, in which native preparations of HMf form a mixture of $(HMfA)_2$ and $(HMfB)_2$ homodimers and heterodimers (Sandman *et al.*, 1994b). It has been observed that the

ratio of HMfA to HMfB increases as the culture enters stationary phase and that the $(HMfA)_2$ and $(HMfB)_2$ homodimers have different DNA-binding properties *in vitro* (Sandman *et al.*, 1994b). Furthermore, it was suggested that the formation of different quaternary structures between the subunits might provide a mechanism for differential control of gene expression (Sandman *et al.*, 1994b). The data presented here indicate that a similar mechanism is unlikely to be in effect in *T. zilligii* and that the level of either subunit could limit the amount of heterodimers available for nucleosome assembly. Consequently, the cell would need to regulate specifically or fine tune the amount of only one histone to control the level of nucleosome formation. The level of nucleosome formation is critical for correct gene expression in eukaryotes; however, it is unknown whether this is also the case in Archaea. The inequality in the levels of HTz1 and HTz2 observed in the spectrum suggest that HTz2 is the limiting histone for heterodimer formation in *T. zilligii*.

To characterize further the nature of this growth phase-dependent degradation of HTz, the expression of the *htz1* gene was examined. The result indicated that the level of *htz1* mRNA was highest during early log phase, then decreased gradually throughout the log phase, such that it was at an undetectable level as the cells entered stationary phase. A tempting interpretation of this observation is that the synthesis of the histone is coupled to cell division or, more specifically, DNA replication. This would be consistent with histone synthesis in most Eukarya studied, in which the histone genes are transcribed primarily in the S-phase and have been shown to be dependent on the presence of replicating DNA (Osley, 1991). This pattern of synthesis is suggested to occur in order to preserve the correct ratio of DNA to histones, which is essential for the appropriate regulation of gene expression. Furthermore, a similar pattern of expression to that observed is also seen in several of the nucleoid proteins in some members of the Bacteria. For example, the DNA-binding proteins FIS and H-NS in *E. coli* and AbrB in *Bacillus subtilis* are expressed primarily in the early log phase, then reduce in level (Free and Dorman, 1995; Xu and Johnson, 1995; O'Reilly and Devine, 1997). Additionally, H-NS expression has been shown to couple directly with DNA synthesis (Free and Dorman, 1995).

The absence of histones in the stationary phase cells suggests a hypothesis that alternative DNA-binding proteins were replacing the histones. These replacement proteins may also be involved in adapting the levels and patterns of gene expression for the stress conditions encountered. This simple hypothesis could explain the observed changes in overall protein levels and the physiological and morphological changes that the cells undergo on entry into stationary phase (Klages and Morgan, 1994). The mass, abundance and likely nucleoid association suggested SRATz, SRBTz, SRCTz, SRDTz and SRETz as candidate growth phase-regulatory proteins. The N-terminal sequences of SRCTz and SRDTz indicated that these proteins were homologous to the Sac10b and LRP families of DNA-binding proteins respectively. Sac10b has been shown by electron microscopy to bind co-operatively to DNA in a RecA-like manner and was suggested to be involved in an unknown recombination/repair process (Dijk and Reinhardt, 1986). The observation that the homologue of Sac10b decreases in level in *T. zilligii*, as well as its appearance in every archaeal genome sequenced to date, leads us to suggest that this protein family may be a critical factor in the regulation of gene expression in these organisms. The LRPs are major transcriptional regulators found in both Archaea and Bacteria. In *E. coli*, LRPs control the expression of a large number of operons, can act as both repressers and activators and are suggested to be involved in chromosomal organization (Lin *et al.*, 1992; Wang *et al.*, 1994). Curiously, SRDTz corresponds to only the C-terminal half of the LRPs. As the tertiary structure of the LRPs is presently unknown, we do not know whether this region can bind DNA. In any case, we suggest that SRDTz, either by itself or in association with another protein, is involved in regulating gene expression in the stationary phase response of *T. zilligii*. Furthermore, the high levels of SRDTz accumulated in stationary phase cells also imply that this protein may have a role in chromosomal organization.

The results presented here indicate that Archaea are likely to possess complex control mechanisms to manage the physiological transitions required to adapt to different conditions. In view of the suggestion that members of the Archaea are earth's most primitive life forms, the study of these regulatory factors may help us to understand the evolution of the mechanisms involved in the control of gene expression. We are presently investigating the characteristics and interactions of the archaeal growth phase-dependent proteins further.

Experimental procedures

Source of native HTz

Native purified HTz was kindly provided by Ron Ronimus (Thermophile Research Unit, University of Waikato, NZ) and had been prepared as described previously (Ronimus and Musgrave, 1996a).

Preparation of rHTz1

The *htz1* gene was polymerase chain reaction (PCR) amplified from genomic *T. zilligii* DNA using the primers, 5′-aacgaagaattctctaaggaggtgttgggga-3′ and 5′-agaaggaagct-taaaacggtaacggtctg-3′. These primers contained *Eco*RI and

*Bam*HI recognition sites, respectively, at their 5′ ends. The double-digested *htz1* PCR product was ligated with pKK223-3 expression vector (Amersham-Pharmacia), which had been cut with the same enzymes. The ligation mix was transformed into *E. coli* JM109 (Amersham-Pharmacia). An isolate containing a plasmid of the correct size was identified, and the presence of the rHTz1 gene was confirmed by sequencing. Production and purification of rHTz1 was performed as described previously (Sandman *et al.*, 1994b).

Culturing of Thermococcus zilligii

T. zilligii was grown in 850 ml volumes in 1.0 l Schott bottles as described previously (Klages and Morgan, 1994). The OD_{600} of the samples was measured and converted to cell density by multiplying them by a conversion factor determined empirically by cell counting at several points throughout the growth cycle. Cell counting was performed according to previously described methods (Koch, 1994).

Preparation of total soluble protein samples from T. zilligii

To harvest the cells, the entire contents of the bottle were centrifuged, and the resulting cell pellet was washed in PBS. The cells were disrupted by sonication in 80 mM Tris-HCl, pH 7.6, 20 mM EDTA, 50 mM NaCl and 1 mM phenylmethylsulphonyl fluoride (PMSF), and the resulting lysate was centrifuged to pellet the cellular debris. The supernatant, containing the total soluble proteins, was dialysed (3500 Da MWCO) against 10 mM Tris-HCl, pH 7.6, 150 mM KCl and 7 mM β-mercaptoethanol.

Analysis of total soluble proteins

Proteins were resolved by Tris-tricine SDS–PAGE, as described elsewhere (Schägger and von Jagow, 1987), using the Bio-Rad Protean II Slab Gel electrophoresis system. Gels (0.75 mm thickness) were electrophoresed at 30 V for 1 h, then for approximately 15 h at 30 mA (constant current). The proteins were silver stained according to the method of Morrissey (1981). Gels were scanned at high resolution and analysed using Bio Image Intelligent Quantifier software (version 2.5; BI Systems Corporation).

Western blot procedure

After incubating gels for 30 min in a solution containing 25 mM Tris-HCl, pH 7.6, 192 mM glycine and 20% (v/v) methanol, proteins were electrophoretically transferred to NC-Extra 22 μm nitrocellulose membranes (Sartorius) using a semi-dry blotting apparatus at 200 mA (constant current) for 50 min. Membranes were incubated in 5% (w/v) non-fat milk powder in Tris-buffered saline NP-40 [TBS NP-40; 10 mM Tris-HCl, pH 7.6, 150 mM NaCl and 0.05% (w/v) NP-40] for 90 min, rinsed briefly in TBS NP-40, incubated in 1:100 rabbit anti-HMf (a generous gift from Diane Stroup, Department of Microbiology, Ohio State University, USA) in 1% (w/v) non-fat milk powder in TBS NP-40 for 90 min, washed in three changes of TBS NP-40 for 15 min, incubated in 1:8000 goat anti-rabbit IgG (conjugated to horseradish peroxidase) in 1% non-fat milk powder in TBS NP-40, then washed again in three changes of TBS NP-40 for 15 min. The second antibody was detected using enhanced chemiluminescence (ECL) reagents (Amersham) according to the manufacturer's instructions, with X-OMAT AR film. The film was scanned and analysed using Bio Image Intelligent Quantifier software (version 2.5; BI Systems Corporation).

Electrospray ionization mass spectrometry

Electrospray ionization mass spectrometry was performed on a VG Platform II instrument. Purified native HTz and rHTz1 were diluted in 1:1 (v/v) acetonitrile–water (final concentration approximately 200 ng ml^{-1}). Electrospray mass spectra were recorded in positive-ion mode, using MassLynx version 2.0 software. The sample was delivered to the mass spectrometer source by a SpectraSystem P1000 HPLC pump, at a flow rate of 0.2 ml min^{-1}, with 10 mM ammonium acetate buffer, adjusted to pH 7.0, in 5% (v/v) methanol as the mobile phase. The sample was introduced via a 10 μl sample loop fitted to a Rheodyne injector. Nitrogen was used as both the nebulizing and the drying gas, and the source temperature at the inlet capillary was 60°C. The cone voltage was adjusted to 40 V to optimize the extent to which the proteins were charged, while still maintaining the integrity of any non-covalent quaternary associations.

Isolation of DNA and RNA from T. zilligii

Genomic DNA was isolated from *T. zilligii* cells using the organic solvent extraction method described by Sambrook *et al.* (1989). Total RNA was isolated using TRIZOL LS Reagent (Gibco BRL, Life Technologies), according to the manufacturer's instructions. The concentration and quality of RNA was estimated by A_{260}, A_{260}/A_{280} and A_{260}/A_{230} measurement.

Southern and Northern blot procedures

The *htz1* probe was prepared from a PCR product (kindly provided by Ron Ronimus) corresponding to the coding sequence of the *htz1* gene by random primer labelling using the Rediprime DNA-labelling system (Amersham) according to the manufacturer's instructions.

For Southern blots, restriction endonuclease-digested genomic DNA was separated by electrophoresis through 1% (w/v) agarose gels, then denatured for 30 min in a solution containing 0.5 M NaOH and 1.5 M NaCl. DNA was transferred to Hybond+ nylon membranes (Amersham) by capillary blotting for 12 h in 10 × SSC. Membranes were prehybridized in a solution containing 500 mM Na_2HPO_4, pH 7.2, 1 mM EDTA, 7% (w/v) SDS at 42°C for 1 h, then hybridized in a replacement of the solution containing the *htz1* gene probe for a further 12 h. Membranes were washed three times in 1 × SSC with 0.1% SDS for 15 min at 25°C, then once in 1 × SSC with 0.1% SDS for 10 min at 62°C, before being dried and exposed to X-OMAT AR film (Kodak).

For Northern blots, total RNA samples were denatured by incubation for 15 min at 65°C in an equal volume of a solution

containing 20% formalin, 50% formamide, 0.1 M sodium 3-(N-morpholino)-propanesulphonic acid (MOPS; Sigma), then separated by electrophoresis through 2% (w/v) agarose gels run at 4.2 V cm^{-1}, containing 2.2 M formaldehyde, 20 mM MOPS and 8 mM sodium acetate. After washing the gels for 30 min in 10 × SSC, the RNA was transferred to Hybond+ nylon membranes by capillary blotting for 12 h in 10 × SSC. Nucleic acids were UV cross-linked to the membrane by irradiating the membrane for 45 s with an UV transilluminator. Membrane hybridization was performed at 60°C, using the procedure described for the Southern blots. Membranes were washed three times in 3 × SSC with 0.1% SDS for 10 min at 25°C, before drying and exposure to X-OMAT AR film. The intensity of the bands was quantified by analysis of a scanned image of the film using Bio Image Intelligent Quantifier software (version 2.5; BI Systems Corporation).

Characterization of other growth phase response proteins

Acid-soluble proteins from *T. zilligii* were prepared as described previously (Ronimus and Musgrave, 1996a). The acid-soluble proteins from a 4-day-old culture were concentrated, resolved by SDS–PAGE and electroblotted onto polyvinylidene difluoride (PVDF) membrane (Schägger and von Jagow, 1987; Ursitti *et al.*, 1996). The proteins on the membrane were visualized by staining with amido black (Harper and Speicher, 1996), and the bands corresponding to SRC, SRD and SRETz were excised. Automated N-terminal sequencing was performed by the University of Auckland Protein Sequencing Facility (New Zealand).

Acknowledgements

We thank the Department of Chemistry for use of the electrospray ionization mass spectrometer, and Dr Bill Henderson, Dr Maarten Dinger and Cameron Evans for helpful technical advice and assistance. We thank Dr Ron Ronimus for the purified HTz sample, and Patrick Forterre for unpublished information on HPa sequences.

References

Balke, V.L., and Gralla, J.D. (1987) Changes in the linking number of supercoiled DNA accompany growth transitions. *J Bacteriol* **169**: 4499–4506.

Dijk, J., and Reinhardt, R. (1986) The structure of DNA-binding proteins from Eu- and Archaebacteria. In *Bacterial Chromatin.* Gualerzi. C.O.. and Pon, C.L. (eds). Berlin: Springer Verlag, pp. 185–218.

Dinger, M., and Musgrave, D. (1999) Identification of archaeal genes encoding a novel stationary phase-response protein. *Biochem Biophys Acta* **1490**: 115–120.

Forterre, P., Confalonieri, F., and Knapp, S. (1999) Identification of the gene encoding archeal-specific DNA-binding proteins of the Sac10b family. *Mol Microbiol* **32**: 669–670.

Free, A., and Dorman, C.J. (1995) Coupling of *Escherichia coli hns* mRNA levels to DNA synthesis by autoregulation: implications for growth phase control. *Mol Microbiol* **18**: 101–113.

Harper, S., and Speicher, D. (1996) Detection of proteins on blot membranes. In *Current Protocols in Protein Science.* Coligan, J.E., Dunn, B.M., Ploegh, H.L., Speicher, D.W., and Wingfield, P.T. (eds). New York: John Wiley & Sons, pp. 10.8.1–10.8.7.

Hengge-Aronis, R. (1996) Regulation of gene expression during entry into stationary phase. In *Escherichia coli and Salmonella: Cellular and Molecular Biology.* Neidhardt, F.C., Curtiss, R., Ingrahmam, J.L., Lin, E.C.C., Low, K.B., Magasanik, B., et al. (eds). Washington, DC: American Society for Microbiology Press, pp. 1497–1512.

Kawarabayasi, Y., Sawada, M., Horikawa, H., Haikawa, Y., Hino, Y., Yamamoto, S., et al. (1998) Complete sequence and gene organization of the genome of a hyper-thermophilic archaebacterium, *Pyrococcus horikoshii* OT3. *DNA Res* **5**: 55–76.

Klages, K.U., and Morgan, H.W. (1994) Characterization of an extremely thermophilic sulphur-metabolizing archaebacterium belonging to the Thermococcales. *Arch Microbiol* **162**: 261–266.

Koch, A.L. (1994) Growth measurement. In *Methods for General and Molecular Bacteriology.* Gerhardt, P., Murray, R.G.E., Woord, W.A., and Krieg, N.R. (eds). Washington, DC: American Society for Microbiology Press, pp. 251–262.

Kolter, R., Siegele, D.A., and Tormo, A. (1993) The stationary phase of the bacterial life cycle. *Annu Rev Microbiol* **47**: 855–874.

Laurent-Winter, C., Ngo, S., Dachin, A., and Bertin, P. (1997) Role of *Escherichia coli* histone-like nucleoid-structuring protein in bacterial metabolism and stress response: identification of targets by two-dimensional electrophoresis. *Eur J Biochem* **244**: 767–773.

Light-Wahl, K.J., Schwartz, B.L., and Smith, R.D. (1994) Observation of the noncovalent quaternary associations of proteins by electrospray ionization mass spectrometry. *J Am Chem Soc* **116**: 5271–5278.

Lin, R., D'Ari, R., and Newman, E. (1992) Lambda placMu insertions in genes of the leucine regulon: extension of the regulon to genes not regulated by leucine. *J Bacteriol* **174**: 1948–1955.

Loo, O.R.R., Goodlett, D.R., Smith, R.D., and Loo, J.A. (1993) Observation of a noncovalent ribonuclease S protein/S-peptide complex by electrospray ionization mass spectrometry. *J Am Chem Soc* **115**: 4391–4392.

López-García, P., and Forterre, P. (1997) DNA topology in hyperthermophilic archaea: reference states and their variation with growth phase, growth temperature, and temperature stresses. *Mol Microbiol* **23**: 1267–1279.

McGovern, V., Higgins, N.P., Chiz, R.S., and Jaworski, A. (1994) H-NS over-expression induces an artificial stationary phase by silencing global transcription. *Biochimie* **76**: 1019–1029.

Morrissey, J.H. (1981) Silver stain for proteins in polyacrylamide gels: A modified procedure with enhanced uniform sensitivity. *Anal Biochem* **117**: 307–310.

Musgrave, D.R., Sandman, K.M., and Reeve, J.N. (1991) DNA binding by the archaeal histone HMf results in positive supercoiling. *Proc Natl Acad Sci USA* **88**: 10397–10401.

Newman, E., and Lin, R. (1995) Leucine-responsive regulatory protein: a global regulator of gene expression in *E. coli. Annu Rev Microbiol* **49**: 747–775.

Nyström, T. (1995) Glucose starvation stimulon of *Escherichia coli*: role of integration host factor in starvation survival and growth phase-dependent protein synthesis. *J Bacteriol* **177**: 5707–5710.

O'Reilly, M., and Devine, K.M. (1997) Expression of AbrB, a transition state regulator from *Bacillus subtilis*, is growth phase dependent in a manner resembling that of Fis, the nucleoid binding protein from *Escherichia coli. J Bacteriol* **179**: 522–529.

Osley, M.A. (1991) The regulation of histone synthesis in the cell cycle. *Annu Rev Biochem* **60**: 827–861.

Osuna, R., Lienau, K., Hughes, T., and Johnson, R.C. (1995) Sequence, regulation, and functions of *fis* in *Salmonella typhimurium*. *J Bacteriol* **177**: 2021–2032.

Owen-Hughes, T.A., Pavitt, G.D., Santos, D.S., Sidebotham, J.D., Hulton, J.C.D., Hinton, D., et al. (1992) The chromatin-associated protein H-NS interacts with curved DNA to influence DNA topology and gene expression. *Cell* **71**: 255–265.

Pagel, J., Winkelman, J., Adams, C., and Hatfield, G. (1992) DNA topology-mediated regulation of transcription initiation from the tandem promoters of the ilv GMEDA operon of *Escherichia coli*. *J Mol Biol* **4**: 919–935.

Pereira, S.L., Grayling, R.A., Lurz, R., and Reeve, J.N. (1997) Archaeal nucleosomes. *Proc Natl Acad Sci USA* **94**: 12633–12637.

Przybylski, M., and Glocker, M.O. (1996) Electrospray mass spectrometry of biomacromolecular complexes with noncovalent interactions – new analytical perspectives for supramolecular chemistry and molecular recognition processes. *Angewandte Chemie (International Edition in English)* **35**: 806–827.

Ronimus, R.S., and Musgrave, D.R. (1996a) A gene, *han1A*, encoding an archaeal histone-like protein from the *Thermococcus* species AN1: homology with eukaryal histone consensus sequences and the implications for delineation of the histone fold. *Biochem Biophys Acta* **1307**: 1–7.

Ronimus, R.S., and Musgrave, D.R. (1996b) Purification and characterization of a histone-like protein from the Archaeal isolate AN1, a member of the *Thermococcales*. *Mol Microbiol* **20**: 77–86.

Ronimus, R.S., Reysenbach, A.L., Musgrave, D.R., and Morgan, H.W. (1997) The phylogenetic position of the *Thermococcus* isolate AN1 based on 16S rRNA gene sequence analysis: a proposal that AN1 represents a new species, *Thermococcus zilligii* sp. nov. *Arch Microbiol* **168**: 245–248.

Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989) *Molecular cloning: a Laboratory Manual*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Sandman, K., Krzycki, J.A., Dobrinski, B., Lurz, R., and Reeve, J.N. (1990) HMf, a DNA-binding protein isolated from the hyperthermophilic archaeon *Methanothermus fervidus*, is most closely related to histones. *Proc Natl Acad Sci USA* **87**: 5788–5791.

Sandman, K., Perler, F.B., and Reeve, J.N. (1994a) Histone-encoding genes from *Pyrococcus*: evidence for members of the HMf family of archaeal histones in a non-methanogenic Archaeon. *Gene* **150**: 207–208.

Sandman, K., Grayling, R.A., Dobrinski, B., Lurz, R., and Reeve, J.N. (1994b) Growth-phase-dependent synthesis of histones in the archaeon *Methanothermus fervidus*. *Proc Natl Acad Sci USA* **91**: 12624–12628.

Schägger, H., and von Jagow, G. (1987) Tricine-sodium dodecyl sulfate–polyacrylamide gel electrophoresis for the separation of proteins in the range from 1 to 100 kDa. *Anal Biochem* **166**: 368–379.

Soares, D., Dahlke, I., Li, W., Sandman, K., Hethke, C., Thomm, M., et al. (1998) Archaeal histone stability, DNA binding, and transcription inhibition above 90°C. *Extremophiles* **2**: 75–81.

Tabassum, R., Sandman, K.M., and Reeve, J.N. (1992) HMt, a histone-related protein from *Methanobacterium thermoautotrophicum* ΔH. *J Bacteriol* **174**: 7890–7895.

Travers, A., and Muskhelishvili, G. (1998) DNA microloops and microdomains: a general mechanism for transcription activation by torsional transmission. *J Mol Biol* **279**: 1027–1043.

Tupper, A.E., Owen-Hughes, T.A., Ussery, D.W., Santos, D.S., Fergusson, D.J.P., Sidebottom, J.M., *et al.* (1994) The chromatin-associated protein H-NS alters DNA topology in vitro. *EMBO J* **13**: 258–268.

Ursitti, J., Mozdzanowski, J., and Speicher, D. (1996) Electroblotting from polyacrylamide gels. In *Current Protocols in Protein Science*. Coligan, J.E., Dunn, B.M., Ploegh, H.L., Speicher, D.W., and Wingfield, P.T. (eds). New York: John Wiley & Sons, pp. 10.7.1–10.7.14.

Wang, Q., Wu, J., Friedberg, D., Plakto, J., and Calvo, J. (1994) Regulation of the *Escherichia coli lrp* gene. *J Bacteriol* **176**: 1831–1839.

Xu, J., and Johnson, R.C. (1995) Identification of genes negatively regulated by Fis: Fis and RpoS comodulate growth-phase-dependent gene expression in *Escherichia coli*. *J Bacteriol* **177**: 938–947.

ELSEVIER

Biochimica et Biophysica Acta 1490 (2000) 115–120

BIOCHIMICA ET BIOPHYSICA ACTA
BBA

www.elsevier.com/locate/bba

Short sequence-paper

Identification of archaeal genes encoding a novel stationary phase-response protein

Marcel E. Dinger, David R. Musgrave *

Department of Biological Sciences, University of Waikato, Private Bag 3105, Hamilton, New Zealand

Received 23 August 1999; received in revised form 12 October 1999; accepted 25 October 1999

Abstract

A novel stationary phase-response protein has been identified in the acid-soluble protein extract of the thermophilic archaeon, *Thermococcus zilligii*. N-Terminal sequencing data were used to identify likely genes for homologues of the protein in the complete genome sequences of various archaeal species. The corresponding genes were identified and analysed. The genes encode a protein ranging from 83 to 92 amino acids in length, with a calculated p*I* ranging from 4.6 to 9.7. The amino acid sequences of the genes were highly conserved, even between members belonging to the different archaeal kingdoms. The computed secondary structure of the protein indicates it consists of four large helical regions separated by short coiled regions. We propose this protein as a candidate regulator of gene expression in stationary phase.

Keywords: Archaeon; Histone-fold; Histone; Stationary phase; Stress response; *Thermococcus*

In their natural environment, microorganisms must endure multiple stresses, including changes in temperature, osmolarity and nutrient availability [1]. Microorganisms respond to these stresses by undergoing adaptive change in gene expression, which leads to consequent physiological and morphological differentiation of the cells [2]. One mechanism for the sometimes global changes in gene expression invoked by stress is through alterations in the topology of the organism's genome. The change in DNA topology may be mediated directly, by the environmental stimuli themselves [3,4], or indirectly via changes in the levels or activity of DNA topoisomerases and proteins involved in DNA compaction [5,6]. In the hyperthermophilic archaeons, *Sulfolobus islandicus* and *Thermococcus* sp. GE31, plasmid topology has been observed to vary in response to changes in temperature and growth phase [7]. In *Methanothermus fervidus*, the relative level of the archaeal histone dimers $(HMfA)_2$ and $(HMfB)_2$ alters with growth phase, and due to the differing topological effects these dimers have on DNA, it has been suggested that this may be a mechanism to regulate gene expression [8]. Recently, we have found that the levels of histones in *Thermococcus zilligii*, which constitute approx. 5% of the total soluble protein in exponentially growing cells, dramatically decrease as cells entered stationary phase and could not be detected by late stationary phase (Dinger et al., submitted for publication). In addition, several other proteins increased in level as the cultures aged. Due to the important role that histones play in the compaction of DNA

* Corresponding author. Fax: +64 (7) 8384324;
E-mail: musgrave@waikato.ac.nz

0167-4781/00/$ – see front matter
PII: S0167-4781(99)00216-X

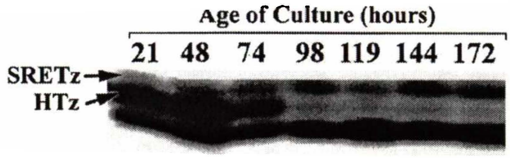


Fig. 1. Separation of total acid-soluble proteins from *T. zilligii*. Total protein was extracted from cell cultures of varying age, acid precipitated and subjected to SDS-PAGE. The positions of *T. zilligii* stationary phase-response protein E (SRETz) and *T. zilligii* histone (HTz) are indicated. Gel electrophoresis and silver staining were performed as described [10,29].

and gene expression, it was suggested that some of these proteins may take over the role of DNA compaction in the stationary phase and therefore influence gene expression. Here we report the identification and sequence analysis of one of these proteins.

Total cellular proteins were extracted from *T. zilligii* cultures of various ages, ranging from 21 h (mid-log) to 172 h (late stationary). In order to enrich for proteins that were likely to be associated with the nucleoid, the acid-soluble proteins were isolated [9]. The acid-soluble proteins were separated by SDS-PAGE and visualised by silver staining. A clearly resolved protein, tentatively named SRETz (stationary phase-response protein E) that migrated with molecular mass of between 7 and 8 kDa was observed to increase in level as the culture entered stationary phase, with a concomitant decrease in the histone HTz (48 h) (Fig. 1). The acid-soluble proteins from a 4-day-old culture were concentrated, resolved by SDS-PAGE and electroblotted onto PVDF membrane [10,11]. The proteins on the membrane were visualised by staining with amido black [12], and the band corresponding to SRETz was excised.

Using automated N-terminal sequencing, residues 3–12 of SRETz were determined to be ELIQQIVQVL; the first two residues could not be clearly identified. To locate potential protein homologues, the sequence was compared against the NCBI non-redundant database using the tblastn algorithm with the PAM70 amino acid similarity matrix [13]. The search returned only two sequences from Archaea: AE00118 (position 5334–5317) from *Archaeoglobus fulgidus* and AP000007 (position 185075–185098) from *Pyrococcus horikoshii*. The latter sequence corresponded to the 5′ end of a potential open reading frame, which encoded a hypothetical protein of 85 amino acids in length and a calculated molecular mass of 9667 Da (Fig. 1). The high homology of the N-terminus of SRETz with the N-terminus of the hypothetical protein, the similar molecular weight to SRETz, and the close phylogenetic relationship between *Pyrococcus* and *Thermococcus*,

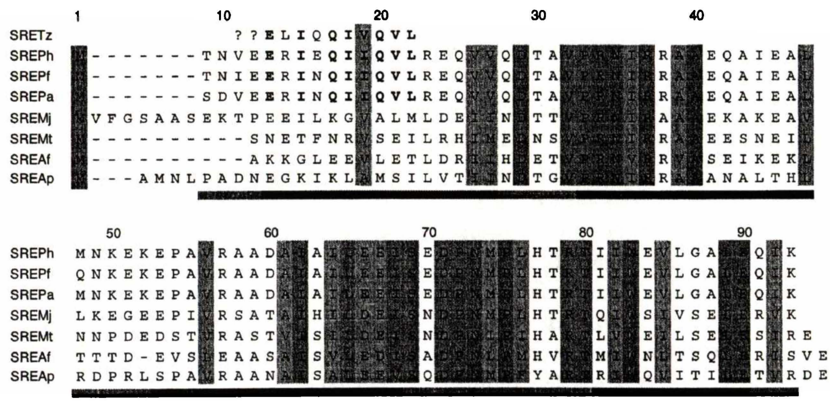


Fig. 2. Amino acid alignment of the putative stationary phase-response protein E (SRE) sequences from *P. horikoshii* (SREPh) [15], *P. furiosus* (SREPf), *P. abyssi* (SREPa), *M. jannaschii* (SREMj) [16], *M. thermoautotrophicum* (SREMt) [17], *A. fulgidus* (SREAf) [18], and *A. pernix* (SREAp) [30]. The N-terminal sequence (residues 3–12) of SRE from *T. zilligii* (SRETz) is shown, with homologous regions to the SRE sequences from other members of the Thermococcales highlighted in bold. Conserved residues are shown with dark shading, and similar residues with light shading. Residues were considered similar if they were hydrophobic-non-polar (A, V, I, L, P, M, F, W), polar non-charged (G, S, T, C, Y, N, Q), basic (H, R, K), or acidic (E, D) [31,32]. The consensus secondary structure, determined by the DSC algorithm [24], is aligned along the bottom of the sequences. α-Helical residues are shown in black and coiled residues in grey.

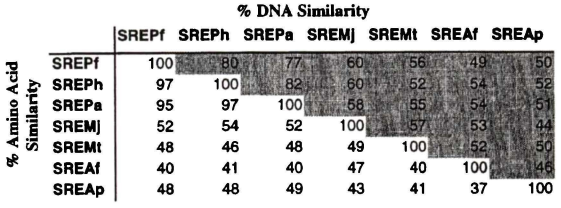


Fig. 3. Distance matrix indicating percent similarity between the SRE amino acid sequences (bottom) and encoding nucleotide sequences (top). The matrix was computed from the multiple alignment shown in Figs. 2 and 4, using ClustalW [33].

led to the conclusion that the hypothetical protein was a homologue of SRETz.

A BLAST search using the amino acid sequence of the SRETz homologue of *P. horikoshii* (SREPh), revealed putative SRE homologues in *Pyrococcus furiosus*, *Pyrococcus abyssi*, *Methanococcus jannaschii*, *Methanobacterium thermoautotrophicum*, *A. fulgidus* and *Aeropyrum pernix* (Fig. 2). None of the genomes contained more than one copy of the gene. To further represent the degree of similarity between the sequences, a distance matrix was constructed using the amino acid and DNA sequence alignment (Fig. 3). Clearly the protein is highly conserved amongst the Archaea, with the most distant sequences, SREAf from *Archaeoglobus* and SREAp from *Aeropyrum*, still possessing 37% similarity. Coupled with the existence of the SRE sequence in every archaeal genome sequence published to date, the sequence conservation strongly suggests that the protein performs a fundamental function and was present before the division of the crenarchaeal and euryarchaeal lineages [14]. However, since no convincing homology to the SRE sequences could be identified in bacterial or eucaryal genomes, it is likely that the *sre* gene evolved after the radiation of the three domains.

Interestingly, the SRE homologue of *A. fulgidus* (SREAf) had not been identified as an open reading frame and consequently could only be identified using a tblastn search. Furthermore, the detailed examination of the nucleotide sequences suggests that the open reading frames for the other organisms had been identified incorrectly. Fig. 4 shows the open reading frames that we predict from the SRE DNA sequences, together with the predicted start codons (ATG) in the published genome analyses for *P. horikoshii*, *M. jannaschii*, *M. thermoautotrophicum* and *A. fulgidus* [15–18]. Clearly, the identified location of the ribosome binding site (RBS) sequences [19] suggests that the open reading frames we propose are more substantive.

The SRE DNA sequences were analysed for other potential regulatory elements (Fig. 4). Using the TATA box consensus sequence, TTTAWA [20], only *sreMj* was found to possess a putative TATA binding motif corresponding to this consensus. However, several AT rich regions were detected 25–30 bp upstream from the ribosome-binding site of the other *sre* genes. The centre of the *sreMj* TATA box was located approx. 27 bp upstream from a potential transcriptional initiator element, as indicated by a pyrimidine-purine dinucleotide [21]. The absence of the TATA box in the other sequences suggests that the expression of these genes requires alternative basal transcription machineries. This is perhaps not surprising considering that the induction of SRETz occurs at the onset of stationary phase. A similar observation has been noted in the gene encoding the T_{ind} product of *Sulfolobus shibatae* virus 1 (SSV1), which is highly induced upon UV irradiation, but has no TATA box upstream from the start site [22]. Finally, each gene had long pyrimidine rich stretches following the stop codon (TGA or TAA), which is a typical characteristic of many thermophilic archaeal terminators [19].

The theoretical p*I* of the translated SRE sequences was determined [23]. The SRE homologues from members of the Euryarchaeota, SREPh, SREPf, SREPa, SREMj, SREMt, and SREAf, were all predicted to be acidic with p*I* values calculated as 4.7, 4.8, 4.7, 5.0, 4.6, and 4.9, respectively. Assuming that SRETz would have a similar p*I*, this was an unexpected result, since acidic proteins should have precipitated during the preparation of the acid wash extract. One explanation for this observation is that the protein associates with a basic protein and consequently remains soluble in acid conditions. Interestingly, SREAp, the most divergent of the sequences, was predicted to be basic with a calculated p*I* of 9.7. This result suggests that the overall acidic nature of the euryarchaeal proteins is not a critical characteristic in the function of the protein.

The amino acid sequences of SRE from each organism were aligned and analysed (Fig. 2). Using the DSC algorithm (version 1.2; Biomolecular Modelling

Fig. 4. DNA sequence alignment of the genes encoding stationary phase-response protein E from *P. horikoshii* (*srePh*) [15], *P. furiosus* (*srePf*), *P. abyssi* (*srePa*), *M. jannaschii* (*sreMj*) [16], *M. thermoautotrophicum* (*sreMt*) [17], *A. fulgidus* (*sreAf*) [18], and *A. pernix* (*sreAp*) [30]. Conserved residues are indicated in bold print. Predicted regulatory sequences are labelled along the bottom and indicated in the alignment by the shaded regions, with the dark shading indicating identity with the consensus sequences for these elements. The previously identified start codons, as predicted in the respective genome databases, are outlined. The GenBank primary accession numbers and positions of the corresponding sequences are: *srePh*, AP000007 (184994..185360); *sreMj*, U67582 (complement of 6228..5841); *sreMt*, AE000903 (complement of 4360..4047); *sreAf*, AE001112 (5227..5638); *sreAp*, AP000064 (complement of 72105..71715).

Laboratory, London) [24], a consensus secondary structure of SRE was computed from the sequences (Fig. 2). The predicted structure, with an estimated accuracy of 85%, indicates the protein consists of four large helical regions, separated by short coiled regions. Use of the Predator [25], GORIV [26] and MLRC [27] algorithms on individual sequences yielded similar predictions (results not shown). As shown in Fig. 2, the first and third coiled regions and the beginning of the subsequent α-helical regions align with the most conserved residues found, suggesting that these structures have a critical role in the functioning and/or correct folding of the protein. This is similar to the histone-fold motif, where the most conserved regions are also those involved in the coiled regions and start of the α-helices, since these are critical in determining the relative positions of the helical segments and in binding DNA [28]. The

correct positioning of the archaeal histone helices is fundamental in quaternary structure formation and the non-specific binding to DNA [28]. In addition, the SRE sequences contain a high proportion of hydrophobic residues in the predicted helical residues (59%). This is also a characteristic of the histone-fold, where the hydrophobic residues in the central helix form the surface for dimerisation [28]. Consequently, we suggest that the SRE protein may form or be part of a larger quaternary structure.

The predicted helix-turn-helix structure of the SRE proteins, together with the observation that SRETz appears concurrently with the disappearance of HTz as the culture enters stationary phase, leads to the tempting speculation that SRE may bind to DNA and take over the role of DNA compaction in stationary phase. The low molecular weight of the SRE proteins and the high abundance of SRETz in stationary phase cells further support this hypothesis. The similarity in predicted secondary structure between the SRE proteins and the archaeal histones also opens the possibility that these proteins may associate with similar machineries to the histones, or even with the histones themselves, in the transition to stationary phase. Consequently, we propose this protein as a candidate regulator of gene expression in stationary phase. To investigate this possibility, we are presently in the process of characterising the function of this protein.

Acknowledgements

Preliminary sequence data for *P. furiosus* were obtained from the Center of Marine Biotechnology (University of Maryland, USA) website at http://www.umbi.umd.edu. Sequencing of *P. furiosus* was accomplished with support from the US Department of Energy. We thank Patrick Forterre (Université Paris-Sud, France) for preliminary sequence data for *P. abyssi*.

References

[1] R. Kolter, D.A. Siegele, A. Tormo, Annu. Rev. Microbiol. 47 (1993) 855–874.

[2] Hengge-Aronis, R. (1996) in: F.C. Neidhardt, R. Curtisss, J.L. Ingrahmam, E.C.C. Lin, K.B. Low, B. Magasanik, W.S. Reznikoff, M. Riley, M. Schaechter, H.E. Umbarger (Eds.), *Escherichia coli* and *Salmonella*: Cellular and Molecular Biology, 2nd edn., ASM Press, Washington, DC, pp. 1497–1512.

[3] C.S.J. Hulton, A. Seirafi, J.C.D. Hinton, J.M. Sidebotham, L. Waddell, G.D. Pavitt, T. Owen-Hughes, A. Spassky, H. Buc, C.F. Higgins, Cell 63 (1990) 631–642.

[4] J. McClellan, P. Boublikova, E. Palecek, D. Lilley, Proc. Natl. Acad. Sci. USA 87 (1990) 8373–8377.

[5] C.F. Higgins, J.C. Hinton, C.S. Hulton, T. Owen-Hughes, G.D. Pavitt, A. Seirafi, Mol. Microbiol. 4 (1990) 2007–2012.

[6] M. Almirón, A. Link, D. Furlong, R. Kolter, Genes Dev. 170 (1992) 5169–5170.

[7] P. Lòpez-Garcìa, P. Forterre, Mol. Microbiol. 23 (1997) 1267–1279.

[8] K. Sandman, R.A. Grayling, B. Dobrinski, R. Lurz, J.N. Reeve, Proc. Natl. Acad. Sci. USA 91 (1994) 12624–12628.

[9] R.S. Ronimus, D.R. Musgrave, Mol. Microbiol. 20 (1996) 77–86.

[10] H. Schägger, G. von Jagow, Anal. Biochem. 166 (1987) 368–379.

[11] J. Ursitti, J. Mozdzanowski, D. Speicher, in: J.E. Coligan, B.M. Dunn, H.L. Ploegh, D.W. Speicher, P.T. Wingfield (Eds.), Current Protocols in Protein Science, John Wiley and Sons, New York, 1996, pp. 10.7.1–10.7.14.

[12] S. Harper, D. Speicher, in: J.E. Coligan, B.M. Dunn, H.L. Ploegh, D.W. Speicher, P.T. Wingfield (Eds.), Current Protocols in Protein Science, John Wiley and Sons, New York, 1996, pp. 10.8.1–10.8.7.

[13] S. Altschul, T. Madden, A. Schaffer, J. Zhang, Z. Zhang, W. Miller, D. Lipman, Nucleic Acids Res. 25 (1997) 3389–3402.

[14] C.R. Woese, O. Kandler, M.L. Wheelis, Proc. Natl. Acad. Sci. USA 87 (1990) 4756–4759.

[15] Y. Kawarabayasi, M. Sawada, H. Horikawa, Y. Haikawa, Y. Hino, S. Yamamoto, M. Sekine, S. Baba, H. Kosugi, A. Hosoyama, Y. Nagai, M. Sakai, K. Ogura, R. Otsuka, H. Nakazawa, M. Takamiya, Y. Ohfuku, T. Funahashi, T. Tanaka, Y. Kudoh, J. Yamazaki, N. Kushida, A. Oguchi, K. Aoki, H. Kikuchi, DNA Res. 5 (1998) 55–76.

[16] C. Bult, O. White, G. Olsen, L. Zhou, R. Fleischmann, G. Sutton, J. Blake, L. Fitzgerald, R. Clayton, J. Gocayne, A. Kerlavage, B. Dougherty, J. Tomb, M. Adams, C. Reich, R. Overbeek, E. Kirkness, K. Weinstock, J. Merrick, A. Glodek, J. Scott, N. Geoghagen, J. Weidman, J. Fuhrmann, D. Nguyen, T. Utterback, J. Kelley, J. Peterson, P. Sadow, M. Hanna, M. Cotton, K. Roberts, M. Hurst, B. Kaine, M. Borodovsky, H. Klenk, C. Fraser, H. Smith, C. Woese, J. Venter, Science 273 (1996) 1058–1073.

[17] D.R. Smith, Doucette-Stamm, C. Deloughery, H. Lee, J. Dubois, T. Aldredge, R. Bashirzadeh, D. Blakely, R. Cook, K. Gilbert, D. Harrison, L. Hoang, P. Keagle, W. Lumm, B. Pothier, D. Qiu, R. Spadafora, R. Vicaire, Y. Wang, J. Wierzbowski, R. Gibson, N. Jiwani, A. Caruso, D. Bush, H. Safer, D. Patwell, S. Parbhakar, S. McDougall, G. Shimer, A. Goyal, S. Pietrokovski, G.M. Church, C.J.

Daniels, J. Mao, P. Rice, J. Nölling, J.N. Reeve, J. Bacteriol. 179 (1997) 7135–7155.
[18] H.P. Klenk, R.A. Clayton, J.F. Tomb, O. White, K.E. Nelson, K.A. Ketchum, R.J. Dodson, M. Gwinn, E.K. Hickey, J.D. Peterson, D.L. Richardson, A.R. Kerlavage, D.E. Graham, N.C. Kyrpides, R.D. Fleischmann, J. Quackenbush, N.H. Lee, G.G. Sutton, S. Gill, E.F. Kirkness, B.A. Dougherty, K. McKenney, M.D. Adams, B. Loftus, S. Peterson, C.I. Reich, L.K. McNeil, J.H. Badger, A. Glodek, L. Zhou, R. Overbeek, J.D. Gocayne, J.F. Weidman, L. McDonald, T. Utterback, M.D. Cotton, T. Spriggs, P. Artiach, B.P. Kaine, S.M. Sykes, P.W. Sadow, K.P. D'Andrea, C. Bowman, C. Fujii, S.A. Garland, T.M. Mason, G.J. Olsen, C.M. Fraser, H.O. Smith, C.R. Woese, J.C. Venter, Nature 390 (1997) 364–370.
[19] J. Brown, C. Daniels, J. Reeve, CRC Crit. Rev. Microbiol. 16 (1989) 287–338.
[20] R. Wolf-Deiter, U. Hüdepohl, Z. Zillig, Proc. Natl. Acad. Sci. USA 87 (1990) 9509–9513.
[21] J. Hain, W. Reiter, U. Hudepohl, W. Zillig, Nucleic Acids Res. 20 (1992) 5423–5428.
[22] C. Schleper, K. Kubo, W. Zillig, Proc. Natl. Acad. Sci. USA 89 (1992) 7645–7649.
[23] B. Bjellqvist, G. Hughes, C. Pasquali, N. Paquet, F. Ravier, J. Sanchez, S. Frutiger, D. Hochstrasser, Electrophoresis 14 (1993) 1023–1031.
[24] R. King, M. Sternberg, Protein Sci. 5 (1996) 2298–2310.
[25] D. Frishman, P. Argos, Proteins 27 (1997) 329–335.
[26] J. Garnier, J. Gibrat, B. Robson, Methods Enzymol. 266 (1996) 540–553.
[27] Y. Guermeur, C. Geourjon, P. Gallinari, G. Del age, Bioinformatics 15 (1999) 413–421.
[28] G. Arents, E.N. Moudrianakis, Proc. Natl. Acad. Sci. USA 92 (1995) 11170–11174.
[29] J.H. Morrissey, Anal. Biochem. 117 (1981) 307–310.
[30] Y. Kawarabayasi, Y. Hino, H. Horikawa, S. Yamazaki, Y. Haikawa, K. Jin-no, M. Takahashi, M. Sekine, S. Baba, A. Ankai, H. Kosugi, A. Hosoyama, S. Fukui, Y. Nagai, K. Nishijima, H. Nakazawa, M. Takamiya, S. Masuda, T. Funahashi, T. Tanaka, Y. Kudoh, J. Yamazaki, N. Kushida, A. Oguchi, H. Kikuchi, DNA Res. 6 (1999) 83–101, 145–152.
[31] S. Karlin, G. Ghandour, Proc. Natl. Acad. Sci. USA 82 (1985) 8597–8601.
[32] A.L. Lehninger, Biochemistry, Worth Publishers, New York, 1982.
[33] J.D. Thompson, D.G. Higgins, T.J. Gibson, Nucleic Acids Res. 22 (1994) 4673–4680.

ASTROBIOLOGY
Volume 2, Number 3, 2002

Archaeal Genome Organization and Stress Responses: Implications for the Origin and Evolution of Cellular Life

DAVID MUSGRAVE, XIAOYING ZHANG,* and MARCEL DINGER

ABSTRACT

For DNA to be used as an informational molecule it must exist in the cell on the edge of stability because all genomic processes require local controlled melting. This presents mechanistic opportunities and problems for genomic DNA from hyperthermophilic organisms, whose unpackaged DNA could melt at optimal temperatures for growth. Hyperthermophiles are suggested to employ the novel positively supercoiling topoisomerase enzyme reverse gyrase (RG) to form positively supercoiled DNA that is intrinsically resistant to thermal denaturation. RG is presently the only archaeal gene that is uniquely found in hyperthermophiles and therefore is central to hypotheses suggesting a hypothermophilic origin of life. However, the suggestion that RG has evolved by the fusion of two pre-existing enzymes has led to hypotheses for a lower temperature for the origin of life. In addition to the action of topoisomerases, DNA packaging and the intracellular ionic environment can also manipulate DNA topology significantly. In the Euryarchaeota, nucleosomes containing minimal histones can adopt two alternate DNA topologies in a salt-dependent manner. From this we hypothesize that since internal salt concentrations are increased following an increase in temperature, the genomic effects of temperature fluctuations could also be accommodated by changes in nucleosome organization. In addition, stress-induced changes in the nucleoid proteins could also play a role in maintaining the genome in the optimal topological state in changing environments. The function of these systems could therefore be central to temperature adaptation and thus be implicated in origin of life scenarios involving hyperthermophiles. Key Words: Archaeal histones—Genome packaging—Genome topology—Stationary phase. Astrobiology 2, 241–253.

INTRODUCTION

DNA SEQUENCE COMPARISONS of genes encoding the small subunit rRNA component of ribosomes (16S rRNA in prokaryotes and 18S rRNA in eukaryotes) have revealed that cellular life can be placed into one of three domains: Bacteria, Archaea, or Eucarya (Woese, 1987; Woese *et al.*, 1990). Archaea are morphologically and metabolically similar to Bacteria, but their genes involved in the information processing systems, including transcription, translation, DNA replication, and recombination, most closely resemble eucaryal orthologues (Rivera *et al.*, 1998). However, the eucaryal-like nature of the informational genes in Archaea could also be the result of dif-

Department of Biological Sciences, University of Waikato, Hamilton, New Zealand.
*Present address: Nucleics, University of New South Wales, Sydney, Australia.

ferential rates of evolution for metabolic and informational processing genes (Forterre and Phillipe, 1999). The phylogenetic trees based on 16S rRNA sequences also highlight two phenotypically diverse groups in the archaeal domain: the Euryarchaeota and the Crenarchaeota (Woese *et al.*, 1990). One of the major differences between them is that most euryarchaeal species, like all of the Eucarya, have a histone-containing, nucleosome-based DNA condensation system (Ronimus and Musgrave, 1996a; Sandman and Reeve, 2001). This observation, together with the existence of a minimal eucaryal information processing system, predicts that the archaeal and eucaryal domains shared an important evolutionary history subsequent to the universal ancestor stage (Olsen and Woese, 1997).

For many biologists, an interesting outcome of phylogenetic trees is that they predict that the last universal common ancestor was a hyperthermophilic archaeon because these organisms are the closest extant organisms to the root of these trees. However, a vast body of work has cast doubt on the validity of these trees, particularly the position of the root (Forterre and Phillipe, 1999). Doolittle (1999) goes further and argues for a net-like rather than a tree-like structure to describe prokaryotic evolution, describing prokaryotic taxa as being "imprecisely bounded and ephemeral." In 1998, Woese proposed a genetic annealing model for early cellular life in which life originated as a genetically plastic progenote. He described this by saying "the universal ancestor is not a discrete entity. It is, rather a diverse community of cells that survives and evolves as a biological unit." This harks back to the suggestion of Reanney (1974), several decades ago, that we should more realistically view prokaryotes as one global superorganism within which genes are exchanged at different frequencies. The extent and importance of horizontal transfer in early evolution are, in part, demonstrated by the mosaic nature of the genomes of extant prokaryotes. For example, the fully sequenced genome of the deeply rooted thermophilic bacterium *Thermotoga maritima* has shown that 24% of its genes have greater sequence similarity with archaeal genes than any other bacterial or eucaryal gene and that these genes are in defined clusters with conserved gene order and have an atypical mol percent G+C compared with the rest of the genes in the genome (Nelson *et al.*, 1999). However, Forterre and Phillipe (1999) have argued that the mosaicism found in many genomes may be due to differential gene loss or nonorthologous replacement.

The potential nature of the last universal common ancestor can also be predicted by understanding the physical conditions that could have provided environments for the development and evolution of cellular life. The discovery of deep-sea hydrothermal vents (Corliss *et al.*, 1979) has provided such an environment, and the physiology of the Archaea and Bacteria that inhabit hydrothermal vents continues to challenge our preconceptions about the physical limits of conventional life and Darwin's suggestion, found in a private letter to Hooker, of a "warm little pond" for the site of the origin of life itself. Investigations of hyperthermophilic microorganisms have formed the basis of the "hot" origin of life for a number of reasons other than because phylogenetic studies place them close to the root of the universal tree of life. Firstly, they exist in conditions expected to have been common during formation of the earth. Secondly, they could have survived meteorite impacts at the earth's surface. And finally, many Archaea have novel resistance mechanisms, such as extreme radiation resistance, that could have been used to protect them in these extreme environments (Peak *et al.*, 1995; Gerard *et al.*, 2001).

Comparative genomic studies of prokaryotes have shown that only one gene, encoding a type I topoisomerase enzyme, reverse gyrase (RG), is uniquely found in hyperthermophiles. RG is a topoisomerase that has the unique potential to use ATP to produce positively supercoiled DNA (Kikuchi and Asai, 1984), and it has been correlated with the presence of positively supercoiled plasmid and viral DNA in hyperthermophilic organisms. Positively supercoiled DNA is suggested to have greater resistance to thermal denaturation than the negatively supercoiled DNA found in all other cells (Lopez-Garcia and Forterre, 2000). However, both the formation of archaeal nucleosomes and the extreme intracellular salt concentrations also have a significant affect on the structural stability of DNA (Anderson and Bauer, 1978; Ronimus and Musgrave, 1996b).

Here we propose that the ability of archaeal nucleosomes to form both positive and negative supercoils in a salt-dependent manner (Musgrave *et al.*, 2000) and a novel stress response in hyperthermophilic Archaea (Dinger *et al.*, 2000) give us additional clues to how these organisms can

adapt their genome topology to changing environments and act as an alternate system to RG, allowing hypotheses for hyperthermophilic origins of life to be revisited.

MATERIALS AND METHODS

Reference strains

Pyrococcus abyssi GE5 was obtained from Patrick Forterre, University of Paris XI, Paris, France, and *Thermococcus zilligii* was obtained from the Thermophile Research Unit, University of Waikato, Hamilton, New Zealand.

Culturing of T. zilligii *and* P. abyssi

T. zilligii was grown in 850-mL volumes, in 1.0-L Schott bottles as previously described (Klages and Morgan, 1994), and *P. abyssi* cells were cultivated in "YPS" medium (Erauso *et al.*, 1993). The OD_{600} of the samples was measured and converted to cell density by multiplying the OD_{600} by a conversion factor empirically determined by cell counting as described (Koch, 1994).

Isolation of RNA

Total RNA was isolated using TRIZOL LS Reagent (GibcoBRL, Life Technologies, Ltd.), according to the manufacturer's instructions. RNA was recovered after phenol (pH 4.0)/chloroform and chloroform extraction, then precipitated and resuspended in RNase-free water, and stored at −70°C. The concentration and quality of RNA were estimated by absorbance (A_{260}, A_{260}/A_{280}, and A_{260}/A_{230}) measurement.

Northern blot procedures

For northern blots, total RNA samples were denatured by incubation for 15 min at 65°C in an equal volume of a solution containing 20% formalin, 50% formamide, and 0.1 *M* sodium 3-(*N*-morpholino)propanesulfonic acid (MOPS) and then separated by electrophoresis through 2% (wt/vol) agarose gels containing 2.2 *M* formaldehyde, 20 m*M* MOPS, and 8 m*M* sodium acetate at 4.2 V cm^{-1}. After washing the gels for 30 min in 10× saline–sodium citrate (SSC), the RNA was transferred to Hybond$^+$ nylon membranes by capillary blotting for 12 h in 10× SSC. Nucleic acids were UV-cross-linked to the membrane by irradiating the membrane for 45 s with an UV transilluminator. Membrane hybridization was performed essentially as described by Dinger and Musgrave (2000).

Electroblotting from polyacrylamide gels onto PVDF membranes for N-terminal sequencing

Electroblotting of proteins from polyacrylamide gels onto a retentive membrane was used for western blotting and N-terminal protein sequencing (Ausubel *et al.*, 1998). For N-terminal sequencing Trans-Blot membranes (Bio-Rad) were used to transfer proteins from sodium dodecyl sulfate-containing polyacrylamide electrophoresis gels.

The transfer procedure was as described by Ausubel *et al.* (1998). After blotting, the gel was stained with silver or Coomassie Brilliant Blue to check transfer efficiency and photographed. The PVDF membrane was stained with 10% (wt/vol) amido black (naphthol blue black 10B, Sigma) in 10% (vol/vol) acetic acid, destained with 10% acetic acid/30% ethanol, and then drip-dried. The desired band on the membrane was excised and sent for protein N-terminal sequencing. Sequencing was carried out using Edman degradation chemistry on an Applied Biosystem Procise 492 protein sequencer at the Sequencing Facility, Auckland University, Auckland, New Zealand.

Radioactive labeling of DNA probes and ladders

DNA probes were prepared from polymerase chain reaction (PCR) products by random primer labeling using the Rediprime™ II DNA labeling system (Amersham Pharmacia Biotech), following the manufacturer's instructions. The labeled DNA was denatured by adding 7.32 μL of 10 *M* NaOH 10 min prior to use. Single-stranded oligonucleotides were end-labeled using T4 polynucleotide kinase as described by Sambrook *et al.* (1989), stored at 4°C, and used within 3 months of preparation.

Preparation of archaeal histone proteins

Native HTz was purified as previously described (Ronimus and Musgrave, 1996b). Recombinant HMfA and HMfB were purified from *Escherichia coli* as described (Sandman *et al.*, 1990), and recombinant MkaH was purified from *E. coli* as described (Fahrner *et al.*, 2001).

Preparation of DNA substrates

pUC18 plasmid was isolated by alkaline lysis (Sambrook *et al.*, 1989) and CsCl density gradient centrifugation followed by treatment with proteinase K (Sigma), phenol/chloroform extraction, and ethanol precipitation. Relaxed pUC18 plasmids were produced by treating negatively supercoiled plasmid DNA with Topo V (Fidelity Systems) at 70°C or 85°C or with wheat germ Topo I (BRL) at 37°C followed by phenol/chloroform extraction and ethanol precipitation. pUC18 topoisomers of different superhelical densities were produced by relaxation in the presence of ethidium bromide as previously described (Pugh *et al.*, 1989) except that Topo V was used. Positively supercoiled pUC18 control topoisomers were produced by incubation of negatively supercoiled pUC18 DNA with RG (Kikuchi and Asai, 1984).

DNA topology assays

Topology assays and two-dimensional gels to determine the handedness of topoisomers were performed as described (Musgrave *et al.*, 2000).

Gel visualization and photography

Nucleic acids, stained with ethidium bromide, were visualized under UV light, with a TFx-35 UV transilluminator (Life Technologies). The gels were photographed using a COHU high-performance CCD camera and printed on a Sony UP-D890 Digital Graphic Printer. A digital imaging program, Scion Image (release beta 2 12/5/97, Scion Corp.) was used for manipulation of the images.

RESULTS

Recently it has become clear that in eukaryotes chromatin and the packaging of DNA in nucleosomes are intimately involved in the control of gene expression (Wu and Grunstein, 2000). Also, the discovery of histone acetyltransferases, histone deacetylases, and other chromatin remodeling complexes, which can specifically modify the structure of chromatin, has suggested that eukaryotic chromatin is structurally dynamic (Jenuwein and Allis, 2001). The structural plasticity of chromatin is also evidenced by the production of nucleosomes containing H2A.Z, a novel histone that promotes nucleosome instability when incorporated into chromatin, and the ability of nucleosomes to wrap DNA in two opposite configurations forming positive and negative toroidal supercoils (Hamiche *et al.*, 1996; Hamiche and Richard-Foy, 1998; Santisteban *et al.*, 2000). We have suggested that archaeal nucleosomes are also structurally plastic because of their ability to form both positive and negative supercoils in a salt-dependent manner and, in *T. zilligii*, possibly by direct modifications to the nucleoid (Dinger and Musgrave, 2000; Musgrave *et al.*, 2000).

To examine the effect that archaeal chromatin has on global gene expression control we have tested stationary-phase changes in the archaeal nucleoid and have examined the level of transcripts of the histone genes and genes encoding nucleoid proteins at various growth phases in *P. abyssi*, another member of the Thermococcales. To test whether both the *P. abyssi* histone genes (*hpa1* and *hpa2*) and the stress response genes (*srbPa1* and *srbPa2*) were being transcribed in a growth phase-dependent manner, as was found for the homologous genes in *T. zilligii* (Dinger, 1998), total RNA samples were isolated from cells at different growth phases as depicted in Fig. 1. The RNA samples shown in Fig. 2 were subject to northern blotting and hybridization analysis. Individual duplicate N^+ membranes containing total RNA samples were hybridized with one of the probes listed in Table 1. Figure 2 shows that RNA degradation was minimal, as indicated by the equivalent intensities and lack of smearing of the 23S and 16S rRNA bands for each RNA sample.

The northern hybridization results for *hpa1*, *hpa2*, and *srbPa1*/*srbPa2* are shown in Figs. 3–5, respectively. A single transcript of 206 bp for *hpa1* and *hpa2* and two transcripts of the predicted sizes of *srbPa1* (585 bp) and *srbPa2* (417 bp) were detected. Southern blots were hybridized at the same time in the same hybridization tube, with the same probe used for northern hybridization as a control for the integrity of the probes used and to confirm the *P. abyssi* genome sequence annotation (Figs. 3B and 4B). The northern hybridization autoradiographs were quantified, and the relative level of each transcript was determined in relation to the level of 23S rRNA in each sample from Fig. 2 (Figs. 3C, 4C, and 5B). The intensity of *hpa1* and that of *hpa2* did not show any significant change at any of the points in the growth phase that were tested.

In contrast, the level of both *srbPa1* and *srbPa2*

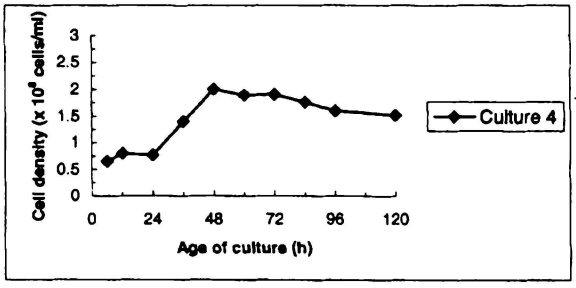


FIG. 1. **Growth curve of a representative batch culture of *P. abyssi* cells grown in "YPS" medium at 80°C.** The cell density of the culture versus culture age is indicated.

transcripts increased significantly in stationary-phase cells, indicative of induction of *srbPa1* and *srbPa2* gene transcription in these cells. Both *srbPa1* and *srbPa2* transcripts were able to be detected using the *srbPa1* gene probe because these two genes are ~40% identical, but because a single probe was used the relative levels of *srbPa1* and *srbPa2* could not be determined.

Three other genes of *P. abyssi* (*srcPa, srdPa,* and *srePa,* described as encoding stationary-phase response proteins in *T. zilligii*) were expressed constitutively (results not shown). Constitutive expression of these genes was surprising because *T. zilligii* and *P. abyssi* are both members of a same archaeal family, the Thermococcales, and each contain the same suite of highly conserved stationary-phase stress genes (Dinger and Musgrave, 2000). The significantly different expression of the histone and stationary-phase stress genes in these species suggests that the response to stationary-phase stress in these organisms is significantly different. The results also suggest that the depletion of nucleosomes is not a significant factor in the control of global gene expression in *P. abyssi,* as has been suggested for *T. zilligii.*

The dynamic behavior of archaeal nucleosomes has been demonstrated by the ability of the archaeal histones to wrap DNA in both positive and negative supercoils in a salt-dependent manner. This is also seen in Fig. 6, which shows the toroidal wrapping of a number of different histones at a salt concentration in which both positive and negative topoisomers are produced. The results for the HMf histone from *Methanothermus fervidus* and the HTz histone from *T. zilligii* show that, at a salt concentration of 150 m*M*, negative supercoils are produced at protein/DNA mass ratios below ~0.8 (lanes a–e for HMfB, lanes a–f for HMfA+B and HTz). Furthermore, positive supercoils are produced at protein/DNA mass ratios above this (lanes f–j for HMfB and lanes g–j and g–i for HMfA+B and HTz, respectively). We describe the ability of the archaeal nucleosome to adopt two conformations as "nucleosome flipping" in order to suggest that one conformation can be converted to the other, as is seen in the production of both conformations in the same tube. This is most evident in lane g of Fig. 6, in which topoisomers of pUC18 are produced by the binding of the HTz histone isolated from *T. zilligii.* In contrast, only negative supercoils were produced using the histone MkaH from *Methanopyrus kandleri.* MkaH is an unusual archaeal histone in that it is two histone monomers fused in a single polypeptide (Fahrner *et al.,* 2001). We have proposed that this fusion prevents the rearrangement of the monomers that are necessary for the structural changes that allow two nucleosome conformations to occur (Musgrave *et al.,* 2000). Models aimed at explaining the significance of nucleosomal flipping are presented in Figs. 7 and 8. In 1987, Liu and Wang proposed that the tracking of a macromolecule along a DNA molecule would result in the production of positive supercoils in front of the advancing complex and that the equivalent number of negative

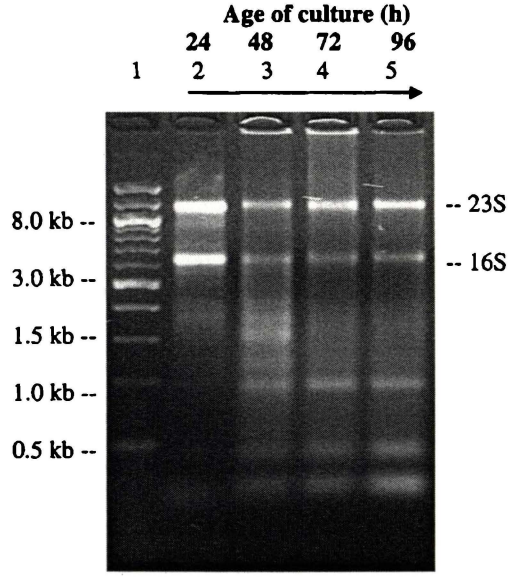


FIG. 2. **Total RNA samples extracted from *P. abyssi* cells at various stages of growth, resolved by TBE agarose gel electrophoresis.** Lane 1 contains a 1-kb DNA ladder. Lanes 2–5 each contain 2 μg of total RNA, isolated from cells at the times indicated.

TABLE 1. PRIMERS USED FOR PCR AMPLIFICATION OF *P. ABYSSI* GENE PROBES USED IN NORTHERN ANALYSES

HPa1-F	ATGGGAGAGTTG CCAATT
HPa1-R	TCAGCTCTTAATA GCCAA
HPa2-F	ATGGCTGAGTTGC CAATT
HPa2-R	GTCAGCTCCTAAT TGCAA
SrbPa1-F	GTGATGCATATGA GGATCGCGGTTCC AACTAA
SrbPa1-R	AAAAAGGCTCTTC CGCACCACCACCA GCCGTAGAGCCACC

The oligonucleotide primers listed were used to PCR-amplify the histone and stress response protein B genes from *P. abyssi* genomic DNA. PCR products were purified, labeled with γ-^{33}P, and used as a probes for the *P. abyssi hpa* and *srb* transcripts in northern hybridizations using total RNA extracted from *P. abyssi* cells of various ages. The start and stop codons of the respective open reading frames are indicated in boxes, and restriction recognition sites are underlined.

supercoils would be produced behind. This could be envisaged to occur when DNA is replicated or transcribed by the tracking of DNA or RNA polymerase. The production of positive supercoils ahead of a DNA tracking complex is expected to modulate gene expression by decreasing the stability of nucleosomes because nucleosomes have been shown to be unstable on positively supercoiled DNA templates *in vitro*. Therefore, transcriptionally mediated supercoiling is thought to be another way in which the repressive effects of nucleosomes on transcription could be overcome by clearing nucleosomes from in front of the advancing RNA polymerase. This could also contribute to polar effects, when the transcription of one gene is seen to affect the expression of a gene

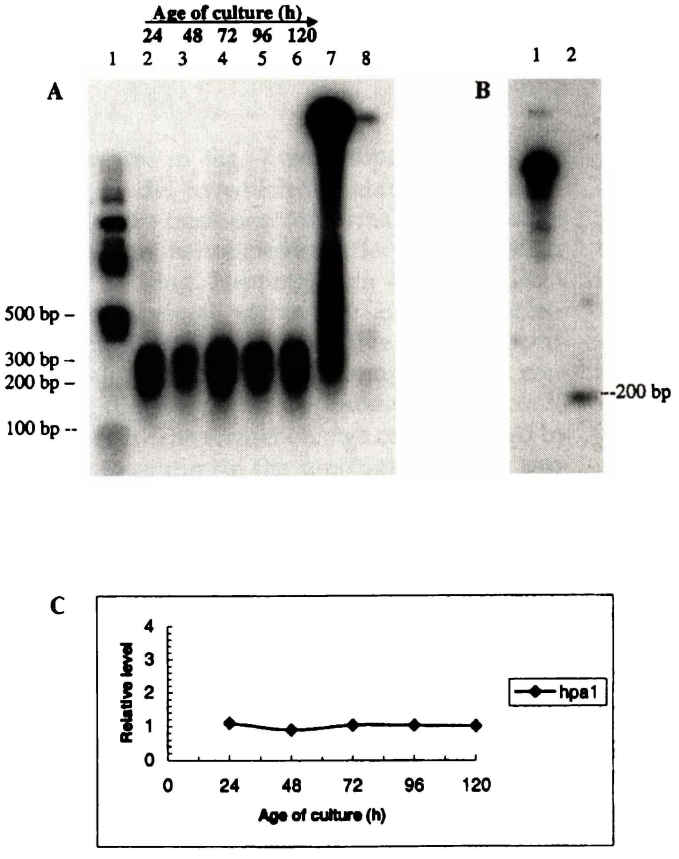


FIG. 3. A: Northern hybridization of total RNA from *P. abyssi* cells after various periods of growth, with the *hpa1* gene probe. Lane 1 contains 1.5 μg of γ-^{33}P-5′ end-labeled DNA marker, with sizes indicated. Lanes 2–6 each contain 10 μg of total RNA from cells of the ages indicated. Lane 7 contains 1 ng of *hpa1* PCR product, and lane 8 contains 1 ng of *hpa2* PCR product, as a positive control. **B:** Southern hybridization controls for *hpa1*. Lane1 contains 3 μg of *Sac*I-digested *P. abyssi* genomic DNA, and lane 2 contains 0.5 ng of the *hpa1* PCR product. **C:** Levels of *hpa1* mRNA relative to the corresponding 23S rRNA levels versus culture age. Levels were determined by densitometric analysis of the northern blots shown in (A) and were compared with the level of 23S rRNA stained with methylene blue.

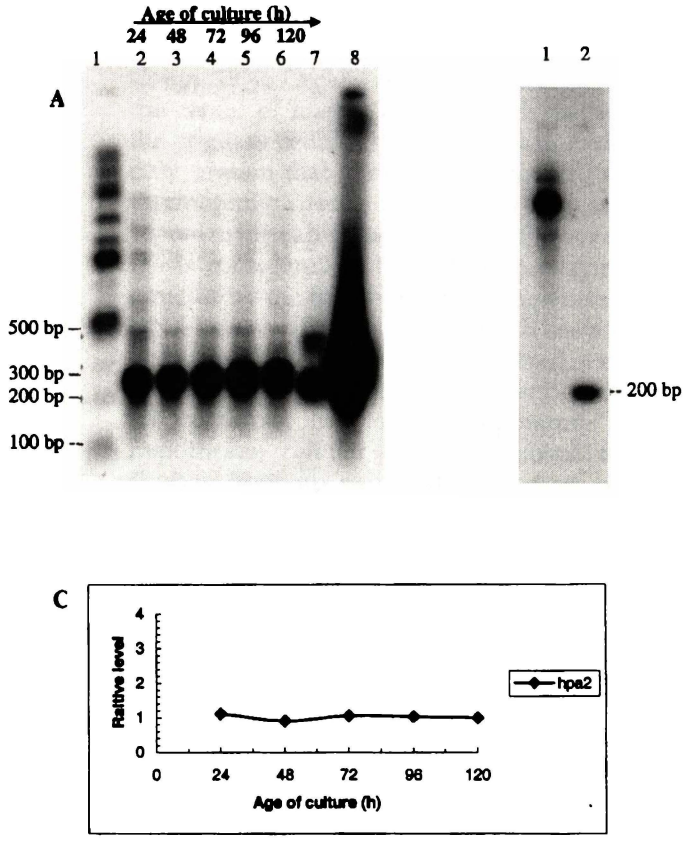


FIG. 4. A: Northern hybridization of total RNA from *P. abyssi* cells after various periods of growth, with the *hpa2* gene probe. Lane 1 contains 1.5 μg of γ-^{33}P-5′ end-labeled DNA marker, with sizes indicated. Lanes 2–6 each contains 10 μg of total RNA from cells of the ages indicated. Lane 7 contains 1 ng of *hpa1* PCR product, and lane 8 contains 1 ng of *hpa2* PCR product as positive controls. B: Southern hybridization controls for *hpa2*. Lane 1 contains 3 μg of *Sac*I-digested *P. abyssi* genomic DNA, and lane 2 contains 0.5 ng of *hpa2* PCR product. C: Relative levels of *hpa2* mRNA to the corresponding 23S rRNA levels versus their culture ages. The mRNA levels were determined as described for Fig. 3.

downstream. In Fig. 7 we propose a topology sensing model in which the ability for nucleosomes to flip from one conformation to another could negate the topological effects produced by protein tracking. Flipping from a negative to a positive supercoil within the archaeal nucleosome could minimize the positive supercoils produced ahead of a tracking complex, and, conversely, flipping in the opposite direction could maximize it. The former change could be used by the cell to minimize the topological effects produced by transcription or replication, whereas the latter change could maximize the ability of the advancing complex to clear nucleosomes ahead of it and hence modify the topology of downstream genes. A temperature sensing model is proposed in Fig. 8 in which nucleosome flipping could be used by the cell to accommodate changes in the temperature of its surrounding environment. Many hyperthermophilic archaeal species contain concentrations of salts in their cytoplasm as high as 1 *M*, and increased internal salt concentrations are correlated with increased growth temperatures in archaeal cells (Hensel and Konig, 1988; Kurr *et al.*, 1991). Since the wrapping of DNA in archaeal nucleosomes is sensitive to salt concentration (Musgrave *et al.*, 2000) and since negative supercoils are substituted for positive supercoils as the salt concentration is increased, we envisage that nucleosome flipping could be utilized by the cell to bring about a rapid change in the topology of genomic DNA when cells are required to adapt to different growth temperatures. The ability of both archaeal and eucaryal nucleosomes to wrap DNA in both positive and negative supercoils is also another indicator of the shared evolutionary history of the Euryarchaeota and eukaryotes. Nucleosome flipping of the eucaryal H3/H4 core tetramer also suggests that the evolution of the eucaryal nucleosome, by the addition of flanking H2A/H2B dimers on each side, has stabilized the octamer in a negative orientation. This may be required to limit its structural flexibility, given the seemingly unlimited posttranslational modifications that are possible for eucaryal histone proteins and the

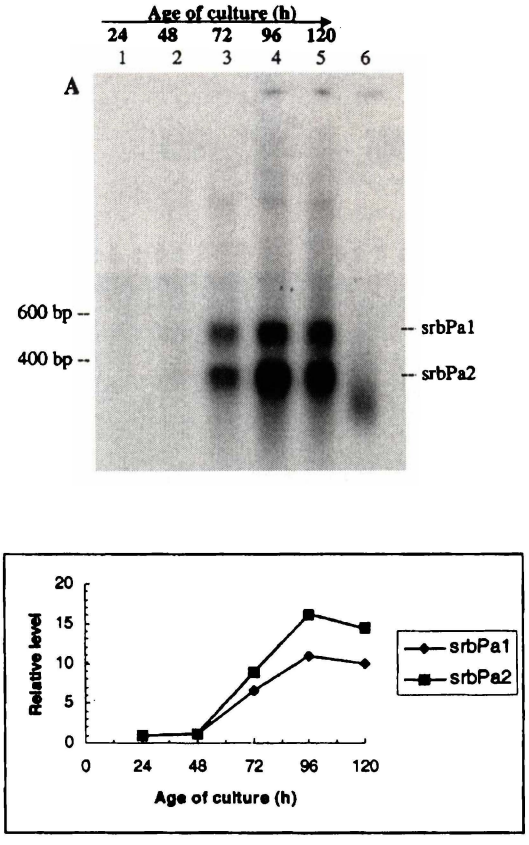


FIG. 5. **A:** Northern hybridization of total RNA from *P. abyssi* cells after various periods of growth, with the *srbPa1* gene probe. Lanes 1–5 each contain 10 μg of total RNA corresponding from cells of the ages indicated. Lane 6 contains 1 ng of *srbPa1* PCR product as a positive control. **B:** Relative levels of *srbPa1* and *srbPa2* mRNA to the corresponding 23S rRNA levels versus their culture ages. The mRNA levels were determined as described in Fig. 3.

fact that topoisomerases capable of increasing the superhelicity of DNA have not been found in the Eucarya.

DISCUSSION

RG is a type I DNA topoisomerase that has the unique ability to use ATP to drive the production of positive supercoils. It is proposed that the positive supercoils produced by RG result from the tracking of the enzyme along the DNA in a helicase-like fashion and the consequent removal of negative supercoils by its topoismerase I-like activity, resulting in the production of net positive supercoils in the DNA (Declais *et al.*, 2000). This view of RG is influenced by its low ATP requirement, and by the observation that the two domains can function separately (Collin *et al.*, 1988; Collin, 1990; Declais *et al.*, 2000). RG has been at the center of many "hot origin" hypotheses for the origin of cellular life because it is, to date, the only protein that is exclusively found in hyperthermophilic species and is thus seen to be essential to life at high temperature. However, since RG is considered to be a modern enzyme, constructed by the fusion of a topoisomerase I and a DNA helicase (Declais *et al.*, 2000), it has been proposed that life could not have evolved at the temperatures inhabited by hyperthermophilic microorganisms. In "lower temperature" origin hypotheses, cellular life is proposed to have evolved both up temperature as well as down temperature (Forterre, 1995). However, because of the lack of a robust genetic system in hyperthermophilic organisms, it has yet to be proven that RG is essential for life at all, let alone essential for hyperthermophily. Also, positively supercoiled DNA could be produced in a number of other ways in the absence of RG. Firstly, the negative wrapping of DNA by archaeal histones, as seen in Fig. 6, produces positive supercoils in nucleosome-free DNA. Secondly, positive supercoils can be produced by the removal of negative supercoils by a conventional type I topoisomerase. Thirdly, as suggested here, positive wrapping could also be the result of nucleosome flipping caused by an increase in intracellular salt concentrations or the selective removal of negative supercoils by a conventional type I topoisomerase following the tracking of a complex along the DNA. These methods could have been used to stabilize DNA molecules in the early evolution of cellular life at temperatures currently inhabited by hyperthermophiles in the absence of RG. RG is thus proposed to be a modern solution to the DNA stability problem in hyperthermophiles. In addition to modulating topology and the production of inherently stable positive supercoils in nucleosome-free DNA, the wrapping of DNA in nucleosomes has a significant effect on the temperature stability of double-stranded DNA in its own right. The melting temperature of DNA is increased by 20°C by the formation of HTz histone nucleosomes *in vitro* (Ronimus and Musgrave, 1996b). The melting temperature of DNA can also be increased significantly by charge neutralization resulting from an increase in the ionic strength of the cytoplasm or an increase in the concentration of basic proteins, peptides, or other

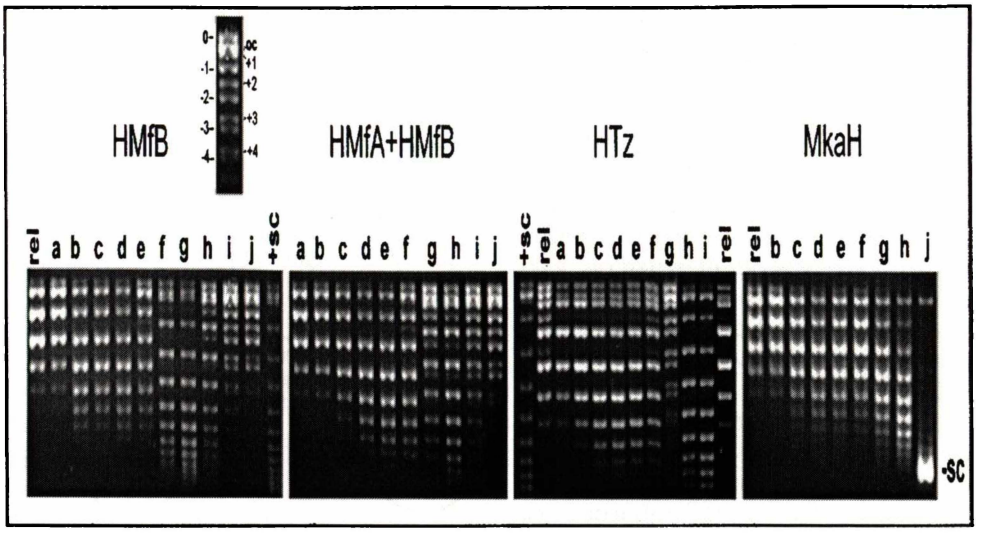


FIG. 6. DNA topology assays of the archaeal histones: recombinant HMfB homodimer from *M. fervidus*, recombinant HMfA/B heterodimers, HTz native heterodimer from *T. zilligii*, and native MkaH from *M. kandleri*. pUC18 (150 ng) relaxed at 70°C in 150 m*M* potassium glutamate was incubated in these conditions with archaeal histone proteins at protein/DNA mass ratios of (a) 0:1, (b) 0.1:1, (c) 0.15:1, (d) 0.3:1, (e) 0.6:1, (f) 0.8:1, (g) 1:1, (h) 1.5:1, and (i) 2:1. Topoisomerase V (20 U) was added, and the incubation was continued for 15 min. Reaction mixes were incubated with 1% sodium dodecyl sulfate for 2 min at 90°C and treated with 25 μg/ml proteinase K for 30 min at 40°C. Topoisomers were separated by 1.5% agarose gel electrophoresis at 1 V/cm for 16 h with buffer circulation. Lanes labeled "rel" are untreated substrate molecules, and those labeled +SC are pUC18 positively supercoiled topoisomer controls.

positively charged species. However, considerations of the *in vivo* stability of DNA must be considered with some caution, as its cellular concentration cannot be approached *in vitro*. For example, a bacterial chromosome must be condensed 1,000-fold to fit inside the cell, resulting in a cellular concentration of DNA of between 14 and 34 μg/mL (Bohrmann *et al.*, 1991).

The ability of archaeal histones to wrap DNA in two alternate topologies, positive and negative toroidal supercoils, allows archaeal nucleosomes to be structurally dynamic. Initially we suggested that positive toroidal supercoils might be necessary to counteract the affect that cytoplasms containing high salt might have on the cellular DNA in hyperthermophiles (Musgrave *et al.*, 1991, 1992). We reasoned that positive wrapping would result in an increase in the negative superhelicity of free DNA and that this wrapping would have the opposite effect to that exerted by intracellular salt. At low salt concentrations all archaeal histones wrap DNA in positive toroidal supercoils. However, as the salt concentration is increased, the wrapping becomes both positive and negative in the same reaction and possibly on the same molecule. This is evidenced by the positively supercoiled molecules produced at low protein/DNA mass ratios and negatively supercoiled molecules produced at higher protein/DNA mass ratios in the topology assay shown in Fig. 6. At 150 m*M* potassium glutamate a switch from positive to negative wrapping occurs with a very small increase in added protein. We suggest here that this structural plasticity could be used by the cell to modulate the topology of chromosomal DNA in a rapid and energetically efficient way. We propose that this "nucleosome flipping" could allow Archaea to rapidly adjust the topology of their genome in response to environmental changes. Because Archaea have been shown to increase their cytoplasmic salt concentrations in response to temperature increases (Hensel and Konig, 1988; Kurr *et al.*, 1991) we propose here that nucleosome flipping could also result from such an increase. The result of this would be to switch the toroidal wrapping in the nucleosomes from positive to negative as the temperature increased. This would cause the free DNA to become more positively supercoiled, a result that is similar to the action of RG, and consequently the free DNA would be less prone to denaturation. An example of a drastic topologi-

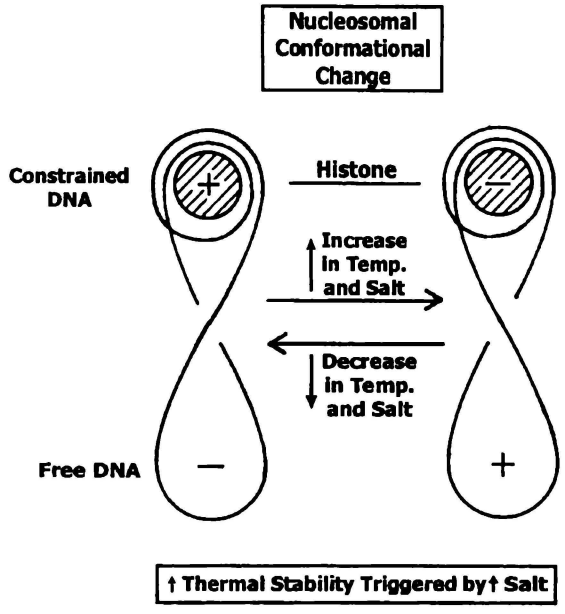


FIG. 7. Topology sensing via protein tracking-induced conformational change of the DNA in archaeal nucleosomes. A model is presented to propose that archaeal nucleosome conformational changes could modulate the topological changes that occur in DNA as a result of the tracking of a protein complex. Archaeal nucleosomes, depicted as half-shaded circles, can wrap the DNA in positive and negative toroidal supercoils. Negative and positive plectonemic supercoils are produced in the free DNA by protein tracking along the DNA. Conformational changes in the mode of wrapping of the DNA by the archaeal nucleosomes, from positive to negative on the left and negative to positive on the right, would negate the plectonemic supercoils in the free DNA.

cal change brought about by a change in DNA packaging is the positive wrapping of the DNA found in the spores of *Bacillus subtilis*. Positive wrapping, as opposed to the negative wrapping of bacterial chromosomal proteins found in vegetative cells, is induced by the small acid-soluble spore proteins (SASPs). This wrapping is suggested to account for the resistance of spore DNA to irradiation by UV light because SASP/DNA complexes *in vitro* prevent the production of DNA-damaging thymine dimers (Nicholson *et al.*, 1991; Griffith *et al.*, 1994).

Another potential function of nucleosome flipping could be to compensate for changes in genome topology that accompany DNA replication and transcription as described in Fig. 7. Transcriptionally mediated supercoiling occurs because the mass of the RNA transcript with its attached ribosomes is too great to be rotated around the DNA helix during transcription. Therefore, for transcription to proceed the DNA must effectively rotate, resulting in the production of supercoils in the DNA (Liu and Wang, 1987; Wu *et al.*, 1988). Because DNA strands are not broken, the positive supercoils produced ahead of RNA polymerase result in an equal number of negative supercoils behind the complex. A change in the toroidal supercoiling in archaeal nucleosomes from negative to positive would decrease the positive superhelicity ahead of tracking complex, and *vice versa*. With eucaryal nucleosomes it has been proposed that the positive supercoils produced ahead of RNA polymerase could help to remove nucleosomes form the DNA, leading to more efficient elongation of the transcript. This model is based on the instability of nucleosomes on positively supercoiled DNA and thermodynamics suggesting that formation of negatively wrapped nucleosomes on negatively supercoiled DNA is favored because of the effective reduction in net negative supercoils. In Eucarya, the dynamic nature of the nucleosome is modulated by remodeling. However, it appears that archaeal nucleosomes are not subject to the same remodeling because archaeal histones do not have the required N-terminal tails, and, in *T. zilligii*, histones with altered mass have not been isolated (Dinger *et al.*, 2000).

In addition to the role that the wrapping of archaeal chromosomes in nucleosomes might play in allowing cellular life at high temperatures, we have described a stationary-phase stress response protein that has a phylogenetic profile that mimics that seen for RG (Dinger *et al.*, 2000). This gene, named *srb*, is found in the genomes of both archaeal and bacterial thermophiles. We previously detected SRB protein as a stationary phase-induced acid-soluble protein in the hyperthermophile *T. zilligii* (Dinger *et al.*, 2000). In accord with this, we show here that transcription of the *srbPa* paralogues is also stationary phase-induced in *P. abyssi*, another hyperthermophilic member of the Thermococcales. SRB protein has been crystallized from *Methanobacterium thermoautotrophicum*, and the structure has been used to search for structural homologues. Although highly significant matches were not made, the closest structural homologues all have roles in nucleic acid metabolism or processing, a role that agrees with a GR repeat found in the C-terminus of some SRB proteins (Cort *et al.*, 2000). In addition to testing the transcription of the *srb* genes in stationary-

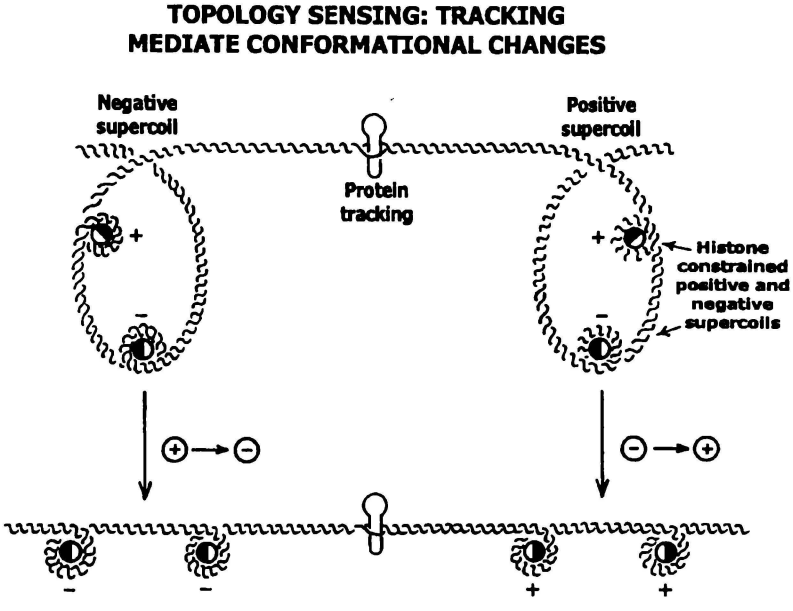


FIG. 8. Temperature sensing via a salt-induced conformational change of the DNA in archaeal nucleosomes. A model is presented to propose that archaeal nucleosome conformational changes could modulate the topological changes that occur in DNA as a result of a change in intracellular salt concentrations induced by a change in temperature of the environment, or another physical parameter that would influence the topology of genomic DNA. Archaeal nucleosomes, depicted as hatched circles labeled (+) or (−), can wrap the DNA in positive and negative toroidal supercoils, respectively. An increase in temperature, salt, or both would cause archaeal nucleosomes to "flip" from a positive toroidal supercoil to a negative toroidal supercoil, and *vice versa*. An increase in temperature would therefore result in the free DNA more positively supercoiled, a response similar to that effected by the topoisomerase RG.

phase cells we have shown that there are contrasting transcription patterns of the histone paralogues in *T. zilligii* and *P. abyssi*. We previously reported that the histone proteins were rapidly depleted from cells in the stationary phase in *T. zilligii* and that the level of the *htz1* gene transcript was decreased as cells entered the stationary phase and was not able to be detected by northern hybridization using stationary-phase cells (Dinger *et al.*, 2000). In this study we show that, in contrast with the similar stationary-phase induction of expression of the *srb* genes, the histone genes (*hpa1* and *hpa2*) are not transcribed in such an obvious stationary phase-repressed manner in *P. abyssi* as was shown for *htz1* and *htz2* (Zhang, 2001). Consistent with this result, the levels of the *sre* gene transcript, which may encode a histone replacement protein, is likewise not stationary phase-induced in *P. abyssi*. From these results we conclude that the nucleiod and hence the topology of genomic DNA is dynamic in the Archaea. Therefore the mechanisms for genome packaging should be considered along with the function of topoisomerases in discussing the role that the physical and genetic stability of the genome plays in cellular life and the evolution of that life. We also conclude that the systems induced to allow Archaea to survive stationary phase show significant diversity given that *T. zilligii* and *P. abyssi* are members of the same archaeal family. This should promote a wider examination of these mechanisms, particularly since the recently described archaeal symbiont *NanoArchaeaum equitans*, whose 16S rRNA sequence suggests it belongs to a new phylum, are found associated predominantly with stationary-phase cells (Huber *et al.*, 2002).

ABBREVIATIONS

MOPS, sodium 3-(*N*-morpholino)propanesulfonic acid; PCR, polymerase chain reaction; RG, reverse gyrase; SASP, small acid-soluble spore protein; SSC, saline–sodium citrate.

REFERENCES

Anderson, P. and Bauer, W. (1978) Supercoiling in closed circular DNA: dependence upon ion type and concentration. *Biochemistry* 17, 594–601.

Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A., and Struhl, K. (1998) *Current Protocols in Molecular Biology*, John Wiley & Sons, New York.

Bohrmann, B., Villiger, W., Johansen, R., and Kellenberger, E. (1991) Coralline shape of the bacterial nucleoid after cryofixation. *J. Bacteriol.* 173, 3149–3158.

Collin, R.G. (1990) Reverse gyrase from Archaebacteria [Ph.D. Thesis]. University of Waikato, Hamilton, New Zealand.

Collin, R.G., Morgan, H.W., Musgrave, D.R., and Daniel, R.M. (1988) Distribution of reverse gyrase in representative species of eubacteria and Archaeabacteria. *FEMS Microbiol. Lett.* 55, 235–240.

Corliss, J.B., Dymond, J., and Gordo, L.I. (1979) Submarine thermal springs on the Galapagos Rift. *Science* 203, 1073–1078.

Cort, J.R., Yee, A., Edwards, A.M., Arrowsmith, C.H., and Kennedy, M.A. (2000) NMR structure determination and structure-based functional characterization of conserved hypothetical protein MTH1175 from *Methanobacterium thermoautotrophicum*. *J. Struct. Funct. Genom.* 1, 15–25.

Declais, A.C., Marsault, J., Confalonieri, F., de La Tour, C.B., and Duguet, M. (2000) Reverse gyrase, the two domains intimately co-operate to promote positive supercoiling. *J. Biol. Chem.* 275, 19498–19504.

Dinger, M. (1998) Histones and stationary phase of the Thermophilic archaeon *Thermococcus zilligii* [M.Sc. Thesis]. University of Waikato, Hamilton, New Zealand.

Dinger, M.E. and Musgrave, D.R. (2000) Identification of archaeal genes encoding a novel stationary phase-response protein. *Biochim. Biophys. Acta* 1490, 115–120.

Dinger, M.E., Baillie, G.J., and Musgrave, D.R. (2000) Growth phase-dependent expression and degradation of histones in the thermophilic Archaeaon *Thermococcus zilligii*. *Mol. Microbiol.* 36, 876–885.

Doolittle, W.F. (1999) Phylogenetic classification and the universal tree. *Science* 284, 2124–2129.

Erauso, G., Reysenbach, A.L., Godfroy, A., Meunier, J.R., Crump, B., Partensky, F., Baross, J.A., Marteinsson, V., Barbier, G., Pace, N.R., and Prieur, D. (1993) *Pyrococcus abyssi* sp. Nov., a new hyperthermophilic archaeon isolated from a deep-sea hydrothermal vent. *Arch. Microbiol.* 160, 338–349.

Fahrner, R.L., Cascio, D., Lake, J.A., and Slesarev, A. (2001) An ancestral nuclear protein assembly: crystal structure of the *Methanopyrus kandleri* histone. *Protein Sci.* 10, 2002–2007.

Forterre, P. (1995) Thermoreduction, a hypothesis for the origin of prokaryotes. *C. R. Acad. Sci. III* 318, 415–422.

Forterre, P. and Phillipe, H. (1999) Where is the root of the universal tree of life? *Bioessays* 21, 871–879.

Gerard, E., Jolivet, E., Pieur, D., and Forterre, P. (2001) DNA protection mechanisms are not involved in the radioresistance of the hyperthermophilic Archaea *Pyrococcus abyssi* and *P. furiosus*. *Mol. Genet. Genomics* 266, 72–78.

Griffith, J., Makhov, A., Santiago-Lara, L., and Setlow, P. (1994) Electron microscopic studies of the interaction between a *Bacillus subtilis* alpha/beta-type small, acid soluble spore protein with DNA: protein binding is cooperative, stiffens the DNA and induces negative supercoiling. *Proc. Natl. Acad. Sci. USA* 91, 8224–8228.

Hamiche, A. and Richard-Foy, H. (1998) The switch in helical handedness of the histone (H3-H4)2 tetramer within a nucleoprotein particle requires a re-orientation of the H3-H3 interface. *J. Biol. Chem.* 273, 9261–9268.

Hamiche, A., Carot, V., Alilat, M., De Lucia, F., O'Donohue, M.-F., Révet, B., and Prunell, A. (1996) Interaction of the histone (H3-H4)2 tetramer of the nucleosome with positively supercoiled DNA minicircles: potential flipping of the protein from a left- to a right-handed superhelical form. *Proc. Natl. Acad. Sci. USA* 93, 7588–7593.

Hensel, R. and König, H. (1988) Thermoadaptation of methanogenic bacteria by intracellular ion concentration. *FEMS Microbiol. Lett.* 49, 75–79.

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

Jenuwein, T. and Allis, C.D. (2001) Translating the histone code. *Science* 293, 1074–1080.

Kikuchi, A. and Asai, K. (1984) Reverse gyrase: a topoisomerase which introduces positive superhelical turns into DNA. *Nature* 309, 677–681.

Klages, K.U. and Morgan, H.W. (1994) Characterization of an extremely thermophilic sulfur-metabolizing archaebacterium belong to the thermococcales. *Arch. Microbiol.* 162, 261–266.

Koch, A.L. (1994) Growth measurement. In *Methods for General and Molecular Bacteriology*, edited by P. Gerhard, R.G.E. Murray, W.A. Woord, and N.R. Krieg, ASM Press, Washington, DC, pp. 248–277.

Kurr, M., Huber, H., Jannasch, H.W., Fricke, H., and Tricone, A. (1991) *Methanopyrus kandleri*, gen sp Nov. represents a novel group of hyperthermophilic methanogens, growing at 110°C. *Arch. Microbiol.* 156, 239–242.

Liu, L.F. and Wang, J.C. (1987) Supercoiling of the DNA template during transcription. *Proc. Natl. Acad. Sci. USA* 84, 7024–7027.

Lopez-Garcia, P. and Forterre, P. (2000) DNA topology and the thermal stress response, a tale from mesophiles and hyperthermophiles. *Bioessays* 22, 738–746.

Musgrave, D.R., Sandman, K., and Reeve, J.N. (1991) DNA binding by the archaeal histone HMf results in positive supercoiling. *Proc. Natl. Acad. Sci. USA* 88, 10397–10401.

Musgrave, D.R., Sandman, K., Stroup, D., and Reeve, J.N. (1992) DNA binding proteins and genome topology in thermophilic prokaryotes. In *Biocatalysis at Extreme Temperatures*, edited by M.W.W. Adams and R.M. Kelly, American Chemical Society, Washington, DC, pp. 174–188.

Musgrave, D.R., Forterre, P., and Slesarev, A. (2000) Negative constrained supercoiling in archaeal nucleosomes. *Mol. Microbiol.* 35, 341–349.

Nelson, K.E., Clayton, R.A., Gill, S.R., Gwinn, M.L., Haft, D.H., Hickey, E.K., Peterson, J.D., Nelson, W.C., Ketchum, K.A., McDonald, L., Utterback, T.R., Malek, J.A., Linher, K.D., Garret, M.M., Stewart, A.M., Cotton, M.D., Pratt, M.S., Phillips, C.A., Richardson, D., Heidelberg, J., Sutton, G.G., Fleischmann, R.D., Eisen, J.A., White, O., Salzberg, S.L., Smith, H.O., Venter, J.C., and Fraser, C.M. (1999) Evidence for lateral gene transfer between Archaea and Bacteria from genome sequence of *Thermotoga maritima*. *Nature* 399, 323–329.

Nicholson, W.L., Setlow, B., and Setlow, P. (1991) Ultraviolet radiation of DNA complexed with alpha/beta-type small, acid soluble spore proteins from spores of *Bacillus* or *Clostridium* species make spore photoproduct but not thymine dimers. *Proc. Natl. Acad. Sci. USA* 88, 8288–8292.

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

Peak, M.J., Robb, F.T., and Peak, J.G. (1995) Extreme resistance to thermally induced DNA backbone breaks in the hyperthermophilic Archaeaon *Pyrococcus furiosus*. *J. Bacteriol.* 177, 6316–6318.

Pugh, B.F., Schutte, B.C., and Cox, M.M. (1989) Extent of duplex underwinding induced by recA protein binding in the presence of ATP. *J. Mol. Biol.* 205, 487–492.

Reanney, D.C. (1974) On the origin of prokaryotes. *J. Theor. Biol.* 48, 243–251.

Rivera, M.C., Jain, R., Moore, J.E., and Lake, J.A. (1998) Genomic evidence for two functionally distinct gene classes. *Proc. Natl. Acad. Sci. USA* 95, 6239–6244.

Ronimus, R.S. and Musgrave, D.R. (1996a) A gene, han1A encoding an archaeal histone-like protein from the *Thermococcus* species AN1: homology with eukaryal histone consensus sequence and the implications for delineation of the histone fold. *Biochim. Biophys. Acta* 1307, 1–7.

Ronimus, R.S. and Musgrave, D.R. (1996b) Purification and characterisation of a histone-like protein from the archaeal isolate AN1, a member of the *Thermococcales*. *Mol. Microbiol.* 20, 77–86.

Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989) *Molecular Cloning*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Sandman, K. and Reeve, J.N. (2001) Chromosome packaging by archaeal histones. *Adv. Appl. Microbiol.* 50, 75–99.

Sandman, K., Krzycki J.A., Dobrinski, B., Lurz, R., and Reeve, J.N. (1990) HMf, a DNA-binding protein isolated from the hyperthermophilic archaeon *Methanothermus fervidus*, is most closely related to histones. *Proc. Natl. Acad. Sci. USA* 87, 5788–5791.

Santisteban, M.S., Kalashnikova, T., and Smith, M.M. (2000) Histone H2A.Z regulates transcription and is partially redundant with nucleosome remodelling complexes. *Cell* 103, 411–422.

Woese, C.R. (1987) Bacterial evolution. *Microbiol. Rev.* 51, 221–271.

Woese, C.R. (1998) The universal ancestor. *Proc. Natl. Acad. Sci. USA* 95, 6854–6859.

Woese, C.R., Kandler, O., and. Wheelis, M.L. (1990) Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proc. Natl. Acad. Sci. USA* 87, 4576–4579.

Wu, J.S. and Grunstein, M. (2000) 25 years after the nucleosome model: chromatin modifications. *Trends Biochem. Sci.* 25, 619–623.

Wu, H.Y., Shyy, S.H., Wang, J.C., and Liu, L.F. (1988) Transcription generates positively and negatively supercoiled domains in the template. *Cell* 53, 433–440.

Zhang, X. (2001) Examination of the stationary phase response in the EuryArchaea [M.Phil. Thesis]. University of Waikato, Hamilton, New Zealand.

Address reprint requests to:
Dr. David Musgrave
Department of Biological Sciences
University of Waikato
P.O. Box 3105
Hamilton, New Zealand

E-mail: musgrave@waikato.ac.nz

EFFECTS OF STATIONARY PHASE IN THE THERMOPHILIC ARCHAEON *THERMOCOCCUS ZILLIGII* ON GLOBAL PROTEIN PATTERNS AND EXPRESSION OF HISTONES

Marcel DINGER, David MUSGRAVE.
Department of Biological Sciences, The University of Waikato, Hamilton, New Zealand.

A feature common to all known microorganisms from each of the three domains of life, is their ability to respond to stress conditions, such as nutrient deprivation, by differentiating into a specialised state to maximise their survival chances (1). This division of the growth cycle in microorganisms, referred to as the stationary phase, is an essential component of their "feast and famine" lifestyle in the environment, when they are typically exposed to nutrients relatively briefly and have to endure long periods in extremely limiting conditions (1). Although, the stationary phase response has been studied extensively in members of the Bacteria and the Eucarya, practically no investigation has been undertaken in members of the Archaea. Histones have been identified in several members of the Archaea. These archaeal histones have primary sequences similar to the eukaryal histone fold domain, and are thought to resemble the archetypal ancestor of the eukaryal nucleosome core histones (2). There is preliminary evidence that the synthesis of the archaeal histones is dependent on growth phase. The overall level and the proportion of individual monomers changes with the onset of stationary phase (3). Moreover, recent studies have indicated that archaeal histones do not associate with certain sequences and it was suggested that histones may regulate the level of gene expression (2). Given the fundamental regulatory and packaging role of the histones in Eucarya and evidence that both the mode of interaction with DNA and the potential specificity of packaging of archaeal histones can vary, an investigation into the effects of growth phase on global protein patterns and the expression of histones, in the thermophilic archaeon, *Thermococcus zilligii*, was begun. In this paper, total cellular protein extracts were isolated from *T. zilligii* cultures at various stages of growth from mid-logarithmic to late-stationary phase, then analysed and compared. The levels of the HTz protein and the *htz1* transcript were examined.

Keywords: Archaea, Stationary phase, Archaeal histones.

(1) Kolter, R., Siegele, D.A., and Tormo, A. (1993). *Annu. Rev. Microbiol.*, **47**, 855-874.
(2) Pereira, S.L., Grayling, R.A., Lurz, R., and Reeve, J.N. (1997). *Proc. Natl. Acad. Sci. USA*, 94, 12633-12637.
(3) Sandman, K., Grayling, R.A. Dobrinski, B., Lurz, R. and Reeve, J.N. (1994). *Proc. Natl. Acad. Sci. USA*, **91**, 12624-12628.

IN SEARCH OF NEW REGULATORY PROTEINS FOR THE CONTROL OF GENE EXPRESSION IN ARCHAEA

Marcel Dinger and David Musgrave.
Department of Biological Sciences, The University of Waikato, Hamilton, New Zealand.

In their natural environment, microorganisms must endure multiple stresses, including changes in temperature, osmolarity and nutrient availability. Microorganisms respond to these stresses by undergoing adaptive changes in gene expression, resulting in the consequent physiological and morphological differentiation of the cells. One mechanism for the sometimes global changes in gene expression invoked by stress is through alterations in the topology of the organism's genome. The DNA topology may be mediated directly by the environmental stimuli itself, or indirectly via changes in the levels or activity of DNA remodelling proteins.

Recently, we have found that the levels of histones in the thermophilic archaeon Thermococcus zilligii dramatically decreased as cells entered stationary phase and could not be detected by late stationary phase. In addition, several other proteins increased in level as the cultures aged. Due to the important role that histones play in the compaction of DNA, it was suggested that some of these proteins may takeover the role of DNA compaction in the stationary phase and influence gene expression. Here we present the sequence analysis, phylogenetic distribution, and preliminary characterisation of three novel stationary phase-induced proteins.

UNDERSTANDING STRESS: THE IDENTIFICATION OF NOVEL STATIONARY-PHASE PROTEINS IN ARCHAEA

Marcel Dinger and David Musgrave.
Department of Biological Sciences, The University of Waikato, Hamilton, New Zealand.

In their natural environment, microorganisms must endure multiple stresses, including changes in temperature, osmolarity and nutrient availability. Microorganisms respond to these stresses by undergoing adaptive changes in gene expression, resulting in physiological and morphological differentiation of the cells.

To further understand the microbial stress-response, we examined the stationary-phase response in the thermophilic archaeon, *Thermococcus zilligii*. Thermophilic Archaea are considered to have characteristics that are most similar to the earliest forms of life on Earth. A deeper understanding of these archetypal stress-response systems will provide insights into the evolution of these complex regulatory systems, and the survival mechanisms of microorganisms in the environment.

Here we report the identification, partial characterisation and possible functions of three novel proteins whose synthesis was induced in stationary-phase. Orthologues of the corresponding genes for these proteins were present in many other phylogenetically disparate species. Elucidating the role of these hitherto unstudied proteins is likely to be very important to the understanding of the microbial stress-response.

UNRAVELLING REGULATION: USING GENOME SEQUENCES TO UNDERSTAND THE ORIGINS OF GENE EXPRESSION CONTROL SYSTEMS

Marcel Dinger and David Musgrave.
Department of Biological Sciences, The University of Waikato, Hamilton, New Zealand.

The discovery that a relatively coherent phylogeny of all life on Earth can be constructed using ribosomal RNA sequences has stimulated renewed interest in the origins of life's most fundamental processes. The availability of numerous complete microbial genome sequences has now made it feasible to interpolate our molecular history and better understand the key events and processes that drove our evolutionary past.

One of the fundamental processes that is critical to all life on Earth today is the regulation of gene expression. In higher organisms, the control of gene expression provides the basis for multicellularity. In microorganisms, the regulation of gene expression allows the cell to respond to changes in its environment, such as nutrient availability, toxin levels, temperature or osmolarity. Since any living system is constantly undergoing change, the development of gene expression controls systems must have been critical for the evolution of increasingly sophisticated life forms.

To broaden the understanding of the origins of gene expression control systems, we have identified three novel stress-response proteins in the thermophilic archaeon, *Thermococcus zilligii*. Using the available genome sequences, we have identified homologues of these proteins in other organisms. By utilising a variety of computational approaches we have constructed a dossier on each of these gene families and provide suggestions for their ancestry and possible roles in the control of gene expression.

References

Aki, T. and Adhya, S. (1997). Repressor induced site-specific binding of HU for transcriptional regulation. *Embo J* **16:** 3666-74.

Ali Azam, T., Iwata, A., Nishimura, A., Ueda, S. and Ishihama, A. (1999). Growth phase-dependent variation in protein composition of the *Escherichia coli* nucleoid. *J Bacteriol* **181:** 6361-70.

Almiròn, M., Link, A., Furlong, D. and Kolter, R. (1992). A novel DNA binding protein with regulatory and protective roles in starved *Escherichia coli*. *Genes Dev* **170:** 5169-5170.

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25:** 3389-402.

Anderson, W.F., Ohlendorf, D.H., Takeda, Y. and Matthews, B.W. (1981). Structure of the cro repressor from bacteriophage lambda and its interaction with DNA. *Nature* **290:** 754-8.

Anglister, J., Grzesiek, S., Ren, H., Klee, C.B. and Bax, A. (1993). Isotope-edited multidimensional NMR of calcineurin B in the presence of the non-deuterated detergent CHAPS. *J Biomol NMR* **3:** 121-6.

Aravind, L. and Koonin, E.V. (1999). DNA-binding proteins and evolution of transcription regulation in the Archaea. *Nucleic Acids Res* **27:** 4658-70.

Arents, G. and Moudrianakis, E.N. (1995). The histone fold: a ubiquitous architectural motif utilized in DNA compaction and protein dimerization. *Proc Natl Acad Sci U S A* **92:** 11170-11174.

Attwood, J.T., Yung, R.L. and Richardson, B.C. (2002). DNA methylation and the regulation of gene transcription. *Cell Mol Life Sci* **59:** 241-57.

Bahloul, A., Boubrik, F. and Rouviere-Yaniv, J. (2001). Roles of *Escherichia coli* histone-like protein HU in DNA replication: HU-beta suppresses the thermosensitivity of dnaA46ts. *Biochimie* **83:** 219-29.

Baillie, G.J. (1994). Examination of the stationary phase response of archaeal isolate AN1. MSc thesis, The University of Waikato, Hamilton, New Zealand. 122 pp.

Balandina, A., Claret, L., Hengge-Aronis, R. and Rouviere-Yaniv, J. (2001). The *Escherichia coli* histone-like protein HU regulates rpoS translation. *Mol Microbiol* **39:** 1069-79.

Bates, A.D. and Maxwell, A. (1993). DNA supercoiling. In: *'In Focus' Series*. Rickwood, D. (ed). IRL Press, Oxford, pp. 17-45.

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

Bell, S.D. and Jackson, S.P. (2000). Mechanism of autoregulation by an archaeal transcriptional repressor. *J Biol Chem* **275:** 31624-9.

Bell, S.D., Magill, C.P. and Jackson, S.P. (2001). Basal and regulated transcription in Archaea. *Biochem Soc Trans* **29:** 392-5.

Benner, S.A. and Ellington, A.D. (1987). The last ribo-organism. *Nature* **329:** 295-6.

Benner, S.A., Ellington, A.D. and Tauer, A. (1989). Modern metabolism as a palimpsest of the RNA world. *Proc Natl Acad Sci U S A* **86:** 7054-8.

Berger, S.L. (2002). Histone modifications in transcriptional regulation. *Curr Opin Genet Dev* **12:** 142-8.

Bernander, R. (1998). Archaea and the cell cycle. *Mol Microbiol* **29:** 955-61.

Bertin, P., Lejeune, P., Laurent-Winter, C. and Danchin, A. (1990). Mutations in *bglY*, the structural gene for the DNA-binding protein H1, affect the expression of several *Escherichia coli* genes. *Biochimie* **72:** 889-891.

Best, S. and Speicher, D.W. (1996). Detection of proteins on blot membranes. In: *Current Protocols in Protein Science*. Coligan, J.E., Dunn, B.M., Ploegh, H.L., Speicher, D.W. and Wingfield, P.T. (eds). John Wiley & Sons, Inc., New York, pp. 10.8.1-10.8.7.

Bollag, D.M. and Edelstein, S.J. (1991). Protein concentration determination. In: *Protein Methods*. Wiley-Liss, Inc., New York, pp. 50-55.

Bonnefoy, E. and Rouviere-Yaniv, J. (1991). HU and IHF, two homologous histone-like proteins of *Escherichia coli*, form different protein-DNA complexes with short DNA fragments. *Embo J* **10:** 687-96.

Bouthier de la Tour, C., Portemer, C., Huber, R., Forterre, P. and Duguet, M. (1991). Reverse gyrase in thermophilic eubacteria. *J Bacteriol* **173:** 3921-3.

Bradford, M.M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem* **72:** 248-254.

Braun, P., Hu, Y., Shen, B., Halleck, A., Koundinya, M., Harlow, E. and LaBaer, J. (2002). Proteome-scale purification of human proteins from bacteria. *Proc Natl Acad Sci U S A* **99:** 2654-9.

Brinkman, A.B., Dahlke, I., Tuininga, J.E., Lammers, T., Dumay, V., de Heus, E., Lebbink, J.H., Thomm, M., de Vos, W.M. and van Der Oost, J. (2000). An Lrp-like transcriptional regulator from the archaeon *Pyrococcus furiosus* is negatively autoregulated. *J Biol Chem* **275:** 38160-9.

Brochier, C. and Philippe, H. (2002). Phylogeny: a non-hyperthermophilic ancestor for bacteria. *Nature* **417:** 244.

Brocks, J.J., Logan, G.A., Buick, R. and Summons, R.E. (1999). Archean molecular fossils and the early rise of eukaryotes. *Science* **285:** 1033-6.

Broyles, S. and Pettijohn, D.E. (1986). Interaction of the *Escherichia coli* HU protein with DNA: Evidence for formation of nucleosome-like structures with altered DNA helical pitch. *J Mol Biol* **187:** 47-60.

Bult, C., White, O., Olsen, G., Zhou, L., Fleischmann, R., Sutton, G., Blake, J., Fitzgerald, L., Clayton, R., Gocayne, J., Kerlavage, A., Dougherty, B., Tomb, J., Adams, M., Reich, C., Overbeek, R., Kirkness, E., Weinstock, K., Merrick, J., Glodek, A., Scott, J., Geoghagen, N., Weidman, J., Fuhrmann, J., Nguyen, D., Utterback, T., Kelley, J., Peterson, J., Sadow, P., Hanna, M., Cotton, M., Roberts, K., Hurst, M., Kaine, B., Borodovsky, M., Klenk, H., Fraser, C., Smith, H., Woese, C. and Venter, J. (1996). Complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschii*. *Science* **273:** 1058-1073.

Caetano-Anolles, G. (2002). Evolved RNA secondary structure and the rooting of the universal tree of life. *J Mol Evol* **54:** 333-45.

Calvo, J.M. and Matthews, R.G. (1994). The leucine-responsive regulatory protein, a global regulator of metabolism in *Escherichia coli*. *Microbiol Rev* **58:** 466-90.

Cam, E.L., Culard, F., Larquet, E., Delain, E. and Cognet, J.A. (1999). DNA bending induced by the archaebacterial histone-like protein MC1. *J Mol Biol* **285:** 1011-21.

Carey, J. (1991). Gel retardation. *Methods Enzymol* **208:** 103-17.

Carlin, A., Shi, W., Dey, S. and Rosen, B.P. (1995). The *ars* operon of *Escherichia coli* confers arsenical and antimonial resistance. *J Bacteriol* **177:** 981-6.

Castresana, J. (2001). Comparative genomics and bioenergetics. *Biochim Biophys Acta* **1506:** 147-62.

Cavalier-Smith, T. (2002). The neomuran origin of archaebacteria, the negibacterial root of the universal tree and bacterial megaclassification. *Int J Syst Evol Microbiol* **52:** 7-76.

Champoux, J.J. (2001). DNA topoisomerases: structure, function, and mechanism. *Annu Rev Biochem* **70:** 369-413.

Chartier, F., Laine, B. and Sautiere, P. (1988). Characterization of the chromosomal protein MC1 from the thermophilic archaebacterium *Methanosarcina* sp. CHTI 55 and its effect on the thermal stability of DNA. *Biochim Biophys Acta* **951:** 149-56.

Chatterjee, J., Miyamoto, C.M. and Meighen, E.A. (1996). Autoregulation of *luxR*: the *Vibrio harveyi* lux-operon activator functions as a repressor. *Mol Microbiol* **20:** 415-25.

Chavali, G.B., Vijayalakshmi, C. and Salunke, D.M. (2001). Analysis of sequence signature defining functional specificity and structural stability in helix-loop-helix proteins. *Proteins* **42:** 471-80.

Choli, T., Henning, P., Wittmann-Liebold, B. and Reinhardt, R. (1988). Isolation, characterization and microsequence analysis of a small basic methylated DNA-binding protein from the Archaebacterium, *Sulfolobus solfataricus*. *Biochim Biophys Acta* **950:** 193-203.

Christendat, D., Yee, A., Dharamsi, A., Kluger, Y., Savchenko, A., Cort, J.R., Booth, V., Mackereth, C.D., Saridakis, V., Ekiel, I., Kozlov, G., Maxwell, K.L., Wu, N., McIntosh, L.P., Gehring, K., Kennedy, M.A., Davidson, A.R., Pai, E.F., Gerstein, M., Edwards, A.M. and Arrowsmith, C.H. (2000). Structural proteomics of an archaeon. *Nat Struct Biol* **7:** 903-9.

Cook, W.R., de Boer, P.A. and Rothfield, L.I. (1989). Differentiation of the bacterial cell division site. *Int Rev Cytol* **118:** 1-31.

Cort, J.R., Yee, A., Edwards, A.M., Arrowsmith, C.H. and Kennedy, M.A. (2000). NMR structure determination and structure-based functional characterization of conserved hypothetical protein MTH1175 from *Methanobacterium thermoautotrophicum*. *J Struct Funct Genom* **1:** 15-25.

Crick, F.H. (1976). Linking numbers and nucleosomes. *Proc Natl Acad Sci U S A* **73:** 2639-43.

Csonka, L.N. (1989). Physiological and genetic responses of bacteria to osmotic stress. *Microbiol Rev* **53:** 121-147.

Cuff, J.A. and Barton, G.J. (1999). Evaluation and improvement of multiple sequence methods for protein secondary structure prediction. *Proteins* **34:** 508-19.

Cuff, J.A. and Barton, G.J. (2000). Application of multiple sequence alignment profiles to improve protein secondary structure prediction. *Proteins* **40:** 502-11.

Cuff, J.A., Clamp, M.E., Siddiqui, A.S., Finlay, M. and Barton, G.J. (1998). JPred: a consensus secondary structure prediction server. *Bioinformatics* **14:** 892-3.

D'Ari, R., Lin, R.T. and Newman, E.B. (1993). The leucine-responsive regulatory protein: more than a regulator? *Trends Biochem Sci* **18:** 260-3.

de Boer, P.A., Cook, W.R. and Rothfield, L.I. (1990). Bacterial cell division. *Annu Rev Genet* **24:** 249-74.

de Boer, P.A., Crossley, R.E., Hand, A.R. and Rothfield, L.I. (1991). The MinD protein is a membrane ATPase required for the correct placement of the *Escherichia coli* division site. *Embo J* **10:** 4371-80.

Deckert, G., Warren, P.V., Gaasterland, T., Young, W.G., Lenox, A.L., Graham, D.E., Overbeek, R., Snead, M.A., Keller, M., Aujay, M., Huber, R., Feldman, R.A., Short, J.M., Olsen, G.J. and Swanson, R.V. (1998). The complete genome sequence of the hyperthermophilic bacterium *Aquifex aeolicus*. *Nature* **392:** 353-358.

DeLange, R.J., Green, G.R. and Searcy, D.G. (1981). A histone-like protein (HTa) from *Thermoplasma acidophilum*. I. Purification and properties. *J Biol Chem* **256:** 900-4.

Delihas, N. (1995). Regulation of gene expression by trans-encoded antisense RNAs. *Mol Microbiol* **15:** 411-4.

Delihas, N. and Forst, S. (2001). MicF: an antisense RNA gene involved in response of *Escherichia coli* to global stress factors. *J Mol Biol* **313:** 1-12.

DelVecchio, V.G., Kapatral, V., Redkar, R.J., Patra, G., Mujer, C., Los, T., Ivanova, N., Anderson, I., Bhattacharyya, A., Lykidis, A., Reznik, G., Jablonski, L., Larsen, N., D'Souza, M., Bernal, A., Mazur, M., Goltsman, E., Selkov, E., Elzer, P.H., Hagius, S., O'Callaghan, D., Letesson, J.J., Haselkorn, R., Kyrpides, N. and Overbeek, R. (2002). The genome sequence of the facultative intracellular pathogen *Brucella melitensis*. *Proc Natl Acad Sci U S A* **99:** 443-8.

Depew, D.E. and Wang, J.C. (1975). Conformational fluctuations of DNA helix. *Proc Natl Acad Sci U S A* **72:** 4275-9.

DeSilva, T.M., Ursitti, J.A. and Speicher, D.W. (1996). Protein detection in gels using fixation. In: *Current Protocols in Protein Science*. Coligan, J.E., Dunn, B.M., Ploegh, H.L., Speicher, D.W. and Wingfield, P.T. (eds). John Wiley & Sons, Inc., New York, pp. 10.5.1-10.5.12.

Dijk, J. and Reinhardt, R. (1986). The structure of DNA-binding proteins from Eu- and Archaebacteria. In: *Bacterial Chromatin*. Gualerzi, C.O. and Pon, C.L. (eds). Springer-Verlag, Heidelberg, pp. 185-218.

Dinger, M.E. (1998). Histones and Stationary Phase of the Thermophilic Archaeon *Thermococcus zilligii*. MSc thesis, The University of Waikato, Hamilton. 155 pp.

Dinger, M.E., Baillie, G.J. and Musgrave, D.R. (2000). Growth phase-dependent expression and degradation of histones in the thermophilic archaeon *Thermococcus zilligii*. *Mol Microbiol* **36:** 876-85.

Dinger, M.E. and Musgrave, D.R. (2000). Identification of archaeal genes encoding a novel stationary phase-response protein. *Biochim Biophys Acta* **1490:** 115-20.

Dixon, N.E. and Kornberg, A. (1984). Protein HU in the enzymatic replication of the chromosomal origin of *Escherichia coli*. *Proc Natl Acad Sci U S A* **81:** 424-8.

Doolittle, R.F. (2000). Searching for the common ancestor. *Res Microbiol* **151:** 85-9.

Dorman, C.J. (1996). Flexible response: DNA supercoiling, transcription and bacterial adaptation to environmental stress. *Trends Microbiol* **4:** 214-6.

Drlica, K. (1992). Control of bacterial DNA supercoiling. *Mol Microbiol* **6:** 425-433.

Drlica, K. and Rouvière-Yaniv, J. (1987). Histone-like proteins of bacteria. *Microbiol Rev* **51:** 301-19.

Duguet, M. (1993). The helical repeat of DNA at high temperature. *Nucleic Acids Res* **21:** 463-8.

Emerson, B.M. (2002). Specificity of gene regulation. *Cell* **109:** 267-70.

Enoru-Eta, J., Gigot, D., Thia-Toong, T.L., Glansdorff, N. and Charlier, D. (2000). Purification and characterization of Sa-lrp, a DNA-binding protein from the extreme thermoacidophilic archaeon *Sulfolobus acidocaldarius* homologous to the bacterial global transcriptional regulator Lrp. *J Bacteriol* **182:** 3661-72.

Erauso, G., Reysenbach, A., Godfroy, A., Meunier, J., Crump, B., Partensky, F., Baross, J.A., Merteinsson, V., Barbier, G., Pace, N.R. and Prieur, D. (1993). *Pyrococcus abyssi* sp. nov., a new hyperthermophilic archaeon isolated from a deep-sea hydrothermal vent. *Arch Microbiol* **160:** 338-349.

Ernst, W. (1983). The early Earth and the archaen rock record. In: *Earth's Earliest Biosphere*. Schopf, J. (ed). Princeton University Press, Princeton, N.J., pp. 41-52.

Facchiano, A.M., Ragone, R., Consalvi, V., Scandurra, R., De Rosa, M. and Colonna, G. (1995). Molecular properties of glutamate dehydrogenase from the extreme thermophilic archaebacterium *Sulfolobus solfataricus*. *Biochim Biophys Acta* **1251:** 170-6.

Facchiano, F., Ragone, R., Porcelli, M., Cacciapuoti, G. and Colonna, G. (1992). Effect of temperature on the propylamine transferase from *Sulfolobus solfataricus*, an extreme thermophilic archaebacterium. 1. Conformational behavior of the oligomeric enzyme in solution. *Eur J Biochem* **204:** 473-82.

Fitz-Gibbon, S.T. and House, C.H. (1999). Whole genome-based phylogenetic analysis of free-living microorganisms. *Nucleic Acids Res* **27:** 4218-22.

Fitz-Gibbon, S.T., Ladner, H., Kim, U.J., Stetter, K.O., Simon, M.I. and Miller, J.H. (2002). Genome sequence of the hyperthermophilic crenarchaeon *Pyrobaculum aerophilum*. *Proc Natl Acad Sci U S A* **99:** 984-9.

Fleischmann, R.D., Adams, M.D., White, O., Clayton, R.A., Kirkness, E.F., Kerlavage, A.R., Bult, C.J., Tomb, J.F., Dougherty, B.A., Merrick, J.M. McKenney, K., Sutton, G., Fitzhugh, W., Fields,

C., Jocayne, J.D., Scott, J., Shirley, R., Liu, L-I., Glodek, A., Kelley, J.M., Weidman, J.F., Phillips, C.A., Spriggs, T., Hedblom, E., Cotton., M.D., Utterback, T.R., Hanna, M.C., Nguyen, D.T., Saudek, D.M., Brandon, R.C., Fine, L.D., Fritchman, J.L., Fuhrmann, J.L., Geoghagen, N.S.M., Gnehm, C.L., McDonald, L.A., Small, K.V., Fraser, C.M., Smith, H.O., and Venter, J.C. (1995). Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* **269:** 496-512.

Forterre, P. (2001). Genomics and early cellular evolution. The origin of the DNA world. *C R Acad Sci III* **324:** 1067-76.

Forterre, P., Confalonieri, F. and Knapp, S. (1999). Identification of the gene encoding archaeal-specific DNA-binding proteins of the Sac-10b family. *Mol Microbiol* **32:** 669-670.

Forterre, P. and Philippe, H. (1999). Where is the root of the universal tree of life? *Bioessays* **21:** 871-9.

Galibert, F., Finan, T.M., Long, S.R., Puhler, A., Abola, P., Ampe, F., Barloy-Hubler, F., Barnett, M.J., Becker, A., Boistard, P., Bothe, G., Boutry, M., Bowser, L., Buhrmester, J., Cadieu, E., Capela, D., Chain, P., Cowie, A., Davis, R.W., Dreano, S., Federspiel, N.A., Fisher, R.F., Gloux, S., Godrie, T., Goffeau, A., Golding, B., Gouzy, J., Gurjal, M., Hernandez-Lucas, I., Hong, A., Huizar, L., Hyman, R.W., Jones, T., Kahn, D., Kahn, M.L., Kalman, S., Keating, D.H., Kiss, E., Komp, C., Lelaure, V., Masuy, D., Palm, C., Peck, M.C., Pohl, T.M., Portetelle, D., Purnelle, B., Ramsperger, U., Surzycki, R., Thebault, P., Vandenbol, M., Vorholter, F.J., Weidner, S., Wells, D.H., Wong, K., Yeh, K.C. and Batut, J. (2001). The composite genome of the legume symbiont *Sinorhizobium meliloti. Science* **293:** 668-72.

Garrell, J. and Campuzano, S. (1991). The helix-loop-helix domain: a common motif for bristles, muscles and sex. *Bioessays* **13:** 493-8.

Gerard, E., Labedan, B. and Forterre, P. (1998). Isolation of a minD-like gene in the hyperthermophilic archaeon *Pyrococcus* AL585, and phylogenetic characterization of related proteins in the three domains of life. *Gene* **222:** 99-106.

Gilbert, W. (1986). The RNA World. *Nature* **319:** 618.

Glansdorff, N. (2000). About the last common ancestor, the universal life-tree and lateral gene transfer: a reappraisal. *Mol Microbiol* **38:** 177-85.

Graham, D.E., Overbeek, R., Olsen, G.J. and Woese, C.R. (2000). An archaeal genomic signature. *Proc Natl Acad Sci U S A* **97:** 3304-8.

Grant, R.A., Filman, D.J., Finkel, S.E., Kolter, R. and Hogle, J.M. (1998). The crystal structure of Dps, a ferritin homolog that binds and protects DNA. *Nat Struct Biol* **5:** 294-303.

Grayling, R.A., Becktel, W.J. and Reeve, J.N. (1995). Structure and stability of histone HMf from the hyperthermophilic archaeon *Methanothermus fervidus. Biochemistry* **34:** 8441-8448.

Grayling, R.A., Sandman, K. and Reeve, J.N. (1996). DNA stability and DNA binding proteins. *Adv Protein Chem* **48:** 437-467.

Grote, M., Dijk, J. and Reinhardt, R. (1986). Ribosomal and DNA binding proteins of the thermoacidophilic archaebacterium *Sulfolobus acidocaldarius. Biochim Biophys Acta* **873:** 405-413.

Guerrier-Takada, C., Gardiner, K., Marsh, T.L., Pace, N.R. and Altman, S. (1983). The RNA moeity of RNase P is the catalytic subunit of the enzyme. *Cell* **35:** 849-857.

Hain, J., Reiter, W.D., Hudepohl, U. and Zillig, W. (1992). Elements of an archaeal promoter defined by mutational analysis. *Nucleic Acids Res* **20:** 5423-8.

Hamiche, A., Carot, V., Alilat, M., De Lucia, F., O'Donohue, M.F., Rèvet, B. and Prunell, A. (1996). Interaction of the histone $(H3\text{-}H4)_2$ tetramer of the nucleosome with positively supercoiled DNA minicircles: potential flipping of the protein from a left- to a right-handed superhelical form. *Proc Natl Acad Sci U S A* **93:** 7588-7593.

Hamiche, A. and Richard-Foy, H. (1998). The switch in the helical handedness of the histone $(H3\text{-}H4)_2$ tetramer within a nucleoprotein particle requires a reorientation of the H3-H3 interface. *J Biol Chem* **273:** 9261-9269.

Hartman, H. and Fedorov, A. (2002). The origin of the eukaryotic cell: a genomic investigation. *Proc Natl Acad Sci U S A* **99:** 1420-5.

Hayat, M.A. and Mancarella, D.A. (1995). Nucleoid proteins. *Micron* **26:** 461-480.

Hegde, R.S., Grossman, S.R., Laimins, L.A. and Sigler, P.B. (1992). Crystal structure at 1.7 A of the bovine papillomavirus-1 E2 DNA-binding domain bound to its DNA target. *Nature* **359:** 505-12.

Hengge-Aronis, R. (1996a). Back to log phase: σ^S as a global regulator in the osmotic control of gene expression in *Escherichia coli*. *Mol Microbiol* **21:** 887-893.

Hengge-Aronis, R. (1996b). Regulation of gene expression during entry into stationary phase. In: *Escherichia coli and Salmonella: Cellular and Molecular Biology*. Neidhardt, F.C., Curtisss, R., Ingrahmam, J.L., Lin, E.C.C., Low, K.B., Magasanik, B., Reznikoff, W.S., Riley, M., Schaechter, M. and Umbarger, H.E. (eds). ASM Press, Washington, D.C., pp. 1497-1512.

Hengge-Aronis, R. (2002). Recent insights into the general stress response regulatory network in *Escherichia coli*. *J Mol Microbiol Biotechnol* **4:** 341-6.

Horiike, T., Hamada, K., Kanaya, S. and Shinozawa, T. (2001). Origin of eukaryotic cell nuclei by symbiosis of Archaea in Bacteria is revealed by homology-hit analysis. *Nat Cell Biol* **3:** 210-4.

Hulton, C.S.J., Seirafi, A., Hinton, J.C.D., Sidebotham, J.M., Waddell, L., Pavitt, G.D., Owen-Hughes, T., Spassky, A., Buc, H. and Higgins, C.F. (1990). Histone-like protein H1 (H-NS), DNA supercoiling, and gene expression in Bacteria. *Cell* **63:** 631-642.

Imbert, M., Laine, B., Helbecque, N., Mornon, J.P., Henichart, J.P. and Sautiere, P. (1990). Conformational study of the chromosomal protein MC1 from the archaebacterium *Methanosarcina barkeri*. *Biochim Biophys Acta* **1038:** 346-54.

Ishihama, A. (1999). Modulation of the nucleoid, the transcription apparatus, and the translation machinery in bacteria for stationary phase survival. *Genes Cells* **4:** 135-43.

Jaxel, C., Bouthier de la Tour, C., Duguet, M. and Nadal, M. (1996). Reverse gyrase gene from *Sulfolobus shibatae* B12: gene structure, transcription unit and comparative sequence analysis of the two domains. *Nucleic Acids Res* **24:** 4668-75.

Jessen, T. (1999). Introduction of plasmid DNA into cells. In: *Current Protocols in Molecular Biology*. Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A. and Struhl, K. (eds). John Wiley & Sons, Inc., New York, pp. 1.8.1-1.8.10.

Johansson, J., Balsalobre, C., Wang, S.Y., Urbonaviciene, J., Jin, D.J., Sonden, B. and Uhlin, B.E. (2000). Nucleoid proteins stimulate stringently controlled bacterial promoters: a link between the cAMP-CRP and the (p)ppGpp regulons in *Escherichia coli*. *Cell* **102:** 475-85.

Joyce, G.F. (1989). RNA evolution and the origins of life. *Nature* **338:** 217-24.

Kamashev, D. and Rouviere-Yaniv, J. (2000). The histone-like protein HU binds specifically to DNA recombination and repair intermediates. *Embo J* **19:** 6527-35.

Kaneko, T., Nakamura, Y., Sato, S., Asamizu, E., Kato, T., Sasamoto, S., Watanabe, A., Idesawa, K., Ishikawa, A., Kawashima, K., Kimura, T., Kishida, Y., Kiyokawa, C., Kohara, M., Matsumoto, M., Matsuno, A., Mochizuki, Y., Nakayama, S., Nakazaki, N., Shimpo, S., Sugimoto, M., Takeuchi, C., Yamada, M. and Tabata, S. (2000). Complete genome structure of the nitrogen-fixing symbiotic bacterium *Mesorhizobium loti*. *DNA Res* **7:** 331-8.

Karlin, S. and Ghandour, G. (1985). Multiple-alphabet amino acid sequence comparisons of the immunoglobulin kappa-chain constant domain. *Proc Natl Acad Sci U S A* **82:** 8597-601.

Kasting, J.F. and Siefert, J.L. (2002). Life and the evolution of Earth's atmosphere. *Science* **296:** 1066-8.

Kawarabayasi, Y., Hino, Y., Horikawa, H., Jin-no, K., Takahashi, M., Sekine, M., Baba, S., Ankai, A., Kosugi, H., Hosoyama, A., Fukui, S., Nagai, Y., Nishijima, K., Otsuka, R., Nakazawa, H., Takamiya, M., Kato, Y., Yoshizawa, T., Tanaka, T., Kudoh, Y., Yamazaki, J., Kushida, N., Oguchi, A., Aoki, K., Masuda, S., Yanagii, M., Nishimura, M., Yamagishi, A., Oshima, T. and Kikuchi, H. (2001). Complete genome sequence of an aerobic thermoacidophilic crenarchaeon, *Sulfolobus tokodaii* strain7. *DNA Res* **8:** 123-40.

Kawarabayasi, Y., Hino, Y., Horikawa, H., Yamazaki, S., Haikawa, Y., Jin-no, K., Takahashi, M., Sekine, M., Baba, S., Ankai, A., Kosugi, H., Hosoyama, A., Fukui, S., Nagai, Y., Nishijima, K., Nakazawa, H., Takamiya, M., Masuda, S., Funahashi, T., Tanaka, T., Kudoh, Y., Yamazaki, J., Kushida, N., Oguchi, A., Kikuchi, H. and et al. (1999). Complete genome sequence of an aerobic hyper-thermophilic crenarchaeon, *Aeropyrum pernix* K1. *DNA Res* **6:** 83-101, 145-52.

Kawarabayasi, Y., Sawada, M., Horikawa, H., Haikawa, Y., Hino, Y., Yamamoto, S., Sekine, M., Baba, S., Kosugi, H., Hosoyama, A., Nagai, Y., Sakai, M., Ogura, K., Otsuka, R., Nakazawa, H., Takamiya, M., Ohfuku, Y., Funahashi, T., Tanaka, T., Kudoh, Y., Yamazaki, J., Kushida, N., Oguchi, A., Aoki, K. and Kikuchi, H. (1998). Complete sequence and gene organization of the genome of a hyper-thermophilic archaebacterium, *Pyrococcus horikoshii* OT3. *DNA Research* **5:** 55-76.

Kawashima, T., Amano, N., Koike, H., Makino, S., Higuchi, S., Kawashima-Ohya, Y., Watanabe, K., Yamazaki, M., Kanehori, K., Kawamoto, T., Nunoshiba, T., Yamamoto, Y., Aramaki, H., Makino, K. and Suzuki, M. (2000). Archaeal adaptation to higher temperatures revealed by genomic sequence of *Thermoplasma volcanium*. *Proc Natl Acad Sci U S A* **97:** 14257-62.

Kerr, L.D. (1995). Electrophoretic mobility shift assay. *Methods Enzymol* **254:** 619-632.

Kikuchi, A. and Asai, K. (1984). Reverse gyrase - a topoisomerase which introduces positive superhelical turns into DNA. *Nature* **309:** 677-81.

Klages, K.U. (1991). Characterization of an extremely thermophilic, anaerobic sulphur-dependent Archaebacterium. PhD thesis, The University of Waikato, Hamilton, New Zealand. 175 pp.

Klages, K.U. and Morgan, H.W. (1994). Characterization of an extremely thermophilic sulphur-metabolizing archaebacterium belonging to the Thermococcales. *Arch Microbiol* **162:** 261-266.

Klenk, H.P., Clayton, R.A., Tomb, J.F., White, O., Nelson, K.E., Ketchum, K.A., Dodson, R.J., Gwinn, M., Hickey, E.K., Peterson, J.D., Richardson, D.L., Kerlavage, A.R., Graham, D.E., Kyrpides, N.C., Fleischmann, R.D., Quackenbush, J., Lee, N.H., Sutton, G.G., Gill, S., Kirkness, E.F., Dougherty, B.A., McKenney, K., Adams, M.D., Loftus, B., Peterson, S., Reich, C.I., McNeil, L.K., Badger, J.H., Glodek, A., Zhou, L., Overbeek, R., Gocayne, J.D., Weidman, J.F., McDonald, L., Utterback, T., Cotton, M.D., Spriggs, T., Artiach, P., Kaine, B.P., Sykes, S.M., Sadow, P.W., D'Andrea, K.P., Bowman, C., Fujii, C., Garland, S.A., Mason, T.M., Olsen, G.J., Fraser, C.M., Smith, H.O., Woese, C.R. and Venter, J.C. (1997). The complete genome sequence of the hyperthermophilic, sulphate-reducing archaeon *Archaeoglobus fulgidus*. *Nature* **390:** 364-370.

Koch, A.L. (1994). Growth measurement. In: *Methods for General and Molecular Bacteriology*. Gerhardt, P., Murray, R.G.E., Woord, W.A. and Krieg, N.R. (eds). American Society for Microbiology, Washington, D.C., pp. 251-262.

Kolling, R. and Lother, H. (1985). AsnC: an autogenously regulated activator of asparagine synthetase A transcription in *Escherichia coli*. *J Bacteriol* **164:** 310-5.

Kolter, R., Siegele, D.A. and Tormo, A. (1993). The stationary phase of the bacterial life cycle. *Annu Rev Microbiol* **47:** 855-874.

Koonin, E.V., Makarova, K.S. and Aravind, L. (2001). Horizontal gene transfer in prokaryotes: quantification and classification. *Annu Rev Microbiol* **55:** 709-42.

Kornberg, R.D. and Lorch, Y. (1992). Chromatin structure and transcription. *Annu Rev Cell Biol* **8:** 563-87.

Kunst, F., Ogasawara, N., Moszer, I., Albertini, A.M., Alloni, G., Azevedo, V., Bertero, M.G., Bessieres, P., Bolotin, A., Borchert, S., Borriss, R., Boursier, L., Brans, A., Braun, M., Brignell, S.C., Bron, S., Brouillet, S., Bruschi, C.V., Caldwell, B., Capuano, V., Carter, N.M., Choi, S.K., Codani, J.J., Connerton, I.F., Danchin, A. and et al. (1997). The complete genome sequence of the gram-positive bacterium *Bacillus subtilis*. *Nature* **390:** 249-56.

Kyrpides, N., Overbeek, R. and Ouzounis, C. (1999). Universal protein families and the functional content of the last universal common ancestor. *J Mol Evol* **49:** 413-23.

Kyrpides, N.C. and Ouzounis, C.A. (1999). Transcription in Archaea. *Proc Natl Acad Sci U S A* **96:** 8545-50.

Laine, B., Culard, F., Maurizot, J.C. and Sautiere, P. (1991). The chromosomal protein MC1 from the archaebacterium *Methanosarcina* sp. CHTI 55 induces DNA bending and supercoiling. *Nucleic Acids Res* **19:** 3041-5.

Lake, J.A. and Rivera, M.C. (1994). Was the nucleus the first endosymbiont? *Proc Natl Acad Sci U S A* **91:** 2880-1.

Langer, D., Hain, J., Thuriaux, P. and Zillig, W. (1995). Transcription in Archaea: similarity to that in Eucarya. *Proc Natl Acad Sci U S A* **92:** 5768-5772.

Lavoie, B.D. and Chaconas, G. (1993). Site-specific HU binding in the Mu transpososome: conversion of a sequence-independent DNA-binding protein into a chemical nuclease. *Genes Dev* **7:** 2510-9.

Le Hégarat, F., Salti-Montesanto, V., Hauck, Y. and Hirschbein, L. (1993). Purification and characterization of the Hu-like protein HPB9 from the *Bacillus subtilis* nucleoid. *Biochim Biophys Acta* **1172:** 101-107.

Lehninger, A.L. (1982). Biochemistry. Worth Publishers, New York.

Leigh, J.A. (1999). Transcriptional regulation in Archaea. *Curr Opin Microbiol* **2:** 131-4.

Leipe, D.D., Wolf, Y.I., Koonin, E.V. and Aravind, L. (2002). Classification and evolution of P-loop GTPases and related ATPases. *J Mol Biol* **317:** 41-72.

Leonard, P.M., Smits, S.H., Sedelnikova, S.E., Brinkman, A.B., de Vos, W.M., van der Oost, J., Rice, D.W. and Rafferty, J.B. (2001). Crystal structure of the Lrp-like transcriptional regulator from the archaeon *Pyrococcus furiosus*. *Embo J* **20:** 990-7.

Lewin, B. (1999a). Initiation of Transcription. In: *Genes VII.* Oxford University Press, Oxford, pp. 617-648.

Lewin, B. (1999b). Regulation of Transcription. In: *Genes VII.* Oxford University Press, Oxford, pp. 649-684.

Lin, R., D'Ari, R. and Newman, E.B. (1990). The leucine regulon of *Escherichia coli*: a mutation in *rblA* alters expression of leucine-dependent metabolic operons. *J Bacteriol* **172:** 4259-4235.

Lindahl, M., Svensson, L.A., Liljas, A., Sedelnikova, S.E., Eliseikina, I.A., Fomenkova, N.P., Nevskaya, N., Nikonov, S.V., Garber, M.B., Muranova, T.A. and et al. (1994). Crystal structure of the ribosomal protein S6 from *Thermus thermophilus*. *Embo J* **13:** 1249-54.

Loewen, P.C. and Hengge-Aronis, R. (1994). The role of the sigma factor σ^S (KatF) in bacterial global regulation. *Annu Rev Microbiol* **48:** 53-80.

Lòpez-Garcìa, P. and Forterre, P. (1997). DNA topology in hyperthermophilic archaea: reference states and their variation with growth phase, growth temperature, and temperature stresses. *Mol Microbiol* **23:** 1267-1279.

Lòpez-Garcìa, P. and Forterre, P. (1999). Control of DNA topology during thermal stress in hyperthermophilic archaea: DNA topoisomerase levels, activities and induced thermotolerance during heat and cold shock in *Sulfolobus*. *Mol Microbiol* **33:** 766-77.

Lopez-Garcia, P., Forterre, P. and Philippe, H. (1999). The root of the tree of life in the light of the covarion model. *J Mol Evol* **49:** 496-508.

Luger, K., Mäder, A.W., Richmond, R.K., Sargent, D.F. and Richmond, T.J. (1997). Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* **389:** 251-260.

Lurz, R., Grote, M., Dijk, J., Reinhardt, R. and Dobrinski, B. (1986). Electron microscopic study of DNA complexes with proteins from the Archaebacterium *Sulfolobus acidocaldarius*. *Embo J* **5:** 3715-3721.

Maeder, D.L., Weiss, R.B., Dunn, D.M., Cherry, J.L., Gonzalez, J.M., DiRuggiero, J. and Robb, F.T. (1999). Divergence of the hyperthermophilic archaea *Pyrococcus furiosus* and *P. horikoshii* inferred from complete genomic sequences. *Genetics* **152:** 1299-305.

Mager, W.H. and De Kruijff, A.J. (1995). Stress-induced transcriptional activation. *Microbiol Rev* **59:** 506-31.

Mai, V.Q., Chen, X., Hong, R. and Huang, L. (1998). Small abundant DNA binding proteins from the thermoacidophilic archaeon *Sulfolobus shibatae* constrain negative DNA supercoils. *J Bacteriol* **180:** 2560-3.

Marc, F., Sandman, K., Lurz, R. and Reeve, J.N. (2002). Archaeal histone tetramerization determines DNA affinity and the direction of DNA supercoiling. *J Biol Chem* (in press).

Martin, W. and Muller, M. (1998). The hydrogen hypothesis for the first eukaryote. *Nature* **392:** 37-41.

Martinèz, A. and Kolter, R. (1997). Protection of DNA during oxidative stress by the nonspecific DNA-binding protein Dps. *J Bacteriol* **179:** 5188-5194.

Mayer, C., Kohrer, C., Grobner, P. and Piendl, W. (1998). MvaL1 autoregulates the synthesis of the three ribosomal proteins encoded on the MvaL1 operon of the archaeon *Methanococcus vannielii* by inhibiting its own translation before or at the formation of the first peptide bond. *Mol Microbiol* **27:** 455-68.

McGovern, V., Higgins, N.P., Chiz, R.S. and Jaworski, A. (1994). H-NS over-expression induces an artificial stationary phase by silencing global transcription. *Biochimie* **76:** 1019-1029.

Mojzsis, S.J., Arrhenius, G., McKeegan, K.D., Harrison, T.M., Nutman, A.P. and Friend, C.R. (1996). Evidence for life on Earth before 3,800 million years ago. *Nature* **384:** 55-9.

Moulton, V., Gardner, P.P., Pointon, R.F., Creamer, L.K., Jameson, G.B. and Penny, D. (2000). RNA folding argues against a hot-start origin of life. *J Mol Evol* **51:** 416-21.

Musgrave, D., Forterre, P. and Slesarev, A. (2000). Negative constrained DNA supercoiling in archaeal nucleosomes. *Mol Microbiol* **35:** 341-9.

Musgrave, D.R., Sandman, K.M. and Reeve, J.N. (1991). DNA binding by the archaeal histone HMf results in positive supercoiling. *Proc Natl Acad Sci U S A* **88:** 10397-10401.

Muskhelishvili, G. and Travers, A.A. (1997). The stabilization of DNA microloops by FIS - a mechanism for torsional transmission in transcription activation and DNA inversion. *Nucleic Acids Mol Biol* **11:** 179–190.

Napoli, A., van der Oost, J., Sensen, C.W., Charlebois, R.L., Rossi, M. and Ciaramella, M. (1999). An Lrp-like protein of the hyperthermophilic archaeon *Sulfolobus solfataricus* which binds to its own promoter. *J Bacteriol* **181:** 1474-80.

Napoli, A., Zivanovic, Y., Bocs, C., Buhler, C., Rossi, M., Forterre, P. and Ciaramella, M. (2002). DNA bending, compaction and negative supercoiling by the architectural protein Sso7d of *Sulfolobus solfataricus*. *Nucleic Acids Res* **30:** 2656-62.

Nelson, K.E., Clayton, R.A., Gill, S.R., Gwinn, M.L., Dodson, R.J., Haft, D.H., Hickey, E.K., Peterson, J.D., Nelson, W.C., Ketchum, K.A., McDonald, L., Utterback, T.R., Malek, J.A., Linher, K.D., Garrett, M.M., Stewart, A.M., Cotton, M.D., Pratt, M.S., Phillips, C.A., Richardson, D.,

Heidelberg, J., Sutton, G.G., Fleischmann, R.D., Eisen, J.A., Fraser, C.M. and et al. (1999). Evidence for lateral gene transfer between Archaea and bacteria from genome sequence of *Thermotoga maritima*. *Nature* **399:** 323-9.

Newman, E.B. and Lin, R. (1995). Leucine-responsive regulatory protein: a global regulator of gene expression in *E. coli*. *Annu Rev Microbiol* **49:** 747-75.

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. and DasSarma, S. (2000). Genome sequence of *Halobacterium* species NRC-1. *Proc Natl Acad Sci U S A* **97:** 12176-81.

Nikolov, D.B., Chen, H., Halay, E.D., Usheva, A.A., Hisatake, K., Lee, D.K., Roeder, R.G. and Burley, S.K. (1995). Crystal structure of a TFIIB-TBP-TATA-element ternary complex. *Nature* **377:** 119-28.

Nolling, J., Breton, G., Omelchenko, M.V., Makarova, K.S., Zeng, Q., Gibson, R., Lee, H.M., Dubois, J., Qiu, D., Hitti, J., Wolf, Y.I., Tatusov, R.L., Sabathe, F., Doucette-Stamm, L., Soucaille, P., Daly, M.J., Bennett, G.N., Koonin, E.V. and Smith, D.R. (2001). Genome sequence and comparative analysis of the solvent-producing bacterium *Clostridium acetobutylicum*. *J Bacteriol* **183:** 4823-38.

Olsen, G.J. and Woese, C.R. (1996). Lessons from an archaeal genome: what are were learning from *Methanococcus jannaschii*. *Trends Genet* **12:** 377-379.

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

Owen-Hughes, T.A., Pavitt, G.D., Santos, D.S., Sidebotham, J.D., Hulton, J.C.D., Hinton, D. and Higgins, C.F. (1992). The chromatin-associated protein H-NS interacts with curved DNA to influence DNA topology and gene expression. *Cell* **71:** 255-265.

Pabo, C.O. and Lewis, M. (1982). The operator-binding domain of lambda repressor: structure and DNA recognition. *Nature* **298:** 443-7.

Pace, N.R. (1997). A molecular view of microbial diversity and the biosphere. *Science* **276:** 734-740.

Pan, C.Q., Finkel, S.E., Cramton, S.E., Feng, J.A., Sigman, D.S. and Johnson, R.C. (1996). Variable structures of Fis-DNA complexes determined by flanking DNA-protein contacts. *J Mol Biol* **264:** 675-95.

Penny, D. and Poole, A. (1999). The nature of the last universal common ancestor. *Curr Opin Genet Dev* **9:** 672-7.

Perdew, G.H., Schaup, H.W. and Selivonchick, D.P. (1983). The use of a zwitterionic detergent in two-dimensional gel electrophoresis of trout liver microsomes. *Anal Biochem* **135:** 453-5.

Pereira, S.L., Grayling, R.A., Lurz, R. and Reeve, J.N. (1997). Archaeal nucleosomes. *Proc Natl Acad Sci U S A* **94:** 12633-12637.

Platko, J.V. and Calvo, J.M. (1993). Mutations affecting the ability of *Escherichia coli* Lrp to bind DNA, activate transcription, or respond to leucine. *J Bacteriol* **175:** 1110-7.

Poole, A.M., Jeffares, D.C. and Penny, D. (1998). The path from the RNA world. *J Mol Evol* **46:** 1-17.

Prithivirajsingh, S., Mishra, S.K. and Mahadevan, A. (2001). Functional analysis of a chromosomal arsenic resistance operon in *Pseudomonas fluorescens* strain MSP3. *Mol Biol Rep* **28:** 63-72.

Rehaber, V. and Jaenicke, R. (1992). Stability and reconstitution of D-glyceraldehyde-3-phosphate dehydrogenase from the hyperthermophilic eubacterium *Thermotoga maritima. J Biol Chem* **267:** 10999-1006.

Reiter, W.D., Hudepohl, U. and 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 U S A* **87:** 9509-13.

Robinson, H., Gao, Y.G., McCrary, B.S., Edmondson, S.P., Shriver, J.W. and Wang, A.H. (1998). The hyperthermophile chromosomal protein Sac7d sharply kinks DNA. *Nature* **392:** 202-5.

Ronimus, R.S. and Musgrave, D.R. (1996). Purification and characterization of a histone-like protein from the Archaeal isolate AN1, a member of the *Thermococcales. Mol Microbiol* **20:** 77-86.

Ronimus, R.S., Reysenbach, A.L., Musgrave, D.R. and Morgan, H.W. (1997). The phylogenetic position of the *Thermococcus* isolate AN1 based on 16S rRNA gene sequence analysis: a proposal that AN1 represents a new species, *Thermococcus zilligii* sp. nov. *Arch Microbiol* **168:** 245-248.

Rosenstein, R., Peschel, A., Wieland, B. and Gotz, F. (1992). Expression and regulation of the antimonite, arsenite, and arsenate resistance operon of *Staphylococcus xylosus* plasmid pSX267. *J Bacteriol* **174:** 3676-83.

Rowlands, T., Baumann, P. and Jackson, S.P. (1994). The TATA-binding protein: a general transcription factor in eukaryotes and archaebacteria. *Science* **264:** 1326-1329.

Ruepp, A., Graml, W., Santos-Martinez, M.L., Koretke, K.K., Volker, C., Mewes, H.W., Frishman, D., Stocker, S., Lupas, A.N. and Baumeister, W. (2000). The genome sequence of the thermoacidophilic scavenger *Thermoplasma acidophilum. Nature* **407:** 508-13.

Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989). Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Sambrook, J. and Russell, D.W. (2001). Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Sandman, K., Grayling, R.A., Dobrinski, B., Lurz, R. and Reeve, J.N. (1994). Growth-phase-dependent synthesis of histones in the archaeon *Methanothermus fervidus. Proc Natl Acad Sci U S A* **91:** 12624-12628.

Sandman, K., Krzycki, J.A., Dobrinski, B., Lurz, R. and Reeve, J.N. (1990). HMf, a DNA-binding protein isolated from the hyperthermophilic archaeon *Methanothermus fervidus*, is most closely related to histones. *Proc Natl Acad Sci U S A* **87:** 5788-5791.

Sandman, K. and Reeve, J.N. (1998). Origin of the eukaryotic nucleus. *Science* **280:** 501, 503.

Sato, T. and Kobayashi, Y. (1998). The ars operon in the skin element of *Bacillus subtilis* confers resistance to arsenate and arsenite. *J Bacteriol* **180:** 1655-61.

Schägger, H. and von Jagow, G. (1987). Tricine-sodium dodecyl sulfate-polyacrylamide gel electrophoresis for the separation of proteins in the range from 1 to 100 kDa. *Anal Biochem* **166:** 368-379.

Schmid, M.B. (1990). More than just "histone-like" proteins. *Cell* **63:** 451-453.

Schneider, R., Lurz, R., Luder, G., Tolksdorf, C., Travers, A. and Muskhelishvili, G. (2001). An architectural role of the *Escherichia coli* chromatin protein FIS in organising DNA. *Nucleic Acids Res* **29:** 5107-14.

Schneider, R., Travers, A., Kutateladze, T. and Muskhelishvili, G. (1999). A DNA architectural protein couples cellular physiology and DNA topology in Escherichia coli. *Mol Microbiol* **34:** 953-64.

Schopf, J.W. (1993). Microfossils of the Early Archean Apex chert: new evidence of the antiquity of life. *Science* **260:** 640-6.

Schopf, J.W., Kudryavtsev, A.B., Agresti, D.G., Wdowiak, T.J. and Czaja, A.D. (2002). Laser-Raman imagery of Earth's earliest fossils. *Nature* **416:** 73-6.

Schopf, J.W. and Packer, B.M. (1987). Early Archean (3.3-billion to 3.5-billion-year-old) microfossils from Warrawoona Group, Australia. *Science* **237:** 70-3.

Searcy, D.G. (1986). The archaebacterial histone HTa. In: *Bacterial Chromatin.* Gualerzi, C.O. and Pon, C.L. (eds). Springer-Verlag, Heidelberg, pp. 175-184.

Searcy, D.G., Montenay-Garestier, T., Laston, D.J. and Helene, C. (1998). Tyrosine environment and phosphate binding in the archaebacterial histone-like protein HTa. *Biochim Biophys Acta* **953:** 321-333.

Shahrestanifar, M., Saha, D.P., Scala, L.A., Basu, A. and Howells, R.D. (1994). Cloning of a human cDNA encoding a putative nucleotide-binding protein related to *Escherichia coli* MinD. *Gene* **147:** 281-5.

She, Q., Singh, R.K., Confalonieri, F., Zivanovic, Y., Allard, G., Awayez, M.J., Chan-Weiher, C.C., Clausen, I.G., Curtis, B.A., De Moors, A., Erauso, G., Fletcher, C., Gordon, P.M., Heikamp-de Jong, I., Jeffries, A.C., Kozera, C.J., Medina, N., Peng, X., Thi-Ngoc, H.P., Redder, P., Schenk, M.E., Theriault, C., Tolstrup, N., Charlebois, R.L., Doolittle, W.F., Duguet, M., Gaasterland, T., Garrett, R.A., Ragan, M.A., Sensen, C.W. and Van der Oost, J. (2001). The complete genome of the crenarchaeon *Sulfolobus solfataricus* P2. *Proc Natl Acad Sci U S A* **98:** 7835-40.

Shimizu, T., Ohtani, K., Hirakawa, H., Ohshima, K., Yamashita, A., Shiba, T., Ogasawara, N., Hattori, M., Kuhara, S. and Hayashi, H. (2002). Complete genome sequence of *Clostridium perfringens*, an anaerobic flesh-eater. *Proc Natl Acad Sci U S A* **99:** 996-1001.

Silman, I.H., Albu-Weissenberg, M. and Katchalski, E. (1966). Some water-insoluble papain derivatives. *Biopolymers* **4:** 441-8.

Slesarev, A.I., Mezhevaya, K.V., Makarova, K.S., Polushin, N.N., Shcherbinina, O.V., Shakhova, V.V., Belova, G.I., Aravind, L., Natale, D.A., Rogozin, I.B., Tatusov, R.L., Wolf, Y.I., Stetter, K.O., Malykh, A.G., Koonin, E.V. and Kozyavkin, S.A. (2002). The complete genome of hyperthermophile *Methanopyrus kandleri* AV19 and monophyly of archaeal methanogens. *Proc Natl Acad Sci U S A* **99:** 4644-9.

Smith, D.R., Doucette-Stamm, Deloughery, C., Lee, H., Dubois, J., Aldredge, T., Bashirzadeh, R., Blakely, D., Cook, R., Gilbert, K., Harrison, D., Hoang, L., Keagle, P., Lumm, W., Pothier, B., Qiu, D., Spadafora, R., Vicaire, R., Wang, Y., Wierzbowski, J., Gibson, R., Jiwani, N., Caruso, A., Bush, D., Safer, H., Patwell, D., Parbhakar, S., McDougall, S., Shimer, G., Goyal, A., Pietrokovski, S., Church, G.M., Daniels, C.J., Mao, J., Rice, P., Nölling, J. and Reeve, J.N. (1997). Complete genome sequence of *Methanobacterium thermoautotrophicum* ΔH: functional analysis and comparitive genomics. *J Bacteriol* **179:** 7135-7155.

Snel, B., Bork, P. and Huynen, M.A. (1999). Genome phylogeny based on gene content. *Nat Genet* **21:** 108-10.

Soppa, J. (1999). Transcription initiation in Archaea: facts, factors and future aspects. *Mol Microbiol* **31:** 1295-305.

Starich, M.R., Sandman, K., Reeve, J.N. and Summers, M.F. (1996). NMR structure of HMfB from the hyperthermophile, *Methanothermus fervidus*, confirms that this archaeal protein is a histone. *J Mol Biol* **255:** 187-203.

Steger, G., Ham, J., Lefebvre, O. and Yaniv, M. (1995). The bovine papillomavirus 1 E2 protein contains two activation domains: one that interacts with TBP and another that functions after TBP binding. *Embo J* **14:** 329-40.

Stein, J.L., Marsh, T.L., Wu, K.Y., Shizuya, H. and DeLong, E.F. (1996). Characterization of uncultivated prokaryotes: isolation and analysis of a 40-kilobase-pair genome fragment from a planktonic marine archaeon. *J Bacteriol* **178:** 591-9.

Suzuki, K., Wakao, N., Kimura, T., Sakka, K. and Ohmiya, K. (1998). Expression and regulation of the arsenic resistance operon of *Acidiphilium multivorum* AIU 301 plasmid pKW301 in *Escherichia coli*. *Appl Environ Microbiol* **64:** 411-8.

Tatusov, R.L., Koonin, E.V. and Lipman, D.J. (1997). A genomic perspective on protein families. *Science* **278:** 631-7.

Tatusov, R.L., Natale, D.A., Garkavtsev, I.V., Tatusova, T.A., Shankavaram, U.T., Rao, B.S., Kiryutin, B., Galperin, M.Y., Fedorova, N.D. and Koonin, E.V. (2001). The COG database: new developments in phylogenetic classification of proteins from complete genomes. *Nucleic Acids Res* **29:** 22-8.

Thomm, M. (1996). Archaeal transcription factors and their role in transcription initiation. *FEMS Microbiol Rev* **18:** 159-171.

Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F. and Higgins, D.G. (1997). The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* **25:** 4876-82.

Thompson, J.D., Higgins, D.G. and Gibson, T.J. (1994). CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position specific gap penalties and weight matrix choice. *Nucleic Acids Res* **22:** 4673-4680.

Thomsen, J., De Biase, A., Kaczanowski, S., Macario, A.J., Thomm, M., Zielenkiewicz, P., MacColl, R. and 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.

Toulmè, F., Le Cam, E., Teyssier, C., Delain, E., Sautiere, P., Maurizot, J.C. and Culard, F. (1995). Conformational changes of DNA minicircles upon the binding of the archaebacterial histone-like protein MC1. *J Biol Chem* **270:** 6286-6291.

Urnov, F.D. and Wolffe, A.P. (2001). Chromatin remodeling and transcriptional activation: the cast (in order of appearance). *Oncogene* **20:** 2991-3006.

Vitale, G., Fabre, E. and Hurt, E.C. (1996). NBP35 encodes an essential and evolutionary conserved protein in *Saccharomyces cerevisiae* with homology to a superfamily of bacterial ATPases. *Gene* **178:** 97-106.

Voorhorst, W.G., Gueguen, Y., Geerling, A.C., Schut, G., Dahlke, I., Thomm, M., van der Oost, J. and de Vos, W.M. (1999). Transcriptional regulation in the hyperthermophilic archaeon *Pyrococcus furiosus*: coordinated expression of divergently oriented genes in response to beta-linked glucose polymers. *J Bacteriol* **181:** 3777-83.

Walter, M. (1983). Archaen stromatolites: evidence of the Earth's earliest benthos. In: *Earth's Earliest Biosphere.* Schopf, J.W. (ed). Princeton University Press, Princeton, N.J., pp. 187-213.

Wang, J.C. (2002). Cellular roles of DNA topoisomerases: a molecular perspective. *Nat Rev Mol Cell Biol* **3:** 430-40.

Wang, J.C. and Lynch, A.S. (1993). Transcription and DNA supercoiling. *Curr Opin Genet Dev* **3:** 764-8.

Wang, J.Y. and Syvanen, M. (1992). DNA twist as a transcriptional sensor for environmental changes. *Mol Microbiol* **6:** 1861-6.

Wang, Q. and Calvo, J.M. (1993). Lrp, a major regulatory protein in *Escherichia coli*, bends DNA and can organize the assembly of a higher-order nucleoprotein structure. *Embo J* **12:** 2495-501.

Watson, J.D. and Crick, F.H. (1953). Molecular structure of nucleic acids: a structure for deoxyribose nucleic acid. *Nature* **171:** 737-738.

White, O., Eisen, J.A., Heidelberg, J.F., Hickey, E.K., Peterson, J.D., Dodson, R.J., Haft, D.H., Gwinn, M.L., Nelson, W.C., Richardson, D.L., Moffat, K.S., Qin, H., Jiang, L., Pamphile, W., Crosby, M., Shen, M., Vamathevan, J.J., Lam, P., McDonald, L., Utterback, T., Zalewski, C., Makarova, K.S., Aravind, L., Daly, M.J. and Fraser, C.M. (1999). Genome sequence of the radioresistant bacterium *Deinococcus radiodurans* R1. *Science* **286:** 1571-7.

Wingfield, P.T. and Palmer, I. (1996). Folding and purification of insoluble (inclusion body) proteins from *Escherichia coli.* In: *Current Protocols in Protein Science.* Coligan, J.E., Dunn, B.M., Ploegh, H.L., Speicher, D.W. and Wingfield, P.T. (eds). John Wiley & Sons, Inc., New York, pp. 6.5.1-6.5.27.

Woese, C. (1998). The universal ancestor. *Proc Natl Acad Sci U S A* **95:** 6854-6859.

Woese, C.R. (1987). Bacterial evolution. *Microbiol Rev* **51:** 221-271.

Woese, C.R. (2000). Interpreting the universal phylogenetic tree. *Proc Natl Acad Sci U S A* **97:** 8392-6.

Woese, C.R. (2002). On the evolution of cells. *Proc Natl Acad Sci U S A* **99:** 8742-8747.

Woese, C.R., Kandler, O. and Wheelis, M.L. (1990). Towards a natural system of organisms: Proposal for the domains Archaea, Bacteria and Eukarya. *Proc Natl Acad Sci U S A* **87:** 4756-4579.

Woese, C.R. and Olsen, G.J. (1986). Archaebacterial phylogeny: perspectives on the urkingdoms. *Syst Appl Microbiol* **7:** 161-77.

Wolf, S.G., Frenkiel, D., Arad, T., Finkel, S.E., Kolter, R. and Minsky, A. (1999). DNA protection by stress-induced biocrystallization. *Nature* **400:** 83-5.

Wolffe, A.P. (2001). Transcriptional regulation in the context of chromatin structure. *Essays Biochem* **37:** 45-57.

Wolin, E.A., Wolin, M.J. and Wolfe, R.S. (1963). Formation of methane by bacterial extracts. *J Biol Chem* **238:** 2882-2886.

Wood, D.W., Setubal, J.C., Kaul, R., Monks, D.E., Kitajima, J.P., Okura, V.K., Zhou, Y., Chen, L., Wood, G.E., Almeida, N.F., Jr., Woo, L., Chen, Y., Paulsen, I.T., Eisen, J.A., Karp, P.D., Bovee, D., Sr., Chapman, P., Clendenning, J., Deatherage, G., Gillet, W., Grant, C., Kutyavin, T., Levy, R., Li, M.J., McClelland, E., Palmieri, A., Raymond, C., Rouse, G., Saenphimmachak, C., Wu, Z., Romero, P., Gordon, D., Zhang, S., Yoo, H., Tao, Y., Biddle, P., Jung, M., Krespan, W., Perry, M., Gordon-Kamm, B., Liao, L., Kim, S., Hendrick, C., Zhao, Z.Y., Dolan, M., Chumley, F., Tingey, S.V., Tomb, J.F., Gordon, M.P., Olson, M.V. and Nester, E.W. (2001). The genome of the natural genetic engineer *Agrobacterium tumefaciens* C58. *Science* **294:** 2317-23.

Xiong, A., Gottman, A., Park, C., Baetens, M., Pandza, S. and Matin, A. (2000). The EmrR protein represses the *Escherichia coli emrRAB* multidrug resistance operon by directly binding to its promoter region. *Antimicrob Agents Chemother* **44:** 2905-7.

Xu, J. and Johnson, R.C. (1995). Identification of genes negatively regulated by Fis: Fis and RpoS comodulate growth-phase-dependent gene expression in *Escherichia coli*. *J Bacteriol* **177:** 938-947.

Xue, H., Guo, R., Wen, Y., Liu, D. and Huang, L. (2000). An abundant DNA binding protein from the hyperthermophilic archaeon *Sulfolobus shibatae* affects DNA supercoiling in a temperature-dependent fashion. *J Bacteriol* **182:** 3929-33.

Zeikus, J.G., Heege, P.W. and Anderson, M.A. (1979). *Thermoanaerobium brockii* gen. nov., sp. nov., a new chemoorganotrophic, caldoactive, anaerobic bacterium. *Arch Microbiol* **122:** 41-48.

Zeikus, J.G. and Wolfe, R.S. (1972). *Methanobacterium thermoautotrophicus* sp. n., an anaerobic, autotrophic, extreme thermophile. *J Bacteriol* **109:** 707-15.

Zhao, Y., Jeruzalmi, D., Moarefi, I., Leighton, L., Lasken, R. and Kuriyan, J. (1999). Crystal structure of an archaebacterial DNA polymerase. *Structure Fold Des* **7:** 1189-99.

Zillig, W., Stetter, K.O. and Janekovic, D. (1979). DNA-dependent RNA polymerase from the archaebacterium *Sulfolobus acidocaldarius*. *Eur J Biochem* **96:** 597-604.