SULFOLOBUS AS A MODEL ORGANISM FOR THE STUDY OF DIVERSE BIOLOGICAL INTERESTS; FORAYS INTO THERMAL VIROLOGY AND OXIDATIVE STRESS

by

Blake Alan Wiedenheft

A dissertation submitted in partial fulfillment
of the requirements for the degree

of

Doctor of Philosophy

In

Microbiology

MONTANA STATE UNIVERSITY
Bozeman, Montana

November 2006

APPROVAL

of a dissertation submitted by

Blake Alan Wiedenheft

This dissertation has been read by each member of the dissertation committee and has been found to be satisfactory regarding content, English usage, format, citations, bibliographic style, and consistency, and is ready for submission to the Division of Graduate Education.

Dr. Mark Young and Dr. Trevor Douglas

Approved for the Department of Microbiology

Dr.Tim Ford

Approved for the Division of Graduate Education

Dr. Carl A. Fox

STATEMENT OF PERMISSION TO USE

In presenting this dissertation in partial fulfillment of the requirements for a doctoral degree at Montana State University – Bozeman, I agree that the Library shall make it available to borrowers under rules of the Library. I further agree that copying of this dissertation is allowable only for scholarly purposes, consistent with "fair use" as prescribed in the U.S. Copyright Law. Requests for extensive copying or reproduction of this dissertation should be referred to ProQuest Information and Learning, 300 North Zeeb Road, Ann Arbor, Michigan 48106, to whom I have granted "the exclusive right to reproduce and distribute my dissertation in and from microfilm along with the non-exclusive right to reproduce and distribute my abstract in any format in whole or in part."

Blake Alan Wiedenheft

November, 2006

DEDICATION

This work was funded in part through grants from the National Aeronautics and Space Administration Program (NAG5-8807) in support of Montana State University's Center for Life in Extreme Environments (MCB-0132156), and the National Institutes of Health (R01 EB00432 and DK57776). Invaluable technical assistance was provided by D. Willits, M. Allen, S. Brumfield, L. Liepold, Z. Varpness, D. Nielsen, A. Ortmann, B. Widener and nearly every other current or previous members of the Mark Young/Trevor Douglas group. Interdisciplinary collaborations with Professors C. Martin Lawrence, Matthew Lavin, Brian Bothner and John Van der Oost have substantially broadened my scientific scope and challenged me to learn the scientific vernacular that comes with each of their respective disciplines.

Special thanks to my mentors (Drs. Mark Young and Trevor Douglas) whom have demonstrated a contagious zeal for both a life of science and for life "outside" of science. These two individuals have managed to merge the appeal of Yellowstone National Park with existing scientific interests in virology, biomedicine and material science. I hope to use this example of creativity and selfless collaboration as a template for my future scientific endeavors.

Thanks to friends and family, whom frequently tolerated unsolicited lectures on any number of 'burning' scientific topics. The patience of these people is a lesson I would do well to learn. Enduring patience has been a reoccurring theme for one individual in particular. Michelle Flenniken has not only been an invaluable scientific reference, but also a partner in diverse adventures on several continents, more then a few of which have been arguably ridiculous.

TABLE OF CONTENTS

TABLE OF CONTENTS - CONTINUED

TABLE OF CONTENTS – CONTINUED

TABLE OF CONTENTS – CONTINUED

TABLE OF CONTENTS – CONTINUED

.

LIST OF TABLES

LIST OF FIGURES

LIST OF FIGURES - CONTINUED

LIST OF FIGURES - CONTINUED

LIST OF BOXES

ABSTRACT

My research interests have focused on two distinct aspects of *Sulfolobus* biology: virology and oxidative stress. My major contribution to the emerging field of thermal virology has been the isolation, characterization and comparative genomic analysis of a spindle-shaped virus partical (SSV RH) infecting the thermoacidophilic archaeal host *Sulfolobus solfataricus* (*18*). Insights from this comparative genomic analysis have served as a platform for targeted structural studies, as well as providing molecular tools used to follow the viral life cycle in culture and for assessing the ecological significance of these viruses in the environment (*9, 19-24*).

My research endeavors in oxidative stress arose from an early interest in iron metabolism and protective mechanisms that allow life to cope with the paradoxical role that iron plays in biological systems. Pursuit of this interest has lead to the discovery of a new class of proteins termed, "DPS-Like" (*7, 17, 24, 25*). These previously unrecognized proteins function as antioxidants and are widely distributed across both prokaryotic domains of life.

CHAPTER 1

AN ARCHAEAL PERSPECTIVE

Abstract

The advent of DNA sequencing technologies revolutionized microbial taxonomy and led to the identification of a new domain of life, termed Archaea (*26-28*). Archaea are single-celled prokaryotes that are morphologically similar to their prokaryotic counterparts of the Bacterial domain. This gross morphological similarity and their bacterial-like genomic structure (condensed circular double stranded DNA chromosomes) obscured their recognition as a novel domain of life until 1977, when Carl Woese proposed his three-domain system of phylogeny (*29, 30*). This new phylogenetic system, which is based on DNA sequences of small ribosomal RNAs, posits the existence of a new prokaryotic "domain" of life, originally referred to as Archaebacteria and later shortened to Archaea. This radical hypothesis initially received a cool reception by the skeptical scientific community, who questioned the information content and interpretations of ribosomal rDNA sequence data. This classification scheme has since been corroborated by extensive biochemical data that continues to support the existence of this new domain. While considerable progress has been made towards characterizing the Archaea, much remains to be learned. This work contributes to the ongoing characterization of Archaea by exploiting

Sulfolobus solfataricus as a model system for exploring the topics of oxidative stress and thermal virology.

Archaea: Life of the Third Kind

Archaeal species are ubiquitous and well represented in diverse ecological niches. Formerly thought of as primarily extremophiles, Archaeal species are now appreciated as cosmopolitan inhabitants of the planet. PCR-based phylogenetic surveys have placed Archaeal inhabitants in nearly every sampled environment, ranging across the spectrum of extreme habitats (*31-35*) to those of forests and garden soils (*36-39*), the human colon (*40*) and as substantial constituents (~20%) of the world's oceanic plankton community (*41, 42*) (Box 1.1). Until recently, our appreciation of Archaeal physiology has principally been limited to a select group of model extremophiles that could be easily maintained in culture. However, recent advances in culture-independent technologies have provided a greater appreciation of the abundance, diversity and likely the role these unculturable Archaea play in natural microbial ecosystems. Species- and group-specific 16S rRNA-targeted oligonucleotide probes have been developed for *in situ* identification of distinct Archaeal members within the context of any given microbial community, from the most extreme to the benignly mundane

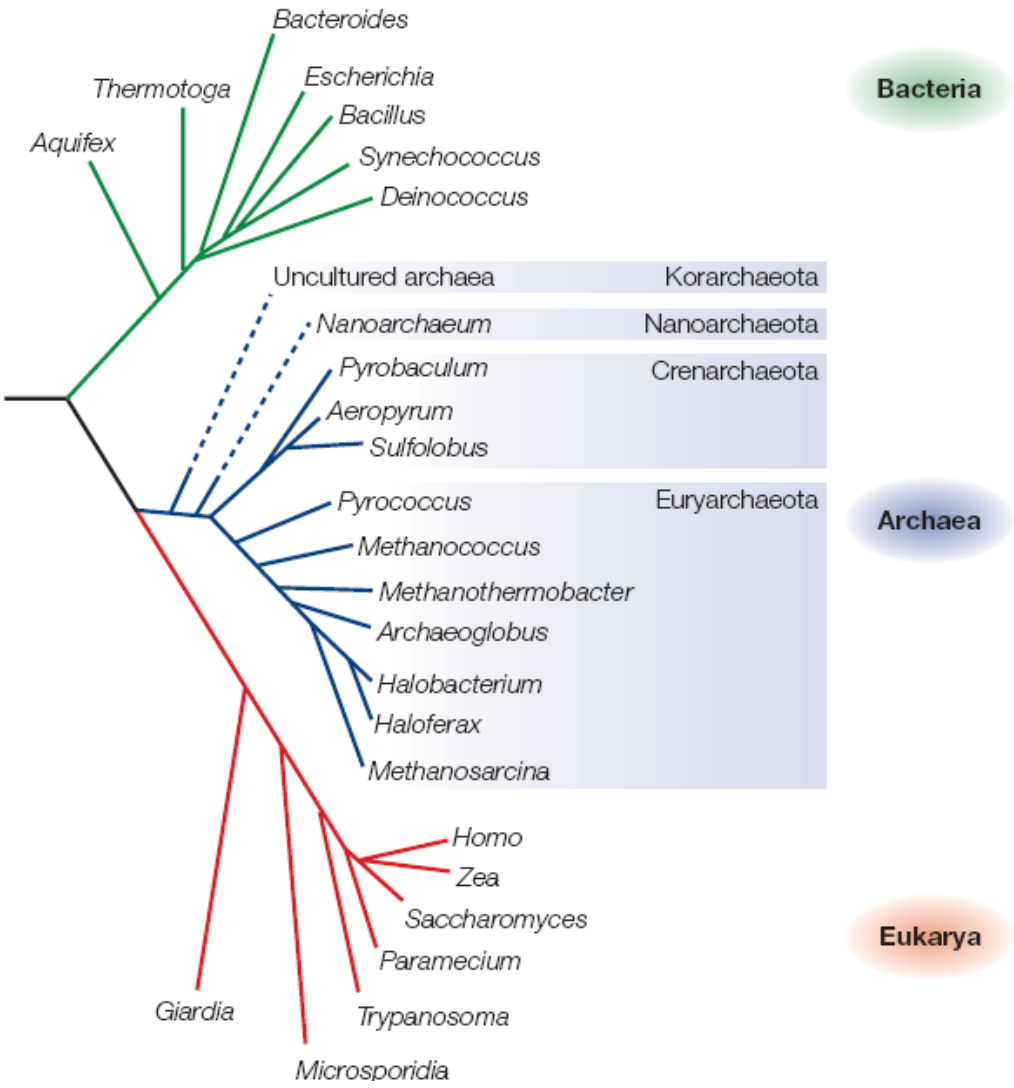


Fig 1.1 Phylogenetic tree of life based on small ribosomal RNA sequence. The three domains of life are presented in Green (Bacteria), Blue (Archaea) and Red (Eukarya). Major phyla (kingdoms) within the Archaeal domain are indicated. Dashed lines indicated uncertain placement of the two proposed kingdoms currently under consideration in the archaeal domain. (Allers *et. al.* 2005)

(REV (*43*). While fluorescent *in situ* hybridization (FISH) allows direct visualization of targeted community members, it offers little insight as to what role these organisms are fulfilling in a given ecosystem. Environmental metagenomic initiatives have, however, succeeded in piecing together substantial regions of uncultivated Archaeal genomes and in the process provided insight into their metabolic potential (*44, 45*). Together, these culture-independent assessments of microbial communities have begun to minimize our dependence on cultureable model organisms. When combined with the benefits of traditional culture based microbiology, we stand to gain a more accurate understanding of the roles that the Archaea play in various ecosystems and their impact on global energy cycles.

Using 16S rDNA sequence information the Archaeal domain can be confidently subdivided into two major kingdoms, the Euryarchaeota and the Crenarchaeota, with two additional kingdoms, the Korarchaeota and the Nanoarchaeota currently under consideration (Figure 1.1). The Euryarchaeota consist of many phylogenetically distinct groups, however culture-based analysis has been primarily limited to methanogens (methane-producing organisms), halophiles (organisms that inhabit high salt environments) and thermophiles (organisms with optimal growth temperatures at or above 50ºC). Founding members of the Crenarchaeota phyla were exclusively extremophiles, most typically those favoring high temperature acidic environments. Cultureable

isolates, most notably those belonging to the orders Thermoproteales, Sulfolobales and the Desulfurococcales, have been developed as model organisms, thereby attracting considerable attention by redefining the limits of life. The outlandish lifestyles of these representatives, thriving in environments previously regarded as sterile, contributed to their reputation as obligate extremophiles. More recently, this stereotype has been upset by extensive culture-independent surveys detecting diverse Archaeal communities in a variety of moderate to low-temperature habitats. PCR-based environmental surveys have contributed considerably to our current appreciation for Archaeal diversity. In fact, the Korarchaeota, a proposed ancient archaeal division, were originally detected by using this approach (Figure 1.1) (*32, 46*). Although efforts to grow organisms from this lineage in pure culture have failed, Korarchaeota of the pJP27 phylotype have now been highly enriched in a laboratory culture at 85ºC. Cells corresponding to the novel Korarchaeote 16S rDNA signature have been identified, by fluorescence hybridization, as rods of 5–10 µm in length (*47*) and the (meta)genome of this mixed culture is being sequenced (*48*).

Continuing interest in microbial biodiversity has lead to an extensive 16S database, a vast majority of this sequence being contributed by PCR-based environmental surveys. This expanding sequence collection affords us the opportunity to continually reevaluate our perception of life with the benefit of new treeing algorithms and refined methods. The final phylogenetic placement

of the proposed Korarchaeota kingdom may serve as a good example of how our perception evolves as more sequence data is evaluated. Recent phylogenetic analysis of Archaeal 16S rDNA signatures, predominantly those derived from environmental sources, indicates that the putative Korarchaeota are a distinct lineage within the Crenarchaeal kingdom rather than a distinct phyla of their own (*49*). Remarkably, accumulating environmental sequence data appears to have unexpectedly collapsed branching in the Archaeal tree, with statistical support for only the two originally proposed Archaeal kingdoms. This is particularly surprising if one considers that the Crenarchaeal and Euryarchaeal kingdoms were originally defined in 1986 on the basis of only a few cultured isolates; this in the absence of extensive environmental data (*50*).

Clearly, culture-independent methods have revolutionized our appreciation of microbial ecology; however it is important to recognize intrinsic limitations of this approach. The discovery of Nanoarchaeota, a recently proposed kingdom of the Archaeal domain, illustrates the limits of PCR-based microbial ecology. *Nanoarchaeum equitans*, currently the only cultured representative of this proposed phylum, was first observed in anaerobic enrichment cultures established from samples collected at a marine hydrothermal system (Kolbeinsey ridge) north of Iceland (*51*). Surprisingly, enrichment cultures designed to selectively enrich for Crenarchaeota belonging to the genus *Igneococcus* contained two distinct morphotypes, one of which was substantially smaller

(~400nm in diameter) than that of a typical cell. The two morphotypes could be separated using either optical tweezers or ultrafiltration (0.45 μm pore size), however only the larger *Igneococcus* cells could be grown in pure culture, notably never giving rise to the morphologically distinct nano-cocci. Initial attempts to amplify small subunit ribosomal DNA using 'universal' 16S primers provided sequence matching only that of *Igneococcus*, suggesting that the small cocci was a cellular derivative or morphotypic intermediate of the larger *Igneococcus* cells. However, these explanations were ruled out by Southern Blot analyses, which revealed two distinct 16S banding patterns, only one of which corresponding to that expected for *Igneococcus* (*51*). DNA fragments corresponding to the new 16S hybridization pattern were cloned and sequenced, revealing a distinct 16S signature not recognized by primers previously thought to be universal. Phylogenetic mapping of the *N. equitans* 16S sequence suggested that this organism was the founding member of a new phylum, termed the Nanoarchaeota, within the Archaeal domain. While 16S primers designed on highly conserved regions of small ribosomal RNA genes have become the standard catch-all approach for assessing microbial diversity; the discovery of *N. equitans* illustrates the limited applicability of PCR-based ecology in discovering novel branches in the tree of life (Box 1.2).

Current State of Culture Based Archaea

Currently, 77 different Archaeal genera - representing more than 230 unique species - are maintained in publicly available culture collections (*52, 53*). Culture based isolates include members from both currently recognized Archaeal kingdoms, the Crenarchaeota and the Euryarchaeota. Members of the proposed Korarchaeota have been stubbornly resistant to pure culture isolation, though co-cultures enriched for Korarcheaote have been established. *N. equitans*, currently the only cultured member of the proposed Nanoarchaeota, is an obligate symbiont and is maintained in a co-culture with its Igneococcal host, however at present this co-culture is not available in any of the public repositories. Thermophiles dominate in Archaeal culture collections and contrary to popular impression, the prevalence of thermophilic isolates is not restricted to the Crenarchaeotal kingdom. Although thermophiles are dominant in both major Archaeal kingdoms, even exclusive among the cultureable Crenarchaeota, there are several non-thermophilic Euryarchaeotal ecotypes (i.e. methanogens and halophiles) maintained in culture. Publicly available microbial culture collections have grown substantially in the past decade and are vital for the ready distribution and continued maintenance of laboratory based cultureable microbes. However, as expansive as these collections are, they represent only a tiny minority (estimated at >0.1%) of the true microbial diversity, as predicted from culture-independent estimates of diversity. From an Archaeal perspective,

a noticeable gap in these collections is the absence of any pure culture isolates of non-thermophilic Crenarchaeota. Culture-independent studies have recognized both abundant and diverse non-thermophilic lineages of Crenarchaeota, however to date we have little appreciation for the role these organisms play in their natural setting. A cultureable representative from this ecotype will almost certainly result in a complete genome sequence, and together these tools will facilitate our understanding of Crenarchaeota in diverse non-thermophilic niches.

Archaeal Genomics: Defining a Domain

The development of high throughput sequencing techniques has given rise to a new era of complete genome sequencing. Aside from the celebrated human genome sequencing projects, these high throughput sequencing efforts have contributed significantly to the field of microbiology. Joint Genome Institute (JGI) integrated microbial genomics page (http://img.jgi.doe.gov/cgi-bin/pub/main.cgi) (updated June 1st, 2006) reports a total of 331 complete prokaryotic genomes. However, only a small fraction, 27 complete and 5 draft sequences, are reported for Archaeal organisms. In 2005, Allers and Mevarech published a comprehensive tally of all complete or ongoing Archaeal genome sequencing projects (*54*). At the time, a total of 44 genomes were listed as ongoing or complete, with an overwhelming majority of these being from

Euryarchaeota. An update of this table less than a year later finds 8 of these previously incomplete or unpublished genomes deposited in the NCBI database (Table 1.1), a promising contribution to the future of Archaeal genomics.

The rapid accumulation of genome sequences has influenced nearly every aspect of biological science and stimulated the creation of a powerful new suite of tools that we collectively refer to using the 'omics' suffix. One such discipline routinely used for discovery is called 'comparative genomics'. Insight into genome evolution, gene function, metabolic potentials and other biochemical pathways can be assessed by comparing whole genome sequences. The first Archaeal genome sequence, that of *Methanocaldococcus jannaschii*, was published in 1996, the fourth complete genome ever released from any domain (*55*). Annotation of the 1.66 Mb genome was facilitated by the then emerging tools associated with the comparative genomics approach and immediately confirmed earlier suggestions that a majority of archaeal genes involved in metabolism and cell division were bacterial orthologs, while those genes associated with information processing systems were strikingly Eukaryal-like. In spite of these findings, ~60% of predicted ORFs were functionally unassignable, using the then available homology-based functional prediction tools. This first glimpse at Archaeal genomics demonstrated the mosaic integration of Bacterial-like and Eukaryal-like components into a single archaeal chromosome and promoted one of the premier archaeal bioinformatisist to comment on how 'the

first sequenced archaeal genome revealed the depth of our ignorance of the biology of this remarkable group of organisms' (*56*).

Table 1.1 Archaea with sequenced genomes or ongoing genome sequencing projects (updated version from *Allers et al* 2005)

Species name	Genome size (Mb) (Date)	Phylum	Growth characteristics and optimal temperature	Genetic potential	Sequence resource
Methanocaldococcus jannaschii	1.66 (1996)	Euryarchaeota	Hyperthermophilic methanogen, anaerobic, 85°C		NCBI
Archaeoglobus fulgidus	2.18 (1997)	Euryarchaeota	Hyperthermophilic, sulphate-reducing, anaerobic, 83°C		NCBI
Methanothermobacter thermautotrophicus	1.75 (1997)	Euryarchaeota	Methanogen, anaerobic, 65°C	+	NCBI
Pyrococcus horikoshii	1.74 (1998)	Euryarchaeota	Hyperthermophilic, anaerobic, 96°C		NCBI
Aeropyrum pernix	1.67 (1999)	Crenarchaeota	Hyperthermophilic, aerobic, 95°C		NCBI
Halobacterium sp. NRC-1*	2.57 (2000)	Euryarchaeota	Halophilic, aerobic, 42°C	+++	NCBI
*Halobacterium salinarum**	~2.5 (2000‡)	Euryarchaeota	Halophilic, aerobic, 42°C	+++	MPG
Thermoplasma acidophilum	1.56 (2000)	Euryarchaeota	Thermoacidophilic, aerobic, 59°C		NCBI
Thermoplasma volcanium	1.58 (2000)	Euryarchaeota	Thermoacidophilic, aerobic, 60°C		NCBI
Pyrococcus abyssi	1.77 (2001)	Euryarchaeota	Hyperthermophilic, anaerobic, 96°C	++	NCBI
Pyrococcus furiosus	1.91 (2001)	Euryarchaeota	Hyperthermophilic, anaerobic, 96°C	++	NCBI
Pyrolobus fumarii	1.85 (2001‡)	Crenarchaeota	Hyperthermophilic, aerobic, 106°C		
Sulfolobus solfataricus	2.99 (2001)	Crenarchaeota	Thermoacidophilic, aerobic, 80°C	+++	NCBI
Sulfolobus tokodaii	2.69 (2001)	Crenarchaeota	Thermoacidophilic, aerobic, 80°C	+	NCBI
Ferroplasma acidarmanus	1.87 (2002§)	Euryarchaeota	Acidophilic, anaerobic, 42°C		ORNL
Methanopyrus kandleri	1.69 (2002)	Euryarchaeota	Hyperthermophilic methanogen, anaerobic, 98°C		NCBI
Methanosarcina acetivorans	5.75 (2002)	Euryarchaeota	Methanogen, anaerobic, 35°C	+++	NCBI
Methanosarcina barkeri	4.86 (2002§)	Euryarchaeota	Methanogen, anaerobic, 35°C	+++	NCBI ±
Methanosarcina mazei	4.10 (2002)	Euryarchaeota	Methanogen, anaerobic, 37°C	+++	NCBI
Pyrobaculum aerophilum	2.22 (2002)	Crenarchaeota	Hyperthermophilic, nitrate-reducing, aerobic, 100°C		NCBI
Hyperthermus butylicus	1.67 (2003‡)	Crenarchaeota	Hyperthermophilic, sulphate-reducing, anaerobic, 100°C		
Methanogenium frigidum	~2.5 (2003§)	Euryarchaeota	Psychrophilic methanogen, anaerobic, 15°C		UNSW
Nanoarchaeum equitans	0.49 (2003)	Nanoarchaeota	Symbiotic hyperthermophile, anaerobic, 90°C		NCBI
Sulfolobus acidocaldarius	2.23 (2003‡)	Crenarchaeota	Thermoacidophilic, aerobic, 80°C	++	NCBI ±
Haloarcula marismortui	4.27 (2004)	Euryarchaeota	Halophile, aerobic, 37°C	+	NCBI
Haloferax volcanii	4.03 (2004‡)	Euryarchaeota	Halophile, aerobic, 45°C	+++	UMBI; TIGR
Methanococcoides burtonii	2.56 (2004§)	Euryarchaeota	Psychrotolerant methanogen,		NCBI ±
Methanococcus maripaludis	1.66 (2004)	Euryarchaeota	Methanogen, anaerobic, 37°C	+++	NCBI
Methanococcus voltae	~1.9 (2004‡)	Euryarchaeota	Methanogen, anaerobic, 37°C	+++	
Natronomonas pharaonis	2.75 (2004‡)	Euryarchaeota	Haloalkaliphilic, aerobic, 40°C		NCBI ±
Picrophilus torridus	1.55 (2004)	Euryarchaeota	Acidophilic, aerobic, 60°C		NCBI
Thermococcus kodakaraensis	2.09 (2004)	Euryarchaeota	Hyperthermophilic, anaerobic, 85°C	++	NCBI ±
Thermoproteus tenax	~1.84 (2004§)	Crenarchaeota	Hyperthermophilic, anaerobic, 86°C		REF. 44
Acidianus brierleyi	~1.9	Crenarchaeota	Thermoacidophilic, aerobic, 70°C		
Halobaculum gomorrense	~2.7	Euryarchaeota	Halophilic, aerobic, 37°C	+	
Haloquadratum walsbyi	~3.18	Euryarchaeota	Halophilic, aerobic, 40°C		NCBI ±
Halorubrum lacusprofundi	~2.6	Euryarchaeota	Psychrotolerant halophile, aerobic, 30°C	+	
Natrialba asiatica	~3.1	Euryarchaeota	Halophile, aerobic, 37°C	+	
Sulfolobus metallicus	~1.9	Crenarchaeota	Thermoacidophilic, aerobic, 80°C		
Cenarchaeum symbiosum		Crenarchaeota	Symbiotic psychrophile, aerobic, 10°C		
Methanococcus thermolithotrophicus		Euryarchaeota	Thermophilic, halotolerant methanogen, anaerobic, 62°C		
Methanosaeta concilii		Euryarchaeota	Methanogen, anaerobic, 37°C		
Methanosarcina thermophila		Euryarchaeota	Thermophilic methanogen, anaerobic, 50°C		
Methanosphaera stadtmanae		Euryarchaeota	Methanogen, anaerobic, 37°C		NCBI ±
Methanospirillum hungateii		Euryarchaeota	Methanogen, anaerobic, 37°C		NCBI ±

Date indicates either completion or publication of genome sequence (no entry indicates a continuing genome project). **Halobacterium* spp NRC-1 and *Halobacterium salinarum* genome sequences are essentially identical. ‡The genome sequence is complete but not published. §The genome sequence is published but remains incomplete. +, growth on solid media; ++, potential for rudimentary genetics, such as transformation and selectable markers; +++, potential for advanced genetics, including shuttle vectors, gene replacement and reporter genes. MPG, Max Plank Gesellschaft; NCBI, National Center for Biotechnology Information; ORNL, Oak Ridge National Laboratory; TIGR, The Institute for Genomic Research; UMBI, University of Maryland Biotechnology Institute; UNSW, University of New South Wales (for URLs, see Online links box).

The genomescape has expanded considerably since the first published Archaeal genome and now includes genetic blueprints from phylogenetically and physiologically diverse life. Increasingly sophisticated computational methodologies have provided an efficient means for genome annotation and functional comparison. Archaeal genomics has served a central role defining this domain of life, largely by employing a gene classification scheme that is based on orthologous genes. Orthologs are homologous genes that evolved from a common ancestral gene that has been separated by speciation events. Phylogenetic information from ancestrally related (Orthologous) genes thus provides a framework for tracking genome evolution. Furthermore, because orthologous genes typically retain function, putative functional assignments can be extended from a single gene to an entire Cluster of Orthologous Genes (COG), which rapidly contributes to the annotation of poorly characterized genomes.

Comparative genomics had been used to identify a common set of approximately 300 COGs expected to be conserved in all Archaeal genomes. Release of the 5th Archaeal genome, that of N. equitans has stripped the core Archaeal gene set to include only 200 COGs (*56, 57*). Given the physiologic diversity and metabolic breadth of organisms in the Archaeal domain it is not expected that many of these genes account for common metabolic pathways, but rather, are more likely involved in fundamental tasks related to replication

and gene expression. Sixteen of these ~200 COGs are found exclusively in Archaea and thus define the minimal genomic signature of organisms belonging to the Archaeal domain of life.

COG analysis illustrates the evolutionary link between the Archaeal and Eukaryal information processing systems. Sixty-one of the 'core' 200 COGs are exclusively shared between Archaea and Eukarya. It is worth noting that only two of these sixty-one COGs (COG1936, a nucleotide kinase, and COG3642, a protein kinase typically fused to a metalloprotease domain) are not expected to be involved with the informational processing apparatus (*56*).

Horizontal gene transfer can confuse efforts to trace evolution by shuffling related genes between unrelated genomes. However, it has been suggested that core genes that are involved in fundamental informational processing systems are not disposed to non-orthologous gene displacement. Said another way, informational processing genes that commonly participate in concert with various other proteins to perform a fundamental operation are generally less likely to be maintained in a chromosome where participants in the analogous process are fundamentally different.

Prokaryotes typically have a single chromosome of condensed circular double stranded DNA. Like Bacteria, Archaea regularly organize functionally related genes in a series know as an operon. Expression of the gene cluster is generally coordinated by a single (sometimes a couple) regulatory element. This

realization has facilitated the annotation of Archaeal genomes. Several conceptually different bioinformatic methods have been developed which consider local genomic context of a gene. The simplest of these allows visual inspection of the gene in the context of surrounding genes, offering insight into order, size and orientation of genes in the neighborhood. Co-regulated members of the operon are expected to interact physically (protein-protein interaction) or physiologically (in a common pathway) and thus homology-based functional prediction tools that are able to confidently assign function to a single gene in an operon can then be extended to assign general function to all gene members in a particular operon. Although composition of operons are sometimes evolutionarily variable, the observed preservation of an operon across phylogenetically diverse organisms can be used to further extend functional assignments in other genomes. Several levels of complexity have been added to the concept of functionally associated genes and package in an interactive web-based server called STRING (search tool for recurring instances of neighboring genes) (*4, 58, 59*). In addition to finding spatially linked gene clusters, this program looks beyond local context assuming that functionally related genes will be either present together or absent together across multiple genomes regardless of context. An extension of this idea rooted in the concept of non-orthologous gene displacement, is that non-homologous genes (evolved independently) can serve the same essential cellular function and thus when one

gene set is present in a given genome, it is unlikely that another unrelated gene set with redundant function will also be present (*60*). Examples of how these tools have been implemented for assessing the oxidative stress response are provided in Chapter 10.

For perhaps the most exciting example of how gene context is contributing to genome annotation we must again be reminded of the evolutionary link between Archaea and Eukarya. In Eukaryotes gene regulation schemes are forbiddingly complex and rarely are context clues of any use in predicting function. However, by exploiting contextual information in Archaea we can extend putative functional information to orthologous genes in eukaryotes, thus Archaeal genome annotation tools are beginning to prove useful in further understanding fundamental Eukaryotic biology. An example of this 'reverse genomics' approach includes new discoveries recently made the in DNA replication apparatus of *Sulfolobus solfataricus*. Some of the proteins that participate in *S. solfataricus* DNA replication exist in still poorly characterized operons. However based on context, general function can be suggested for all members in the operon and this information can be used to direct biochemical validation. Although still in the early days, contextual information of *S. solfataricus* DNA replication genes has been used to implicate orthologous genes in the human genome and are currently being characterized (Bell, personal communication).

Genome sequence efforts have clearly provided a scaffold for understanding Archaeal biochemistry and in so doing provided insights into the other domains of life. However, much of our knowledge about Archaea is fragmented, with far too many ORFs assigned as hypothetical proteins. Understanding the idiosyncrasies of this domain requires us to step beyond the limits of genomics and into a tractable genetic system(s) where the *in vivo* function of any gene can be easily evaluated.

Developing Genetic Systems

An ideal model organism must be amenable to laboratory cultivation, where it should be easy to grow to high cell density in a reasonably short period of time in liquid culture as well as on a solid media. In addition to these culture based concerns, modern molecular biology relies on a complete genome sequence. From the vast abundance of microbial diversity Archaeal biologists are restricted to a select list of Archaea with available genome sequences (Table 1). A practical model must also be amenable to genetic manipulation. More specifically, genes of interest should be easy to knockout, disrupt or silence, while at the same time homologous, heterologous or mutant genes should be easy to introduce; preferably under the control of an inducible regulator.

The search for an ideal model that is easy to grow and facile to genetic manipulation has been a major stumbling block in the progression of Archaeal

biology. Rather than a single model that offers all the above mentioned characteristics (i.e. the *E. coli* of Bacteria), there are currently several organisms, each offering a subset of attractive features. *Halobacterium halobium* was among the first archaeon to be developed for transfection with naked DNA from ΦH (*61*). Previous efforts to track transformation efficiency in Archaeal spp. had been frustrated by the lack of a selectable marker. In the absence of well established selectable markers, which has been the foundation of Bacterial genetics (Box 1.3), Archaeal biologists resorted to following ΦH transfection efficiencies by plaque assays (*61*). This polyethylene glycol-mediated transfection system produces between 10^6 and 10^7 transfectants per μg of DNA, representing the first opportunity to introduce exogenous DNA into any Archaea. Although immediately recognized as a valuable system and adapted to several Halobacterium spp., applicability of this system was largely dependent on spheroplast formation. Cells resistant to the removal or subsequent regeneration of the paracrystalline glycoprotein surface layer (S-layer) have not been receptive to this procedure. The limited versatility of this method has driven the development of alternative transformation systems.

A genetic system for Methanogenic Archaea has been of particular interest given their unique metabolic capacity to convert simple C1 and C2 compounds to methane (CH_4). However, development of a genetic system for Methanogens as been hampered by their extreme oxygen sensitivity, which necessitates elaborate

methods and equipment for their cultivation and manipulation. Further, some methanogens have been resistant to spheroplast formation, a result that originally discouraged the development of an early transformation system. However, in 1991 Gernhardt *et al* reported a major advance in methanogen genetics. A liposome mediated delivery system was used to transform *Methanococcus voltae* with a non-replicating plasmid carrying a puromycin resistance gene (*pac*) (*62*). This strategy has been further developed to include several reporter genes and a *his*A gene to complement histidine auxotrophic mutants in several Methanococcus species (*63*). These tools have been used to identify the components associated with flagellum biosynthesis, in studies related to chromosome partitioning during cell division and the role of histone-like proteins in Archaea (*64-66*). More elaborate tools, based on multiple selectable markers, have since been developed for *Methanosarcina* species, where they have been used to identify and evaluate the versatile substrate capacity of these methanogens (*67, 68*). More recently the need for selectable markers has been minimized by the development of a 'markerless' genetic exchange method (*69*). This strategy relies on the complementation of an *hpt* mutant strain of *Methanosarcina acetivorans* with a single artifical operon that carries both the puromycin resistance gene and a competent *htp* gene, which confers resistant to the toxic base analog 8-aza-2,6-diamino-purine. This cassette is flanked by flip recombinase recognition sites and by sequence homologous to that flanking the

targeted chromosomal gene. Insertion of this construct into the chromosome (targeted by homologous sequence flanking the construct and in the chromosome) replaces the target gene with the *pac-htp* resistance cassette. Following selection, the cassette can be removed by introducing a non-replicating plasmid that carries the flp recombinase gene under the control of a Methanosarcina promoter. The recombinase excises the selection cassette at the engineered flip sites. The cassette can then be recycled to target another gene by engineering new targeting sequences without the need for additional selectable markers. Multi-loci knockouts have been of particular importance in *Methanosarcina* species where a single genome often contains several copies of a homologous gene. Thus the phenotype of a single gene deletion can be masked by compensatory effects of a closely related gene. Though this may be rare, as highly related genes often serve different function, gene duplication can complicate experimental interpretation.

Still another variation of this approach permits investigation of 'essential genes', where knock-outs are lethal. This strategy, again based on homologous recombination, fuses the gene to a controllable promoter that can serve as an on/off switch (or as a rheostat in some instances) that can be used to precisely define the role of an essential gene at any stage in the growth cycle.

Clearly genetic systems in Methanogenic Archaea have matured, overcoming many of the early technical difficulties associated with the genetic manipulation

of strict anaerobes. Today methanogenic biologists are armed with a wide breadth of genetic tools that have been deftly implemented to assess several aspects of methanogenic metabolism, chromatin structure, cell cycle regulation, amino acid biosynthesis and other fundamental biological processes.

In contrast to the well developed genetic systems available for methanogens, genetic manipulation strategies for hyperthermophilic Archaea are just beginning to emerge as a viable option. The relatively small genome size of hyperthermophiles (as compared to that of most mesophiles) is expected to reflect a simple biochemical model, which would offer straightforward interpretation of genetic mutants. Further, interest in hyperthermophiles has been keen since they have been shown to occupy the deepest and shorts branches in most phylogenetic trees, which has lead some to speculate on the antiquity of their biological apparatus (*70*). Aside from fundamental interest in hyperthermophilic life, these organisms also offer substantial biotechnical and industrial opportunities where stable enzymes from thermophiles have been commercially developed (*71, 72*). The exploitation of these interests has been hampered by the paucity of readily available genetic tools. The grim status of genetics in hyperthermophilc Archaea has recently been revived, by the rapid development of a robust targeted gene disruption system for *Thermococcus kodakaraensis (73, 74)*. *T. kodakaraensis* is a sulfur-reducing hyperthermophilic *Euryarchaeota* that is 'naturally' competent to exogenous DNA transformation.

The first iteration of the *T. kodakaraensis* genetic system was based on a random, UV induced mutation in the *pyr*F gene (*75*). This auxotrophic mutant, strain KU25, was used as a hook to target the disruption of both *trp*E and *his*D, yielding reliable amino acid auxotrophs (*74*). This dual selection strategy was exploited to develop a pop-out recombination scheme, similar to that described for methanogens, which affords the opportunity to target multiple loci in a single construct for more advanced functional analyses of genes *in vivo*.

As a whole, far fewer genetic systems exist for Crenarchaeota. In fact, to my knowledge *Sulfolobus solfataricus* represents the sole opportunity for genetic manipulation in this kingdom. *Sulfolobus* sp. were first described in 1972 by Thomas Brock as a new genus of sulfur oxidizing bacteria resident in the geothermal hot springs of Yellowstone National Park (*76*). Although the original taxonomic classification of Sulfolobus has been redressed according to our new appreciation for the three domains of life, this original discovery has proven to be cornerstone in Archaeal biology. *Sulfolobus sp.* are motile thermoacidophilic Crenarcheaota that grow optimally at temperatures between 75-85ºC, within a pH range of 2.5-3.5 and are common inhabitants of high temperature acidic environments around the world. The ubiquitous distribution, easy of cultivation and the availability of 3 complete genome sequences have contributed to prominence of *Sulfolobus* as models among the Crenarchaeota.

Arriving at Sulfolobus as the Model System of Choice

We have selected *Sulfolobus solfataricus* as a model organism for several practical reasons. *Sulfolobus* spp. are aerobic heterotrophs that can be easily cultured to high cell densities (~10^{10} cell/ml) in either batch or fermentor systems, with a reported doubling time of four hours (*77*). The ease at which *Sulfolobus* spp. can be adapted to the laboratory setting has facilitated its development as a model organism. The first *Sulfolobus* genome (*S. solfataricus*, strain P2) was published in 2001 (*78*) and a second genome, that of the closely related *S. tokodaii* immediately followed (*79*). Last year the *Sulfolobus acidocaldarius* genome was published, completing the trifecta and providing an unusual opportunity for comparative genomic analysis (*80*). Genome information for both *S. acidocaldarius* and *S. solfataricus* has since been used to develop cDNA microarrays which are just beginning to provide insights into global gene regulation strategies (*81*). In addition, to these three complete genome sequences; a consortium including researchers at Montana State University has initiated genome sequencing of 10 additional Sulfolobus strains isolated from around the world. These new sequences will truly provide an unprecedented opportunity for genomic comparison and in-so-doing are expected to provide new insight into nearly every aspect of Archaeal biology, from conserved pathways, to rates and mechanisms of genome evolution.

Archaeal genome sequencing efforts did not start with the relatively formidable host chromosome, but rather on the small chromosomes of associated extra-chromosomal elements. In fact the first Archaeal genome sequence published was that of the Sulfolobus spindle-shaped virus 1 (SSV1) (*82*). The novel morphology of this virus particle (spindle-shaped viruses are only found in the Archaeal domain) along with its unusual genetic signature, heightened an already keen interest in thermal virology. Thermal virologists studying *Sulfolobus* have not been disappointed, as this host appears to harbor an unusually high number of extra-chromosomal elements, ~15 of which have been sequenced and are publicly available (http://dac.molbio.ku.dk/dbs/SRSR/). Some of these genetic elements, most notably those based on SSVs, have been developed as genetic tools (*83-86*). SSV-based shuttle-vectors are based on a fusion between a common *E. coli* plasmid (pBR05, pUC18 or pCRII) and the SSV genome. These shuttle-vectors, together with the development of an efficient transformation system for *Sulfolobus* allows for routine genetic manipulations to be done in *E. coli,* while *in vivo* assessment of these manipulations can be made in the natural host. Recently Albers *et al* refined this self-spreading viral-based shuttle-vector system, to include several inducible promoters, along with putative transcriptional terminators. This shuttle-vector includes both a *lacS* reporter gene and a *pyrEF* (uracil auxotroph) selectable marker (*84, 87*). This system has proven useful in the heterologous and homologous expression of His or Strep

tagged proteins in *Sulfolobus*. In addition to this versatile expression system, *Sulfolobus* also boasts a genetic knock-out system (*88, 89*). Unfortunately this system, which is based on homologous recombination, has not proven effective in any sequenced isolate of *Sulfolobus* and published examples of this technique have been limited to a single, uncharacterized strain of *S. solfataricus* referred to as 98/2 (*88, 89*). Although we can only speculate, it seems reasonable that this strain may possess additional recombination mechanisms that render it more conducive to this particular gene disruption strategy. One note of optimism comes from recent experiments done in the Van der Oost lab (personal communication) where RNA isolated from the 98/2 strain was used on *S. solfataricus*, stain P2, microarray platform. Though detailed analysis is forthcoming, preliminary data suggest that the 98/2 strain is highly related to *S. solfataricus* strain P2 and that the already developed *S. solfataricus* (P2) microarray may serve as an adequate segregate platform for *S. solfataricus*, stain 98/2, gene expression studies.

The development of *Sulfolobus* as an important model organism is complemented by our proximity to Yellowstone National Park (YNP). Montana State University is ideally located next to our nation's first national park, which harbors more than 10,000 unique thermal features. *Sulfolobus sp*. are common inhabitants of the abundant acidic hot springs found in YNP and our proximity has allowed us to approach important ecological studies that require high

sampling frequencies. Further, our interests in thermal virology are perfectly matched with the unusually high number of virus-like-particles associated with *Sulfolobus* hosts. These interests, particularly those focused on viral discovery and viral ecology, have also benefited from our ability to frequently access the diverse thermal features found in YNP.

Overview

In this dissertation, I use the P2 strain of *S. solfataricus;* as a model for studying two distinct aspects of archaeal biology; virology and oxidative stress. The first section of this work (Chapters 2-6) is dedicated to topics related archaeal virology. Major question addressed in this section include; what is the phylogenetic relationship of *Fuselloviruses* isolated from four geographical distinct locations (Chapter 3)? Highly conserved regions, identified in this genomic analysis have been exploited as SSV specific markers that are implemented in a two year phylogenetic study amied at assessing viral diversity within and between high temperature hot springs in YNP (Chapter 4).

A new interest in the lab is focused on understanding how prokaryotes respond to viral infection. In chapter 5, I use northern bolts and *in silico* analysis to address questions related to how prokaryotes defend themselves against viral infection. The emerging hypothesis is that host encoded small RNAs that are

homologous to known viral sequences are used in an immune response analogous to RNA interference systems discovered in eukaryotes.

The second section of this dissertation (chapters 7-10) is dedicated to the characterization oxidative stress defense networks in Archaea. This section begins with an introduction to cage-like proteins from the ferritin-like diiron superfamily that are designed to managed the mingled toxicity of iron and oxygen. This interest in antioxidant proteins was born from the discovery of a novel class of antioxidant proteins that we have termed, DPSL (chapters 8 and 9). The discovery of these antioxidants originated with questions related to how Archaea, particularly *S. solfataricus,* manage the paradoxical role that iron plays in biology. These finding have spurred our interests iron and oxidative stress response networks. The microarray experiment described in chapter 10 is directed at identifying components of the oxidative stress defense network in *S. solfataricus*.

Box 1.1 Evolution of an ideal pathogen.

Give the abundance and cosmopolitan distribution of Archaea it is hard to imagine that organisms for this domain have not assumed a pathogenic role in humans. Archaea are capable of establishing long-term colonization and coexistence in the human colonic (*90*), vaginal tract (*91*) and in the oral cavity (*92, 93*), which demonstrates their ability to colonize the human host. Although these relationships have not been clearly defined, they appear to be part of the normal flora rather than active pathogens. Although, to date, no Archaea have been recognized as the causative agent of any disease, there are few if any biologically rational explanations for precluding Archaeal organisms in this evolutionarily predictable behavior. Several independent efforts have recently 'gone looking' for Archaeal pathogens. One such effort has lead to the development of an online database that uses basic alignment tools to search human EST (expressed sequence tags) libraries for Archaeal sequences (*94*). Most efforts however, have focused on PCR-based detection of Archaea from clinically isolated samples. To date, neither approach has been able to implicate Archaea in human disease.

In many respects Archaea would seem to be ideally suited eukaryotic pathogens. The compatibility of related Archaeal/Eukaryal information processing systems would allow an Archaeal pathogen to easily co-opt Eukaryotic components for replication of its own chromosome. The similarity between these two systems is comparable to that observed for viral pathogens. One noteworthy observation is that the ether linked polar lipids isolated from Archaea have been shown to serve as potent immune adjuvants, both *in vitro* and *in vivo* (*95*). Immunogenicity of the unique components of the archaeal cell wall may reflect an inherent hurdle for establishing an infection; however it is hard to imaging that the selective pressures of evolution would not find a clever way around the immune response. The discovery of an archaeal pathogen would certainly provide insights into the shared evolutionary trajectories of organisms from these two domains and recruit a broad scientific interest that would certainly serve to intensify the characterization of organisms in this domain.

Box 1.2 Plight of the Nanoachaeota

Nanoarchaeum equitans is an obligate parasite of *Igneococcus* sp. In culture the tiny nanococci decorate the surface of Igneococus cells, were it is presumed to siphon energy from the metabolic activities of its host cell. The complete genome sequence of *N. equitans* reflects both its stature and its parasitic life style. The *N. equitans* genome is the smallest microbial genome sequenced to date (490,885 bp) encoding a predicted 537 genes. Almost all of the *N. equitans* genes with predictable functions are involved in information processing systems, while those coding for conserved metabolic pathways are conspicuously missing. Although biochemical evidence is forthcoming it is interesting to note that in the absence of intact metabolic pathways the parasite does code for a variety of putative "peptidases, transporters and uncharacterized membrane proteins that could be involved in scavenging substrates, coenzymes, membrane components and other molecules from Igneococcus" (*57, 96*); KS Makarova, unpublished). Another peculiarity of the *N. equitans* genomes is the dearth of conserved operon structure. In fact the fragmentary nature of this genome extends to the gene level where an unusually large number of split genes have been observed. In light of the *N. equitans* parasitic life style, it is worth noting that split genes were first characterized by Roberts and Sharp in adenovirus, a discovery that earned them a Noble Prize in Medicine and Physiology in 1993. Although protein coding split-genes are clearly not unique to *N. equitans*, they do appear to be overrepresented in the genome of hyperthermophiles (*97*). The fragmented nature of the *N. equitans* genome has been cited as evidence of its primal origin (*96*). Another, perhaps more conservative interpretation of this observed split gene frequency, is that spilt genes offer a more efficient means of packing functional content onto a single compact chromosome. That is to say that split genes, whose gene products encode individual conserved domains, could have evolved as a molecular lego set whereby unlinked conserved domains can be mixed-and–matched combinatorially to achieve diverse protein function. This explanation would be consistent with their discovery in viruses and their prevalence in the smaller compact genomes of hyperthermophiles.

Although protein coding split genes are not unique; *N. equitans* offers the first example of split tRNA genes (*98*). Interestingly, each of the split tRNA genes is truncated at a location downstream of the anticodon at a position in which other archaeal tRNAs often contain introns. Joining of tRNA halves is thought to be mediated by a 12-14 nucleotide GC region that is later resolved by a heretofore uncharacterized trans-splicing reaction (*57, 96, 98, 99*).

Clearly the remarkable discovery of *Nanoarchaeum equitans* has given a boost to comparative genomic and evolutionary biology, as well as greater insight into the nature of parasitic Archaea. However the ultimate placement of this linage in the tree of life is less certain. In fact, new phylogenetic

assessments of the *Nanoarchaeum equitans* using a variety of genes suggests that *Nanoarchaeum* represent a distinct linage within the Euryarchaeota, cladding most frequently with the Thermococcales (*100*). Interestingly, this phylogenetic pattern was observed for both the helicase and the topoisomerase domains of reverse gyrase, which is currently the lone hallmark of hyperthermophily (*101*).

Box 1.3 Antibiotic sensitivity of Archaea.

Antibiotics, a term often used generically to describe antibacterial activities. Antibiotics were originally discovered by Alexander Fleming in 1929, when he noted the antibacterial affect of a fungal contaminant (*Penicillium notatum*) on an agar plate of *Staphylococcus aureus*. Since this original discovery a suite of antibacterial compounds have been commercially developed and widely applied in modern medical applications. This collection of antibiotics consists of two major categories, cell wall synthesis inhibitors and protein synthesis inhibitors. Give the distinction between Archaeal and Bacterial cell walls (namely the absence of murein and ester linked phospholipids) it is not surprising that many of these antibacterial compounds are ineffective for Archaea. Furthermore, many clinically developed antibiotics of the later category, owe their action to the inhibition of some step in the complex process of protein synthesis. Given the observed kinship between the Archaeal and Eukaryal information processing apparatus it should not be suppressing that many clinically relevant antibiotics are generally not affective inhibitors of Archaeal protein synthesis. The limited availability of well developed anti-archaeal drugs is compounded by the limited stability of most drugs/compounds when exposed to the extreme growth conditions of many model Archaea. The paucity of effective anti-archaeal drugs has hampered the development of antibiotic based genetic systems in Archaea. However, like some bacterial species, Archaea have also been reported to produce antimicrobial agents. Several Archaeal species have been isolated that producing proteinaceous toxins that selectively inhibit, or kill closely related strains (*102-106*). These toxins are expected to be competitive mechanism targeted at limiting nutrient competition. In Sulfolobus these toxins are 20 kDa proteins that are released from the cell in membrane vesicles. When toxins isolated from membrane vesicles were spotted on lawns of susceptible strains the sulfolobicins cause noticeable growth inhibition halos (*103*). Although these studies are encouraging, the precise composition of these toxic peptides has not been determined. Furthermore, the genes encoding 'toxic peptides' synthesis as well as those encoding the resistance apparatus have not been identified Clearly the occurrence and fundamental nature of these compounds deserve considerably more attention before they can be developed as selectable marker systems.

CHAPTER 2

HOT ARCHAEAL VIRUSES REVEAL DEEP EVOLUTIONARY CONNECTIONS

Abstract

The discovery and characterization of archaeal viruses has provided insight into the fundamental biochemistry and evolution of the Archaea. A wide diversity of viruses have been identified within the Archaeal domain, many of these originating from the high temperature acidic environments like those found in Yellowstone National Park. Archaeal viruses are often morphologically unique and code for genes with little similarity to other known genes in the biosphere, a characteristic that has complicated efforts to trace their evolutionary history. Comparative genomics combined with structural analyses indicate that the spindle-shaped virus lineages might be unique to the Archaea, whereas other icosahedral viruses might share a common lineage with viruses of Bacteria and Eukarya. These studies provide insights into the evolutionary history of viruses in all three domains of life.

Viral Prospecting

Viruses of Archaea were fortuitously discovered in 1974, prior to broad recognition of Archaea as a distinct domain of life. This virus (Hh1) was originally described as a 'spontaneously occurring temperate phage' in *Halobacterium salinarium* cultures, however little characterization beyond this

observation was offered (*107*). Progress in the field continued to be slow until the early 1980's when Wolfram Zillig and colleagues contributed a series of publications defining various aspects of the ΦH life cycle (also a virus of *H. salinarium*), which included a partial genome sequence, biochemical characterization of select proteins and insights into viral gene expression (*108-112*). The 59-kb linear dsDNA genome from this virus was eventually used to develop a transformation system in *H. salinarium* that could be easily followed by plaque assays (*61*). Accumulating examples of halophilic viruses in culture were all of the head-and-tail type morphology, which contributed to a building consensus that this morphotype must be exclusive or at least dominate in this ecotype. This developing paradigm was upset in 1997 by an environmental study that looked directly at VLPs isolated from the Dead Sea (*113*). Viruses in this hypersaline environment proved to be startlingly abundant (10^7 per ml) and commonly spindle-shaped. The following year, Bath and Dyall-Smith reported the isolation and characterization of a lytic (with titers of up to 10^{11} PFU/ml), spindle-shaped virus infecting the halophilic Euryarchaeote, *Haloarcula hispanica* (*114*). The spindle-shaped His1 virion is 44 x 74 nm and packages a linear dsDNA chromosome 14.9 kb in length. It is hard to reconcile the apparent dominance of spindle-shaped virus particles in this hypersaline environment with their limited representation in culture; however viral culture biases have been reported (*22*).

Although, still in the early days, archaeal virology has already made several profound contributions. Transcript maps of the Sulfolobus spindle-shaped virus 1 (SSV1) revealed a consensus transcriptional promoter sequence unlike that found in Bacteria, but surprisingly similar to that recognized by the DNA-dependent RNA polymerases (pol II) in Eukaryotes (*115*). This discovery was consistent with prior discoveries made by Zillig and colleagues, who recognized striking similarities between the multi-subunit structures of archaeal and eukaryal RNA polymerases (*116*). These two findings furthered the credibility of Woese's, then fledgling, three domain hypothesis, and was critical to the broad scientific recognition of Archaea as a distinct domain of life. The novelty of archaeal viruses and the fundamental biological insights they might reveal, continues to drive the field of archaeal virology.

Today the study of viruses infecting Archaea is a rapidly evolving field, with new viruses being isolated at an increasing frequency (*117*). Approximately 50 archaeal viruses have been described so far (see Universal Virus Database of the International Committee on Taxonomy of Viruses), a majority of these within the last decade. Roughly equal numbers of viruses have been isolated from the two major Archaeal kingdoms, the Euryarchaeota and the Crenarchaeota. A majority of these viruses replicate in either hypersaline or thermophilic hosts, which is surprising given the attention paid to methanogenic archaea. The paucity of currently characterized methaonogenic viruses is not likely to reflect their limited

abundance, but rather represents an untapped opportunity for discovery. So far, no viruses have been isolated from the Korarchaeota or the Nanoarchaeota, two recently proposed phyla of the Archaea.

Crenarchaeal Viruses

Although viruses of thermophilic Archaea have recently received considerable attention (14 have been isolated within the past 5 years) detailed biochemical characterizations are still only in their infancy. This point is punctuated by comparison to the approximately 5,100 viruses known to infect bacterial and eukaryal hosts, some of which have been studied for more than a century and have been developed as models for investigating nearly every aspect of virology. However, like viruses from Bacteria and Eukarya, which boast a long history steeped in fundamental biological discovery, Archaeal viruses are also expected to offer new insights into the underpinnings of Archaeal biochemistry.

Our current knowledge of thermal viruses comes from cultured isolates of archaeal hosts. A majority of the 21 thermophilic archaeal viruses isolated to date have come from cultures established from terrestrial hot springs and infect members of the Crenarchaeota, with one thermal archaeal virus isolated from a thermophilic marine euryarchaeote (*118*). Viruses infecting only four genera of crenarchaeal hosts have been isolated (*Acidianus, Sulfolobus, Thermoproteus, Pyrobaculum*); reflecting only a small fraction of the thermophilic archaeal

community identified by PCR based phylogenetic surveys. Viruses of organisms recalcitrant to culture have historically been intractable; however the development of culture independent methods have recently offered new potential (Box 2.1).

One of the most striking characteristics of crenarchaeal viruses is that they possess unusual and diverse particle morphologies (Figure 2.1) (*10*). Several morphotypes have been described from thermophilic Archaea, rod-shaped particles which resemble some RNA plant viruses (*119*), a 100-185 x 70-95 nm droplet-shaped particle with fibers that form a beard-like structure (*120*), a 230 x 75 nm bottle-shaped particle (*121*) and spherical particle morphologies (*122, 123*). Two additional morphotypes, the spindle-shaped (*Fuselloviridae* and *Bicaudaviridae*) (*18, 124, 125*) viruses and the icosahedral virus STIV (*126*),

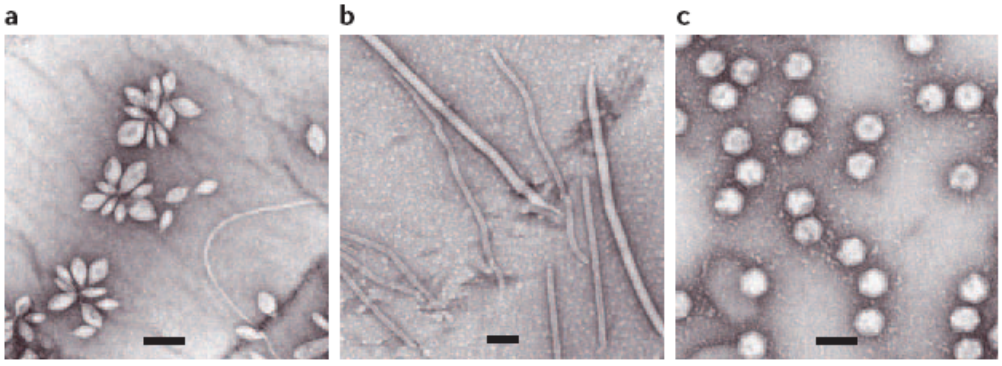


Fig. 2.1 Typical virus morphologies observed in both environmental samples and enrichment cultures established from hot, acidic environments. a) Spindle-shaped virus morphology is restricted to the Archaea. b) Rod and filamentous morphologies are common in both bacteria and archaea. c) Icosahedral viruses are common in all three domains of life. Scale bars denote 100 nm.

have also been described. The phylogenetic distribution of the latter two virus morphologies suggests two very different evolutionary histories.

The *Sulfolobus* spindle-shaped virus 1 (SSV1) was the first crenarchaeal virus to be isolated (*125*). This virus and other members of the *Fuselloviridae*, including SSV2 (*127*), SSV K1 and a YNP isolate SSV RH (*18*), all share a 60 x 90 nm spindle-shaped morphology with flexuous tail fibers that extend from one end. All isolated spindle-shaped viruses (SSVs) infect *Sulfolobus* spp. Although all SSV isolates share the same morphology, variations of this morphotype have been reported. An unusually large (107 x 230 nm) SSV-like virus, the *Sulfolobus tengchongensis* spindle virus 1 (STSV1) was recently isolated from *Sulfolobus tengchongensis* (*128*). This virus is significantly larger than members of the *Fuselloviridae* and probably represents a new viral family. The ~75-kb genome shares little sequence similarity with any genes in the public database and is significantly larger (approximately five times larger) than genomes of the *Fuselloviridae* (*7*). Spindle-like morphologies have been reported outside of the *Sulfolobus* genus. *Acidianus* two-tailed virus (ATV) of the *Bicaudaviridae* family is a roughly spindle-shaped virus (~100 x 150 nm) that undergoes a novel cell-independent structural transition (*124*). On release from its host, the virus particle develops long tails at each end, but only at high temperatures (75–90°C). At lower temperatures, the virus retains its spindle-shaped, tailless morphology. The extracellular growth of tails might represent a survival strategy

for the virus in its harsh environment by promoting host recognition and attachment, especially where host availability might be limited.

Although the spindle-shaped morphology seems to be prevalent among thermophilic crenarchaeotes, there are examples of morphologically similar viruses that infect euryarchaeotes, including one thermophile, the *Pyrococcus abyssi* virus 1 (PAV1), which was isolated from a marine thermophilic euryarchaeote collected from a deep-sea hydrothermal vent (*118*). PAV1 particles are 80 x 20 nm, with tail fibers that extend from one end. In addition to the close morphological resemblance of the PAV1 particle to that of the *Fuselloviridae*, PAV1 packages a similarly organized genome of approximately the same size (~17.5 kb). Spindle-shaped virus morphologies are not limited to thermophilic archaea, as exemplified by the 44 x 74 nm spindle-shaped *His*1 virus isolated from the extreme halophile *Haloarcula hispanica* (*114*) and the spindle-shaped particle isolated from *Methanococcus voltae* A3 (*129*). It is worth noting that this unique virus morphology seems to be restricted to the Archaeal domain. One explanation for this observation is that the SSV morphology evolved after the Archaea separated from the other two domains of life (Figure 2.2).

In contrast to the domain-restricted distribution of spindle-shaped morphologies, recent structural studies focusing on the major capsid protein (MCP) of STIV (*126*) have identified a fold (double-barrel jelly roll) also used in

bacteriophage and eukaryotic viral architectures (*8, 126*). The MCP of STIV is structurally homologous to the MCPs of the algal virus *Paramecium bursaria Chlorella* virus 1 (PBCV1), mammalian adenovirus and the bacteriophage PRD1. Additional support for this evolutionary link has been provided by protein and lipid analysis of STIV particles (*9*). This analysis showed that STIV has an internal lipid membrane composed of a subset of the host's lipids, as seen in PRD1 (*9*). More recently, a halophilic virus, SH1 (*130*), has been isolated which seems to have a similar structural architecture, including the icosahedral shape, the internal lipid membrane and a protein with an ATPase motif that is similar to one occurring in PRD1 (*131*) and homolgous to one predicted to exist in STIV (*9*). Although these viruses have different genome structures and encode for structural genes with low primary sequence similarity, their evolutionary relationship is preserved in the 3D structures of their MCPs. The structural similarity between MCPs and the overall virion architecture indicates that these viral capsid proteins share an ancient ancestor. Once separated, the individual viruses evolved independently. This is the first example of a virus architecture that has been found in all three domains of life and suggests that this virus architecture might have existed prior to the separation of these three domains, over 3 billion years ago (Figure 2.2).

The shared ancestry among icosahedral viruses from the three domains is mirrored by the likelihood of shared ancestry between tailed phage and

herpesviruses. Recent comparative analyses of the capsid proteins of tailed phages infecting bacteria and herpesviruses infecting Eukarya have revealed a common fold in the capsid protein (*132*). This fold provides stability for the capsid and allows for the conformational flexibility seen in the assembly and maturation of these viruses. Further comparisons between the cryo-electron microscopy (cryo-EM) reconstructions of SPO1, a tailed phage, and HSV-1, a

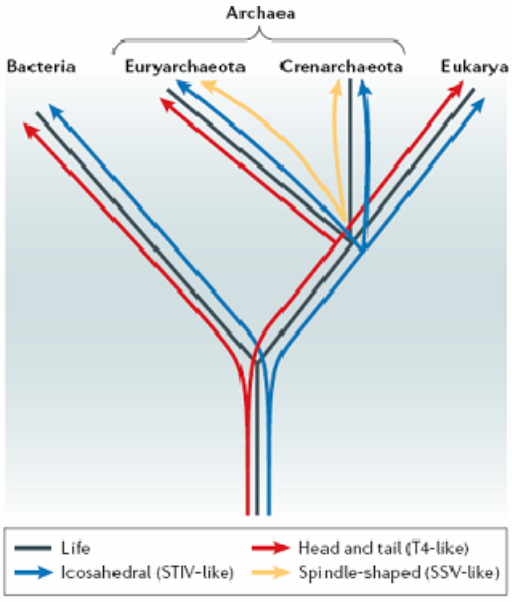


Fig 2.2 Seeing Viral Evolution Through the Branches of a 16S Based Phylogenetic Tree of Life. Icosahedral viruses, similar to Sulfolobus turreted icosahedral virus (STIV), have been detected in all domains of life, and probably evolved before the first branching event. The head-and-tail morphology is detected in the Euryarchaeota and the Bacteria, indicating that this lineage also has an ancient origin. Recentevidence indicates that this lineage is also related to the herpesviruses of the Eukarya, however no representative, has yet to be detected that infects the Crenarchaeota. The spindle-shaped morphology has only been detected within the Archaea and seems to have arisen more recently than the other morphologies.

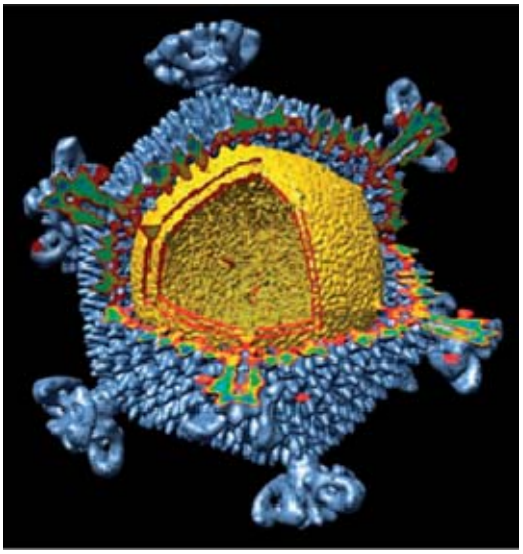


Fig. 2.3 Cryo-transmission-electron microscopy image reconstruction of the *Sulfolobus* turreted icosahedral virus. A cut-away shows the internal lipid layer (yellow) and the integration of the turret structures into the main capsid (*8-10*). Image courtesy of J. Hilmer, Montana State University, using Chimera.

herpesvirus, have strengthened the argument for shared ancestry (*133*). The tailed phage has been detected in bacteria and euryarchaeotes, but so far has not been shown to infect crenarchaeotes. If the tailed phage shares a common ancestry with the herpesviruses, it is possible that there is either a tailed phage or a herpes-like virus that infects members of the Crenarchaeota.

Not surprisingly, all virus particles isolated from high-temperature environments are thermostable, some withstanding near-boiling, acidic conditions. It has been suggested that the unusual morphologies detected in these environments are adaptations for survival of free viruses (*124*). However, it is unclear how these morphologies contribute to the stability of virus particles in hot, acidic environments. Recent atomic resolution structural studies of one

crenarchaeal virus, STIV, indicates that the thermal stability of the particle is a consequence of increased buried surface area between capsid subunits, a tightly packed subunit structure with reduced cavity volume, higher proline content, short solvent-exposed protein loops, an increase in electrostatic interactions and an increased polar surface area (*8*) (Figure 2.3). The occurrence of intracellular disulfide bonds might also aid in the stabilization of the virus-encoded proteins. It is speculated that intracellular disulfide bond formation might also be a general feature of proteins from hyperthermophilic archaea to enhance thermal stability (*134*).

Genomic Structure and Content

All archaeal viruses isolated so far package circular or linear double-stranded DNA genomes, two of which are highly modified (*117*). The genome of *Sulfolobus neozealandicus* droplet-shaped virus (SNDV) has been shown to have Dam-like methylation, possibly owing to a virus-encoded methylase (*120*), whereas STSV1 seems to have several different modifications, including specific modification of cytosine residues, although Dam-like methylation does not seem to occur (*128*). Some of the linear genomes have covalently closed ends (*119*), whereas others have unidentified modifications (*135*). These modifications, like those used by many DNA bacteriophage, are expected to aid in chromosome replication and in preventing degradation by host nucleases.

The comparison of four SSV-like isolates from different hot springs worldwide revealed that out of the approximately 31–35 ORFs encoded by each genome, a common set of 18 ORFs are shared by all four sequenced isolates (*18*). When the genomes of two different Icelandic isolates of the rudivirus SIRV were compared, a similar situation was observed (*136*). The SIRV genomes share clusters of ORFs with high similarity, interspersed with small regions with little or no sequence similarities. The unique genes observed among related viruses might reflect different evolutionary histories or unique adaptations required for a particular host within a particular hot spring environment. As has been suggested for other viruses, it seems that crenarchaeal viral genomes might be composed of a mosaic of genes, probably obtained from different sources at different times (*137*).

As more viral genome sequences become available, it is evident that there are subsets of gene families that are common among different types of crenarchaeal viruses. Based on sequence similarity, ORFs that code for putative glycosyltransferases have been annotated in the genomes of AFV1 (*138*), AFV2 (*139*), ARV1 (*140*), SIFV (*141*), SIRV1 (*142, 143*), SIRV2 (*142*), ATV (*124*) and STSV1 (*128*). More recently, the crystal structure of ORF A179 from STIV reveals a GT-A fold which is common among glycosyltransferases (*144*). Structure based sequence alignments reveal the invariant DXD motif common among these enzymes and adds further support to this proteins role as a

glycosyltransferases. Another sequence shared by diverse thermal virus families is ORF B251 from SSV1. This ORF has limited similarity to DnaA proteins (*145*) and related sequences are found in the genomes of all fuselloviruses, STSV1 and ATV.

This is by no means a complete list of the similarities among the different isolates of viruses that infect thermophilic crenarchaeotes. There are many ORFs that are shared between the *Lipothrixviridae* and the *Rudiviridae* which, along with morphological similarities, has led to the suggestion that the *Lipothrixviridae* and the *Rudiviridae* might form a new virus superfamily (*136*). All of the above examples compare viruses isolated from two closely related genera, *Sulfolobus* and *Acidianus*. The relationship between these viruses might be a result of the close evolutionary history and shared habitats of their hosts. This suggestion is supported by the fact that the two spherical viruses that infect the *Thermoproteales*, which occur in more neutral-pH environments, share several ORFs, but have no significant similarities to any of the ORFs from viruses infecting the *Sulfolobales* (*135*). However, horizontal exchange of viral genes across diverse phylogenic boundaries has been observed for bacteriophage (*137*) and a similar exchange might become evident among these viruses as more sequences become available. Further analysis of viral genes and their products will provide insights into the past histories and current lifestyles of crenarchaeal viruses.

One surprise revealed by these viral genomes, and the genomes of their thermophilic hosts, has been their low GC content, typically around 40%. It is unclear why thermophiles have not exploited the increased thermal stability provided by GC base pairing. However, several other strategies have been implicated in maintaining nucleic acid stability in thermophiles, including increased intracellular ionic concentrations, binding of polyvalent cations or polyamines and nucleoside modifications such as methylation (*146*). Not surprisingly, codon usage in these viral genomes tends to follow the codon bias found in the host organisms and similar to most viral genomes, the crenarchaeal viruses have genes that are tightly arranged on their genome. On average, the viral genes tend to encode proteins that are smaller than their host proteins. Sequence comparisons of crenarchaeal viral genomes using standard alignment algorithms such as BLAST have revealed that few of the putative viral genes share significant similarity with other sequences in the public databases. Where sequence similarity is detected, it tends to be with genes of unknown function in other archaeal viruses or archaeal genomes. Although some of these unique viral sequences may represent proteins with novel function, a majority are expected to be distantly related to proteins in other genomes. Higher order levels of protein structure are preserved over much greater evolutionary distance and thus three-dimensional structures can be used to identify distantly related proteins with similar functions. The utility of this approach has recently been

demonstrated for two SSV1 proteins. The D63 protein (63 residues coded in frame D) from SSV1 is comprised of two antiparallel α-helices that are connected by a short loop (pdb; 1SKV) (*19*). The monomer dimerizes in solution to form an antiparallel four helix bundle. This fold is extremely common and associated with proteins having diverse biological functions. However, new insights into this proteins function are revealed when conserved residues identified by multiple sequence alignments of homologous proteins from SSV2 (D57) and SSV RH (F61) are mapped onto the proteins surface. Conserved residues are preferentially distributed on the N-terminal helix. The N-terminal helices from each monomer are anti-aligned at the dimer interface, resulting in the symmetric display of conserved residues on this surface (*19*). This observation lends itself to the hypothesis that this protein may function as an adaptor molecule that associates with two copies of a monomeric entity or with a single macromolecule that has complementing 2-fold symmetry. One attractive hypothesis recently presented by the Lawrence group, is that D63 regulates copy number of the virus in a Rop-like (Repression of primer) manner. This hypothesis is supported by the genomic context of D63. The D63 gene is co-transcribed with 9 other genes, one of which is a virally encoded integrase (*147-149*). Genes on this transcript may have similar functions related to viral latency. A second structure, also from this transcript (transcript 5) is immediately suggestive of function (*82, 149*). The F93 protein from SSV1 is homodimeric winged helix-turn-helix protein

(wHTH) (pdb; 1TBX) (*20*). A search for nearest structural neighbors using DALI or VAST indicates that F93 is most similar to the transcriptional regulators, SlyA (pdb, 1LJ9) (*150*) and MarR (pdb, 1JGS) (*151*). By analogy to these and other homodimeric winged-helix DNA binding proteins, F93 is expected to bind a specific palindromic or pseudo-palindromic sequence in the viral or host chromosome. Identification of the binding target for this protein will help define the SSV life cycle and give further indications as to the role of genes on this transcript. These two structures and their potential biological implications are discussed in more detail in chapter 6.

Viral Replication Cycles

Although the basic replication cycle (adsorption, replication and release of progeny virus) for any crenarchaeal virus remains to be determined, several trends are emerging.

Entry and Uptake

Viruses of the *Fuselloviridae (18)*, *Lipothrixviridae* (*138, 141*), *Rudiviridae* (*140*) and *Guttaviridae* (*120*) are thought to associate with their host through tail fibers that are present at one or both ends of the virus particle. It is probable that these tail fibers are involved with attachment of the virus particle to components in the host S-layer or the underlying membrane. Similarly, the turret-like projections on the surface of the STIV virion are thought to be

involved with host cell recognition and attachment (*126*). The mechanisms of virus uptake are not known for any crenarchaeal virus. Presumably, a specific virus enters through a pathway in which either the virion is taken up into the cell or the viral genome is delivered into the host cell from the virion, which is attached on the exterior surface of the cell. Whereas some viral infections can be cured from cultured hosts (*141*), others seem to be stably associated with the host cells (*119*). Once inside the cell, some viruses can integrate their genome into the host chromosome, whereas others are maintained as extrachromosomal elements. In the case of SSVs, the full-length genome is integrated site-specifically into the host chromosome. Viral integration is mediated by a virus-encoded integrase of the tyrosine recombinase family (*147, 148, 152*). A sequence in the integrase gene (*att*P) has been shown to target viral integration by complimentary base-pairing between the virus and host chromosomes (*att*A). The four SSV viruses have been shown to integrate into different locations, commonly exploiting the tRNA genes of their host (*18*). The biological significance of SSV integration is not currently understood, however other viruses, such as lambda phage, typically do not require the integrated copy to complete the viral replication cycle. It is not known if integration of the SSV genome is required to complete its viral replication cycle. The SSV genome is also maintained as a negatively supercoiled episomal plasmid (*153*). Given what is known about the role of the integrated form of other viruses, it is probable

that the integrated form of the SSV genome is used as a mechanism by which the viral genome can be maintained during cycles of host replication. A tyrosine recombinase gene has also been identified in ATV. This virus seems to integrate into the host genome and induction results in a lytic cycle that culminates in lysis of the host (*124*). Induction of ATV resulted when cultures were treated with either mitomycin C or ultraviolet (UV) light — two well known inducing agents. Interestingly, virus production was not detected when cells were grown at the optimum growth temperature for the host (85°C) but production was triggered by lowering the temperature of the cultures to 75°C (*124*). Two other virus types have been shown to encode integrase-like genes. STSV1 has a truncated version of what appears to be an integrase gene. However, Southern analysis of infected cells indicates that STSV1 does not integrate (*128*). A putative tyrosine recombinase-like gene has also been identified in STIV ((*126*); D. Mensur personal communication). Southern blots have failed to (H. Þórisdóttir, unpublished observation) detect integration of the STIV genome in virus infected cultures of S. solfataricus, strain P2 (*154*). It is important to note that members of the tyrosine recombinase gene family have a broad range of biological functions in addition to integration of phage into host genomes (*155*), including resolving concatamers of the viral chromosome formed during replication, which might be important for correct packaging of virus particles. Tyrosine recombinases might also aid in the partitioning of plasmids or phage genomes

into daughter cells during cell division, a mechanism that might assist in the spread of viruses without requiring exposure to the extracellular environment.

Transcription and Genome Replication

Many crenarchaeal viruses produce transcripts that seem to be polycistronic, a common virus strategy for co-expressing functionally related genes. An exception to this is STSV1, which seems to transcribe monocistronic RNAs, as 76% of the ORFs are preceded by putative promoter regions (*128*). In two studies of crenarchaeal virus transcription, polycistronic transcripts were detected (*149, 156*). So far, a high degree of differential viral gene expression has not been detected, although these studies have not been carried out with synchronized cultures, which could identify temporal patterns of transcription. Virus production in SSV1 can be induced in cells by UV irradiation (*125, 157*), mitomycin C (*158*) or infection with SIRV1 (*159*). Eight transcripts (T1–T8) were identified in asynchronous cultures of SSV1, whereas a ninth transcript (T*ind*) was only expressed after UV induction, and corresponds to increased virus production (*149*). In cells infected with SIRV1 or SIRV2, transcripts covering all but one ORF were detected 30 minutes after infection. Transcription of some ORFs seem to vary during the infection cycle, including the ORF coding for the coat protein, which was initially transcribed as a polycistronic message, but later detected as a monocistronic message (*156*). Although multiple transcriptional promoters containing the TATA-like core are present in many of the crenarchaeal

virus genomes (*82, 126, 135, 138, 141*) it is not clear how many of these elements are functionally relevant. In general, crenarchaeal virus transcription is expected to mimic that of its host. Currently, it is thought that archaeal gene transcription uses a eukaryotic pol II-like basal transcription apparatus that is regulated by bacterial-like systems in which sequence-specific DNA-binding proteins compete for, or prohibit, extension of 3' ends by the transcriptional machinery.

Although there are few studies of replication strategies for crenarchaeal viruses, replication schemes from viruses of bacteria and eukaryotes might provide some insights. In the case of the *Rudiviridae*, it is postulated that DNA replication is similar to that shown for the eukaryal poxviruses (*136*). In support of this model, the SIRV genome is structurally similar to the vaccinia virus genome, with both having linear, double-stranded DNA that is covalently closed by terminal hairpin structures. The hairpin structures in both SIRV and vaccinia virus genomes are preceded by inverted terminal repeats (called 'flip' and 'flop' in vaccinia virus). Replication is initiated by the introduction of a terminal nick, which exposes a free 3'-hydroxl and allows the free end to fold back on itself forming a duplex with its complementary sequence in the inverted repeat region (*160*). An evolutionary link between these two apparently distant viruses is further supported by the relatively high sequence similarity of homologous ORFs, such as the putative Holliday junction resolvases detected in all three rudiviruses

and required for this method of replication (*136, 140*). A more comprehensive review concerning SIRV replication has recently been contributed by Prangishvili *et al* (*117*).

A potential origin of replication has been identified in STSV1 (*128*). Analysis of the genome has identified a 1.4-kb intergenic region with a low GC content between ORF1 and ORF74. Within this region are two tandem repeats, TR1 (25 bp) with 5.5 copies and TR2 (40 bp) with 2.5 copies, and two sets of inverted repeats which could form stem-loop structures. These characteristics are consistent with this region being the origin of replication. More recently the origin of replication for the SSV1 virus has been experimentally mapped using a 2D-gel technique, similar to the approach used to map the origin of replication initiation (*ori*) in the host cell, *S. solfataricus* (S. Bell, personal communication). A defined viral *ori* is expected to offer considerable insight into the mechanism and evolutionary history of replication in this virus family.

Assembly and Release

The assembly and release of most crenarchaeal viruses known so far does not require cell lysis; the one known exception is the release of ATV, which is mediated by lysis (*124*). Thin sections prepared for transmission electron microscopy (TEM), over a time course of *Sulfolobus* spp*.* cells infected with SSV indicate that virus particles assemble at the host membrane where particles incorporate host lipids, and viral progeny are released without causing cell lysis

(M. Young, unpublished data). By contrast, many of the viruses infecting bacteria and euryarchaeotes tend to lyse cells on release (*161*). Most crenarchaeal viruses seem to maintain a chronic infection, in which particles are produced either continually or during short events resulting in growth inhibition. Long infections with extended release of virus progeny have been proposed to be an adaptation to the extreme external environment (*123*). The stability of a virus in hot, acidic environments might be reduced, therefore continuous release of virus might minimize the time a single virus is exposed to the extracellular environment.

Biogeographic Isolation of Crenarchaeal Viruses

Related crenarchaeal viruses can be isolated from hot springs around the world. The occurrence of morphologically similar viruses from geographically isolated hot springs has raised questions regarding their genetic relationship. It has been suggested that genomic variability among these viruses might be related to their long term geographical isolation from one another and/or differences in biogeochemical composition in each geographical location. Interestingly, studies examining diversity within the host genera, *Sulfolobus*, indicate biogeographical isolation (*162*). If thermophilic crenarchaeotes and their viruses are geographically isolated, it would counter the argument that 'everything (all microbial species) is everywhere' and the environment

determines which species are successful and dominant. Ongoing studies, conducted in this lab, indicate high rates of viral migration between different hot springs within YNP. Understanding the mechanism(s), frequency and rates of viral migrations between distant habitats will be essential to understand the evolution and ecological significance of these and other widely distributed viruses. Determining the total diversity of crenarchaeal viruses is the first step to understanding their distribution around the world. From the study of three hot springs in YNP it is clear that the community size for a single virus type is extremely large, as sequence accumulation curves show that the total diversity was not sampled in the approximately 2,000 clones sequenced for SIRV-like viruses (*163*). These results indicate that the total virus metacommunity size probably exceeds 10^9 sequence types within the SIRV-like viruses in a hot spring environment, underscoring the substantial global biodiversity represented by viruses. The possibility that migration might have an important role in the distribution of viruses within a geothermal area leads to the question of how viruses migrate. Possibilities include movement of either virus particles or infected hosts through underground water connections, by movement of animals between sites or through the atmosphere by steam. Real time PCR performed using universal SSV primers, on water vapor samples collected from steam at the crater hills location, indicates that the steam carries a substantial viral load (M.

Young, unpublished data). Although this data strongly implicates steam in viral dispersion, it is probable that several avenues are involved.

Future Directions and Questions

Although more crenarchaeal viruses have been isolated and examined in recent years, we still have only a rudimentary understanding of this group of unusual viruses. Current and future efforts will be directed at gaining a better understanding of fundamental virology; including mechanisms of attachment, transcription, genome replication and viral release. These studies are expected to facilitate developing genetic systems, providing archaeal virologists with the fundamental genetic tools that have been historically offered by bacteriophage.

The total diversity of the crenarchaeal viruses has yet to be determined and remains an important avenue of research. This question is being studied by further isolation and characterization of viruses and by culture-free techniques. The application of culture-free techniques might enable the study of non-thermophilic crenarchaeal viruses, where no host systems exist. An environmental approach, specifically aimed at detecting thermophilic archaeal RNA viruses replicating in organisms occupying deep branches in the phylogenetic tree, could lead to new insights into the role of RNA in the evolution of life. This would support theories suggesting that primordial life was RNA based. The challenge is to understand the ecological role these viruses have in

their environments and their role in the evolution of life. Although the study of crenarchaeal viruses is still in its infancy, it is a field that is rapidly advancing, and the future holds potential for significant discovery.

Box 2.1 A new approach to thermal virology

Recently, culture-independent methods for detecting virus-like particles (VLPs) directly from hot springs have been developed. Primer sets based on conserved genes identified in comparative genomic studies (*18, 136*) have provided the first opportunity to assess geographical distributions of viral populations, temporal distribution, abundance, and the general ecological significance of specific thermophilic virus populations (*163*). As an alternative to the molecular phylogenetic approach, direct detection of VLPs from hot springs has been accomplished. Epifluorescence microscopy estimates of free virus particles in acidic hot springs range from 10^3–10^5 VLPs ml^{-1} (A. Ortmann, unpublished observations), which are orders of magnitude lower than those observed in marine environments (*164*). This reduced number of VLPs might reflect the adaptation of viral lifestyles that minimize the exposure of virus particles to the harsh environment. Direct filtration of hot spring water that selects for VLP-sized components; have been successfully used for the detection of viruses from hot, acidic environments (J. Fulton, unpublished observations). Although new virus particle morphologies have been observed, preliminary results reveal a community dominated by morphologies previously observed in enrichment cultures. However, sequencing of viral communities directly isolated from the environment indicates an abundance of novel viral genes that share little similarity with viral genes from cultured viral isolates. Although culture-independent methods also have biases, they are likely to expand our understanding of viral diversity in hot environments.

CHAPTER 3

COMPARATIVE GENOMIC ANALYSIS OF THE THERMOACIDOPHILIC ARCHAEAL *FUSELLOVIRIDAE* VIRUSES

Abstract

The complete genome sequences of two *Sulfolobus* spindle-shaped viruses (SSVs) from acidic hot springs in Kamchatka (Russia) and Yellowstone National Park (USA) have been determined. These non-lytic temperate viruses were isolated from hyperthermophilic *Sulfolobus* hosts and both viruses share the spindle-shaped morphology characteristic of the *Fuselloviridae* family. These two genomes, in combination with the previously determined SSV1 genome from Japan and the SSV2 genome from Iceland have allowed us to carry out a phylogenomic comparison of these geographically distributed hyperthermal viruses. Each virus contains a circular dsDNA genome of ~15 kbp encoding approximately thirty-four open reading frames (ORFs). These *Fusellovirus* ORFs show little or no similarity to genes in the public databases. In contrast, eighteen ORFs are common among all four isolates and may represent the minimal gene set defining this viral group. In general, ORFs on one half of the genome are co-linear and highly conserved, while ORFs on the other half are not. One shared ORF among all four genomes is an integrase of the tyrosine recombinase family. All four viral genomes integrate into host tRNA genes. The specific tRNA gene used for integration varies and one genome integrates into

multiple loci. Several unique ORFs are found in the genomes of each isolate and may be a consequence of their evolutionary history, geographic isolation, a requirement for replication in their specific hosts or adaptations to unique features of their respective thermal environments.

Introduction

Comparative genomics is a useful tool for understanding new viral families. One such family, the *Fuselloviridae*, has recently been created to accommodate the approximately 60 X 90 nm spindle shaped viruses found exclusively in the archaeal domain (International Committee on Taxonomy of Viruses, http://www.ncbi.nlm.nih. gov/ICTV/). This viral family currently consists of a single virus, SSV1 (*Sulfolobus* spindle-shaped virus 1), with three others considered tentative members in this genus: SSV2 (*Sulfolobus* spindle-shaped virus 2), SSV3 (*Sulfolobus* spindle-shaped virus 3) and the satellite virus pSSVx (plasmid *Sulfolobus* spindle-shaped viruses x) (*165*). These viruses have circular dsDNA genomes, share a common morphology and are temperate in *Sulfolobus* species that commonly inhabit high temperature (>70ºC) acidic (pH <4.0) environments (*159*). His1 (*Haloarcula hispanica* 1) (*114*), SNDV (*Sulfolobus neozealandicus* droplet-shaped virus) (*120*) and a virus-like-particle isolated from *Methanococcus voltae*, A3 (*129*), are morphologically similar to the

Fuselloviridae. However, their genome topology, genomic structures and host ranges vary dramatically, making their formal classification unclear.

SSV1 is the type virus of the *Fuselloviridae* family and the first high temperature virus to be characterized in detail. SSV1 was originally isolated from *Sulfolobus shibatae* cultured from a sulphurous hot spring in Beppu, Japan (*82, 125, 153, 166*). The virus can also infect virus-free strains of *Sulfolobus solfataricus* originally isolated from a solfataric field near Naples, Italy (*157, 167*). In both hosts, virus production is UV inducible (*125, 157*) and the genome is stably maintained in three different forms. The packaged viral genome is positively supercoiled, while the episomal form of the viral genome has been isolated from *Sulfolobus* as positively supercoiled, negatively supercoiled or relaxed dsDNA (*168*). A provirus is also found integrated into a host tRNA gene (*167, 169, 170*). A 7.4 kbp segment inserted into a *S. solfataricus* arginyl tRNA gene shares extensive sequence similarity with a portion of the SSV1 genome and is likely a remnant of viral integration (*170*). The low G/C content (39.7%) of the viral genome is similar to that of its host. Sequence analysis of the encapsidated 15,465 bp genome (NC_001338)

revealed 34 open reading frames (ORFs) (Figure 3.1 and Table 3.1) (*82*). The predicted ORFs encode protein products that range from ~6-86 kDa and are tightly arranged on the viral genome. Nine transcripts cover all 34 SSV1 ORFs (*149*). This suggests that the viral genes are transcribed via a polycistronic

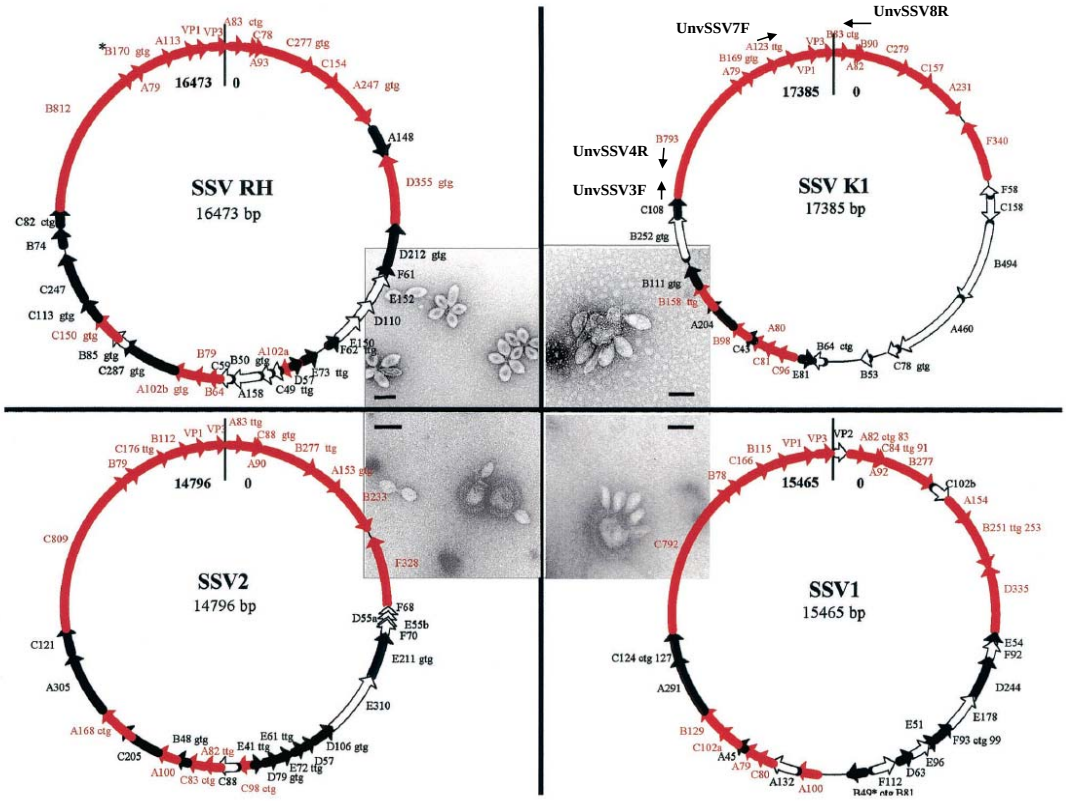


Fig. 3.1 Genome maps of the four SSV isolates. Conserved ORFs shared by all genomes are shown as red arrows. ORFs shared between two or three of the SSV genomes are shown as solid black arrows, and ORFs unique to each isolate are shown as open arrows. TEM images of each virus are positioned next to maps of their respective genome (bars, 100 nm) (3, 34). The alternative initiation codons (asterisks) are indicated directly following the name of each ORF in which they were identified.

strategy. Only four of the 34 ORFs have been assigned a function. The ORFs encoding for three structural proteins VP1, VP2 and VP3 were assigned by sequencing proteins from purified virus particles (*171*), while the fourth ORF was recognized to encode a type I tyrosine recombinase family integrase by sequence similarity to other known recombinases and by direct biochemical analysis (*82, 147, 148, 172*). The remaining 30 ORFs show little or no significant sequence similarity to genes in the public databases. SSV1-based shuttle vectors that can replicate in both *S. solfataricus* and *E. coli* have been described (*85, 86, 173*). These vectors have the potential to greatly expand our ability to express genes in *S. solfataricus.*

The complete genome sequence of SSV2 has recently been determined (NC_005265) (Figure 3.1) (*127*). This virus was isolated from a solfataric hot spring in Reykjanes, Iceland and has been shown to infect *S. solfataricus* (*174*). The SSV2 virus is morphologically indistinguishable from SSV1 and the two viral genomes are similarly arranged. Overall, the SSV2 and SSV1 genomes share 26 ORFs, while 9 ORFs are unique to SSV2 and 8 are unique to SSV1. SSV2, unlike SSV1, does not encode a VP2 structural protein. The genomic comparison of SSV2 with SSV1 indicates that these are two distinct *Fuselloviridae* viruses (*127*). The original culture containing SSV2 also produced a subset of smaller satellite particles (60 X 40 nm) termed pSSVx (AJ243537). The genetic material

packaged by the satellite particle is a hybrid of pRN family plasmids and two ORFs that share sequence similarity to ORFs in SSV1 and SSV2 (*174-176*).

Full-length copies of both SSV1 and SSV2 genomes are found integrated into tRNA genes of their host (*153, 169*). The virally encoded integrase identified in the SSV1 (D335) genome has been shown to function as a site-specific endonuclease and ligase (*147, 148, 172*). *In vitro* the SSV1 integrase is also capable of excisive recombination, but the mechanism of excision *in vivo* is not presently understood (*82, 147*). The viral integrase, like all 130 members of the tyrosine recombinase family, contains the conserved RHRY tetrad (*177, 178*). The SSV1 genome is found integrated into an arginine tRNA gene (*169*), while SSV2 is thought to integrate into a glycyl tRNA gene (*127*). The process of viral integration is modeled to involve base pairing of the *att*A sequence found within the target tRNA gene with the *att*P site found within the viral integrase gene. In the case of SSV1, the *att*A site is a 44bp region that is centered around the anticodon loop of an arginyl tRNA gene (*169*). The *att*P site is an identical sequence located within the 5' half of the viral integrase gene. The SSV1 viral genome integrates in a way that complements the tRNA gene and is presumed to maintain tRNA function (*169*). The recently described SSV2 genome also contains a tyrosine recombinase like integrase gene (ORF F328) (*127*). Forty-nine nucleotides in this gene are identical to sequences found in a glycyl tRNA gene in the *S. solfataricus* P2 genome (*179*). These regions likely represent the

*att*A and *att*P sites for SSV2. The fuselloviruses offer the only viral model for examining site-specific recombination in Archaea. The biological role of SSV integration and *in vivo* mechanisms for SSV excision remain unclear.

SSVs and their *Sulfolobus* hosts are emerging as a model system for examining Archaea and life at high temperatures. We are interested in the evolution of SSVs, the function of SSV encoded gene products and viral adaptations required for replication in high temperature environments. Here we present the genomes of two additional SSV-like viruses, one from Kamchatka, Russia (SSV K1) and the other from Yellowstone National Park, USA (SSV RH). These genomes, in combination with the previously determined SSV1 and SSV2 genomes, provide us with four geographically distinct isolates that we have used in a comparative genomic analysis. In addition, we have determined the integration sites of the two new genomes into *S. solfataricus*. This work is the first such comparative analysis within the archaeal domain.

Material and Methods

Environmental Sampling

Liquid samples were collected from an acidic hot spring (pH 4.0, 75˚C) at 54˚ 26.357' north, 160˚ 8.573' east in the Valley of the Geysers, Kamchatka, Russia. The pH of the Kamchatka samples was adjusted to 5.0 and stored in anaerobic vials for transport as described previously (*159*). Liquid samples collected from

acid chloride hot springs (pH 3.2, 81ºC) in the Norris Geyser Basin of Yellowstone National Park (YNP), USA (44° 43.653' north, 110° 42.862' west) were transported aerobically and placed into enrichment culture within six hours as previously described (*10*).

Enrichment Cultures of Environmental Samples

Enrichment cultures were established by inoculating 20 mls of minimal medium supplemented with 0.1% tryptone and adjusted to pH 3.2 (*159*) with 1 ml of environmental sample. Liquid cultures were grown aerobically in long neck Erlenmeyer flasks placed in shaking oil bath incubators at 80°C for five to seven days. Turbid cultures were streaked on 0.6% Gelrite™ gellan gum plates supplemented with 0.2% tryptone (*159*). Single colony clones were isolated and screened for plaque formation as described previously (*159, 173*). Single colony isolates were used to establish 25 ml cultures. Seven days after inoculation, culture supernatants were visually screened by transmission electron microscopy (TEM; Leo 912 AB) for the presence of virus-like particles as previously described (*10*). Cultures found to be producing large quantities of virus-like particles were scaled up to 250 ml cultures. Total DNA was extracted from 1.5 ml aliquots (*173*). This DNA was used as a template for polymerase chain reaction (PCR) based amplification of the 16S rDNA gene, which was subsequently cloned and sequenced as previously described (*10*).

Virus Purification and Viral Nucleic Acid Isolation

Cells were isolated from cultures by low speed centrifugation and their extra-chromosomal DNA was isolated by previously described methods (*153, 173*). Virus was precipitated from culture supernatants by addition of polyethylene glycol 8000 [10% (wt/volume) final concentration] and stirred at 4°C overnight. The precipitated virus was collected by low speed centrifugation and resuspended in a minimal volume of NNM buffer (20mM $NaPO_4$, pH 5.5, 100mM NaCl, 1 mM $MgCl_2$). After low speed centrifugation to remove material that did not resuspend, the supernatant was mixed with CsCl to a final concentration of 39% (wt/volume) and spun at 69,000Xg for 24 hours in a SW 41 rotor (Beckman, Fullerton, CA). The dominant band was removed and dialyzed against NNM buffer overnight at 4°C. Virus was analyzed by TEM and UV/Vis. spectroscopy. Nucleic acid was isolated from purified virus by the SDS/proteinase K method (*180*).

Construction of Viral DNA Library and Sequencing

DNA was mechanically sheared by nebulization as outlined by the manufacturer (Shotgun Cloning kit, Invitrogen, San Diego, CA). Sheared DNA was ligated into pCR 4Blunt–TOPO (Invitrogen) and transformed into *Escherichia coli* XL-2 MRF′ (Stratagene, San Diego, CA). Plasmid DNA from colonies containing inserts greater than 500 base pairs were prepared for sequencing according to the manufacturer's instructions (Eppendorf, Perfect Preps,

Westbury, NY). Purified DNA was sequenced using universal M13 primers according to Big Dye III termination sequencing protocols on an ABI 3700 automated DNA sequencer (Applied Biosystems, Foster City, CA). After sequence assemblies (see below), any remaining gaps were PCR amplified using primers that flanked each gap. These PCR products were subsequently cloned and sequenced to provide at least three-fold sequence coverage of the genome, except for one small region of about 100 nt that is only two-fold. In this instance both strands are sequenced and of a high quality.

Sequence Analysis

Vector stripping and contig assemblies were accomplished with Sequencher Version 4.1 (Gene Codes Corp., Ann Arbor, MI). ORFs were initially identified using ORF Finder (http://www.ncbi.nlm.nih.gov/gorf/orfig.cgi). ORFs were confirmed and others were identified by manually scanning for TTG, GTG or ATG start codons between stop codons in all six frames. All possible ORFs were subjected to BLAST analysis against the non-redundant (Genbank) database (http://www.ncbi.nlm. nih.gov/BLAST/) (*181*). ORFs predicted to encode products of less than 50 amino acids and having no significant match to the database were discarded. Small ORFs (<50 amino acids) that have sequence similarity to ORFs in other SSV genomes were subjected to further analysis. Additional ORFs (<100 a.a.) were eliminated if they had no homologue in the database or in the other SSV genomes and if they overlapped a larger ORF by

more than 50%. Large ORFs (>100 a.a.) having no SSV homologue were allowed to overlap by up to 30 amino acids before the smaller ORF was eliminated. The SSV RH ORFs identified by this method were compared to those ORFs predicted by Glimmer 2.0 (*182*) and RBS finder (*183*). Numbering of the SSV RH and SSV K1 genomes was standardized using the first nucleotide following the stop codon of VP3 as nucleotide one and the last nucleotide in the stop codon of VP3 as the final nucleotide of the genome. All identified ORFs are listed in the column below their respective genome and homologous ORFs are in rows (Table 3.1). Open reading frames are also numbered from this first nucleotide and are named according to their frame and number of amino acids. All ORFs in the frame starting with the +1 nucleotide are labeled "A", +2 labeled "B", +3 labeled "C" and those on the opposite strand D, E and F respectively. ORFs of the same length and in the same frame are distinguished by a lower case letter (i.e. F61a). Tandem repeats were identified using Tandem Repeats Finder (http://c3.biomath.mssm.edu/trf.html) (*184*).

SSV Genomic Comparisons

TFASTX was used to compare each predicted gene product to the six frame translations of all four SSV genomes (Biology Workbench, http://workbench.sdsc.edu/). Predicted gene products sharing sequence similarity were considered to be possible protein homologues and were

TABLE 3.1 Identified ORFs in SSV isolates.

ORFs[a] in:			
SSV RH (*n* = 37)	**SSV K1 (*n* = 31)**	**SSV2 (*n* = 35)**	**SSV1 (*n* = 34)[b]**
			VP2 (B74)
A83 (TTG)	**B83 (CTG)**	**A83 (TTG)**	**A82 (CTG), 83**
C78	**A82**	**C88**	**C84 (TTG), 91**
A93	**B90**	**A90**	**A92**
C277 (GTG)	**C279**	**B277 (TTG)**	**B277**
			C102b
C154	**C157**	**A153 (GTG)**	**A154**
A247 (GTG)	**A231**	**B233**	**B251 (TTG), 253**
A148			
D355 (GTG)	**F340**	**F328**	**D335**
	F58		
	C158		
	B494		
	B460		
	C78 (GTG)		
	B53		
	B64 (CTG)		
			E54
			F92
		F68	
		D55a	
		E55b	
		F70	
D212 (GTG)		E211 (GTG)	D244
E152			
D110			
E150			
			E178
		E310	
	E81	D106 (GTG)	F93 (CTG), 99

TABLE 3.1 (continued). Identified ORFs in SSV isolates.

			E96
F61		D57	D63
			F112
F62 (TTG)		E61 (TTG)	
		E72 (TTG)	B49 (CTG), 81
E73 (TTG)		D79 (GTG)	E51
A102a	**C96**	**C98 (CTG)**	**A100**
D57		E41 (TTG)	
C49 (TTG)			
B50 (GTG)			
A158			
C59			
		C88	
			A132
B64	**C81**	**A82 (TTG)**	**C80**
B79	**A80**	**C83**	**A79**
	C43	B48 (GTG)	A45
A102b (GTG)	**B98**	**A100**	**C102a**
C287 (GTG)	A204	C205	
B85 (GTG)			
C150 (GTG)	**B158 (TTG)**	**A168 (CTG)**	**B129**
C113 (GTG)	B111 (GTG)		
	B252 (GTG)		
C247		A305	A291
B74			
C82 (CTG)			
	C108	C121	C124 (CTG), 127
B812	**B793**	**C809**	**C792**
A79	**A79**	**B79**	**B78**
B170 (GTG)	**B169 (GTG)**	**C176 (TTG)**	**C166**
A113	**A123 (TTG)**	**B112**	**B115**
VP1 (A89)	**VP1 (B137)**	**VP1 (C88)**	**VP1 (C144)**
VP3 (C96)	**VP3 (A93)**	**VP3 (A92)**	**VP3 (A92)**

[a] Homologous ORFs are in the same row, and ORFs that are common to all four SSV genomes are in boldface. Alternative initiation codons identified are in parentheses.

[b] Alternative initiation codons were not considered in the original SSV1 annotation. The originally annotated names are included here so as not to confuse them with the new names (numbers) that we identified by using the alternative initiation codon indicated.

subjected to further alignments using CLUSTALW (*185*). If amino acid identities between sequences were less then 25%, then the length of the two proteins, their coding direction and their location on the genome were considered. Co-linear proteins (found in at least two of the four genomes) of similar length were considered functionally equivalent genes.

Phylogenetic Comparisons of Common ORFs

Amino acid sequences of each of the eighteen common SSV ORFs were aligned using CLUSTALX (*186*). Maximum parsimony and neighbor-joining analyses were conducted using test version 4.0b10 of PAUP* (*187*). Bootstrap analysis with resampling was performed on 10,000 replicates in each analysis. Maximum likelihood analysis was performed on a concatemer consisting of all the nucleotides in the set of 18 conserved ORFs from each genome. MrBayes analysis was run by sampling every 10,000 generations until the chain reached apparent stationarity (*188, 189*).

Analysis of SSV Integration Sites in the *S. solfataricus* Chromosome

Potential viral genome integration sites were predicted in a two-step process. First, the integrase ORFs containing the presumed *att*P sequences were individually aligned to a database containing all of the annotated tRNA genes in the *S. solfataricus* P2 genome (*78*) using FASTA alignments (http://workbench.sdsc.edu/). From these alignments, putative *att*P sites were identified and aligned to the entire *S. solfataricus* P2 genome using BLASTN

(http://www.ncbi.nlm.nih.gov/cgi-bin/Entrez/geblast?gi=180) (*181*). Sets of PCR primers were designed to flank all regions where the putative viral *att*P site aligned to the *S. solfataricus* P2 sequence with *e* value less than $2.0e^{-09}$ (Table 3.2). A PCR-based integration assay was designed such that one PCR primer corresponds to the viral sequence near the putative site of integration and a second primer flanks the tRNA gene of interest. Sites that were not annotated as tRNA genes were tested using the same strategy. Fifty milliliter cultures of *S. solfataricus* P2, were grown to mid-log phase, and then infected with filtered (0.2 µm) supernatants of either SSV RH or SSV K1 containing cultures. Infected cultures were then grown to stationary phase, at which point the cells were harvested by low speed centrifugation (6000Xg for 10 min) and total DNA was prepared as previously described (*173*). This DNA was used as the template in our PCR-based integration assays. PCR reactions were cycled 35 times and the annealing temperatures were generally five degrees below the lowest predicted melting temperature in the primer set. The resulting PCR products were sequenced as described above.

Structural Modeling of tRNAs

tRNAscan was used to model the structure of *S. solfataricus* tRNAs before and after integration (http://www.genetics.wustl.edu/eddy /tRNAscan- SE/).

Results and Discussion

We have isolated two new SSV-like viruses and determined their genome sequences. SSV RH (Sulfolobus spindle-shaped virus Ragged Hills) was isolated from an acid chloride (pH 3.2) hot spring (81ºC) in the Ragged Hills region of the Norris Geyser basin in Yellowstone National Park (USA). SSV K1 (Sulfolobus spindle-shaped virus Kamchatka 1) was isolated from a hot (75ºC) acidic (pH 4.0) pool in the Geyser Valley region of the Uzhno-Kamchatsky National Park on the Kamchatka peninsula (Russia). Small subunit rDNA sequence analysis indicated that the hosts of both viral isolates are closely related to *S. solfataricus*. In addition to their natural hosts, these viruses can also infect a virus-free strain of *S. solfataricus* (strain P2), from Pisciarelli Italy, like the previously characterized SSV1 and SSV2 viruses (*127, 157, 174*).

All of these viruses share a unique spindle shaped morphology that has only been seen in the archaeal domain. These virus particles are all about 60 X 90 nm with sticky tail fibers extending from one end. These tail fibers are presumed to be involved in viral attachment to the host and are likely the cause of virus clustering into rosette formations as seen in culture (Figure 3.1). Although the spindle shaped morphology is predominant, the virus structure appears to be malleable and is also able to form elongated or cigar shaped morphologies (*125, 157, 159, 174*).

SSV RH and SSV K1 virions contain double stranded circular DNA genomes of approximately 15 kbp. The entire genome sequence of both viruses was determined by random shotgun sequencing. The genome of SSV RH is 16,473 bp (NC_005360) and SSV K1 is 17,384 bp (NC_005361), both of which are larger than the SSV1 (15,465bp) and SSV2 (14,796bp) genomes (*82, 127*). All four viral genomes have a G+C content of ~38% like their *S. solfataricus* host (Figure 3.1, Table 3.1).

Genome analysis of these two new SSV-like viruses, in combination with previously sequenced isolates, indicates a clear relationship among all four viral isolates (Figure 3.1; Tables 3.1). Thirty-eight ORFs have been identified in SSV RH, while the larger SSV K1 genome contains only 31 ORFs (Figure 3.1, Table 3.1). In comparison, 34 and 35 ORFs were identified in the SSV1 and SSV2 genomes, respectively. This is one more ORF than was previously reported for SSV2 (*127*). The additional ORF in SSV2, ORF E41, was identified by having significant sequence identity (63% at the amino acid level) to ORF D57 in SSV RH. Eighteen of the SSV RH ORFs are shared by all four isolates (common ORFs are colored red in Figure 3.1 and are bold in Table 3.1), while seven other SSV RH ORFs are shared between only two of the other viral isolates. SSV RH contains eleven unique ORFs. Nine of these ORFs are found clustered in groups of 2 to 4, while the two remaining ORFs are found as isolated, unique ORFs. SSV K1 contains eight unique ORFs. Seven of these are found in a single cluster

following the integrase gene. Three of the SSV K1 ORFs are shared with SSV1 and SSV2, while two SSV K1 ORFs are shared exclusively with SSV RH (Figure 3.1 and Table 3.1). A comparison of all four viral genomes reveals that each genome shares a co-linear organization (Figure 3.1). Typical of many viral genomes, the predicted ORFs are tightly arranged on the genome with little non-coding sequence. There is little overlap between adjacent ORF's. SSV1 ORFs C84 and A92 overlap by 235 nucleotides and are the exception to this rule. Homologues of these two overlapping ORFs are found in all four viral genomes, indicating a functional role for this overlap or for the DNA sequence conserved in both ORFs (Figure 3.1). ORFs F70, D55a, D55b and F68 of SSV2 overlap by approximately 100 bp each, but these ORFs are not conserved in the other three SSV genomes. Like SSV1 and SSV2, the SSV RH and SSV K1 genomes encode potential products ranging from ~5 to 90 kDa with an average of ~16 kDa. Sequence alignments of the SSV ORFs show little or no sequence similarity to sequences in the public databases. However, among the four genomes, many ORFs share sequence similarity and are co-linearly organized. In general, ORFs on one half of the genome are more highly conserved among all viral four isolates and are arranged in the same orientation, while genes on the other half of the genome are poorly conserved. The highest identity of a predicted ORF product shared by all isolates is 84% (SSV RH B170 homologues) and the lowest is 13% (SSV RH ORF A102a homologues). ORFs with low sequence identity

were considered putative homologues only if they are of a similar size, similar location on the genome and oriented in the same direction.

The asymmetric clustering of cysteine codons that has been observed in the SSV1 and SSV2 genomes is also obvious in the genomes of the two new SSV-like viruses. The consistent patterning of cysteine codons across all four SSV genomes suggests that they share a common ancestor. It has been suggested that this ancestor was a fusion between two genomes with different histories, one that contained cysteine and another that did not (*82, 127*).

Three of the genes that lie on the conserved half of the genome have been assigned function. Two are viral coat proteins (VP1 and VP3) and the third is a viral integrase (*82, 171*). VP1 and VP3 were identified by N-terminal sequencing of proteins isolated from purified SSV1 particles (*171*) and the integrase was identified by sequence similarity (*82, 148, 177*). The N-terminal sequence of purified VP1 was found to be identical to the C-terminal seventy-three amino acids predicted in SSV1 ORF C144. The absence of the N-teminus indicates that VP1 is proteolytically processed (*171*). VP1 homologues have been identified in all four SSV genomes. The residues surrounding the putative proteolytic processing site are conserved in all four viral genomes, with the C-terminal ends of all four genes being highly conserved and the N-terminal ends being very divergent. The protease responsible for this processing has yet to be identified,

but these alignments suggest a common processing mechanism of the VP1 proteins in all four SSV isolates.

VP2 is one of only three proteins identified in the mature SSV1 virion (*171*). VP2 like the other two structural proteins (VP1 and VP3) was identified by N-terminal sequencing of proteins isolated from purified virus particles. This protein is composed largely of basic residues and was shown to be a DNA binding protein. The DNA binding activity and the presence of VP2 in assembled virus particles suggested that VP2 was an essential protein involved in packaging of the SSV1 viral genome. Surprisingly, there is no VP2 homologue in the other three SSV genomes or in the *S. solfataricus* (P2) genome (*78*). It remains unclear how these viruses package their genomes in the absence of a VP2 homologue. Interestingly, two ORFs (SSV1 A154 and B251 and their homologues in the other genomes) are also conserved in the satellite virus pSSVx (*174*). These ORFs probably encode gene products or contain sequences involved in packaging of the viral genomes, however direct evidence for these possible functions is currently lacking. The predicted products encoded by the remaining 29-36 ORFs (depending on the isolate) share no significant similarity to proteins with known function.

A striking feature revealed by this comparison is a set of 18 ORFs common to all four SSV isolates. We propose that the products encoded by this set of 18 common ORFs may represent viral functions common to all *Fuselloviruses* and

clearly reflect a common evolutionary history, despite their geographic isolation. This set of common genes may also represent the minimal replicon of the *Fuselloviridae*. Maximum-parsimony and neighbor-joining trees generated for each of the 18 conserved ORFs do not result in a consistent branching pattern. However, a maximum-likelihood tree generated using a concatemer of nucleotide sequences from all 18 ORFs in each SSV genome suggests that SSV1 and SSV K1 are more genetically similar, while SSV RH and SSV1 are the most genetically different.

This common set of 18 ORFs shared by all four SSV isolates is not contiguous. As mentioned earlier, half the SSV genome is highly conserved while the other half is more divergent. ORFs not common to all four SSV genomes rarely interrupt ORFs in the conversed half of the genome (Figure 3.1). The other half of the genome is largely composed of ORFs that are common to only two or three of the SSV genomes. ORFs in any one SSV genome that have no significant sequence similarity to ORFs in the other SSV isolates are indicated as open arrows in Figure 3.1. These unique ORFs are likely the consequence of their individual evolutionary history, geographic isolation, requirements for replication in their specific hosts, or adaptations to unique features of their respective thermal environments.

Two separate imperfect direct repeats have been identified in the SSV RH genome, while no repetitive sequences were identified in the SSV K1 genome

using the same method. The first repeat (6261-6337 in the SSV RH genome) has a consensus sequence of TTCTTCAGTTCTCAACAAC and occurs 3.9 times. The second repeat (6243-6327) has a consensus sequence of TCTCACAACTTCTTCAGTTTC and is repeated 3.7 times. Both of the sequences have a periodicity of ~21 nt and are located between ORFs, F61a and F62. These repeats are unique to SSV RH and not related to repetitive sequences previously identified in the SSV1 or SSV2 genomes (*82, 127*). The function or significance of these repetitive regions are currently not understood, but they may be signal sequences involved in the recognition or the regulation of viral replication.

We have identified tyrosine recombinase-like genes in the genomes of the two new SSV isolates. Both SSV RH (ORF D355) and SSV K1 (ORF F340) viral integrase-like genes contain nucleotide sequences that are duplicated in tRNA genes in the *S. solfataricus* P2 genome (Figure 3.2). These sequences represent potential *att*P and *att*A sites for directing integration of the viral genome. The integrase-like gene in SSV RH contains a contiguous stretch of 59 nucleotides that shares striking similarity to nucleotides in all five of the leucyl tRNA genes of *S. solfataricus* P2. Fifty-eight of these fifty-nine nucleotides are exactly duplicated in the fifth leucyl tRNA (L5) gene, while this sequence is less conserved in the other four leucyl tRNA genes (Figure 3.2A). The integrase-like gene in SSV K1 contains a contiguous stretch of 49 nucleotides that are exactly

duplicated in the only aspartic acid tRNA gene annotated in the *S. solfataricus* P2 genome. However, 45 of these 49 bases are duplicated in the *S. solfataricus* P2 glutamic acid tRNA genes E1 and E2 (Figure 3.2B). Based on these duplicated sequences we assumed that the SSV RH and SSV K1 genomes should integrate into different tRNA genes of the *S. solfataricus* P2 genome. In an effort to characterize the occurrence and location of SSV RH and SSV K1 integration, we designed PCR primers that flank potential integration sites on the *S. solfataricus* P2 genome (Figure 3.3 and Table 3.2). These primers were used together and in combination with viral Primers that flank the putative *att*P sites. SSV integration is modeled in Figure 3.3A and the PCR based approach is illustrated in Figure 3.3B. PCR products produced by pairing a viral specific primer with a host specific primer were sequenced. SSV RH integrates site-specifically into the fifth leucyl tRNA geneof *S. solfataricus* P2 genome (NC_002754). Viral integration occurs at the exclusion of the other leucyl tRNA genes (data not shown). Viral integration occurs in such a way that the tRNA gene sequence is conserved except for a single T to C nucleotide change in the *S. solfataricus* P2 genome. This nucleotide is located at the 5' end of the predicted T-stem and may disrupt base pairing in the structure. SSV K1 integration appears to be more promiscuous. The SSV K1 genome integrates into three different tRNA genes; D, E1 and E2. All three of these tRNA genes contain putative *att*A sites

A.

```
Sulf_P2_L5    GGGGCGGACTGAGGCTCCGCTGGCGAAGGCCTGCACGGGTTCA
RH_ORF_attP   GGGGCGGACTGAGGCTCCGCTGGCGAAGGCCTGCGCGGGTTCA

Sulf_P2_L5    AATCCCGTCCCCCGCA
RH_ORF_attP   AATCCCGTCCCCCGCA
```

B.

```
Sulf._P2_E2   TGCCGCGGCCGGGATTTGAACCCGGGTCACGGGCTCGAGAGGCCCGCAT
Sulf._P2 E1   TGCCGCGGCCGGGATTTGAACCCGGGTCACGGGCTCGAAAGGCCCGCAT
Sulf._P2_D    CGCCGCGGCCGGGATTTGAACCCGGGTCAGGGGCTCGACAGGCCCCCAT
SSV_K1_aatP   CGCCGCGGCCGGGATTTGAACCCGGGTCAGGGGCTCGACAGGCCCCCAT
```

Fig. 3.2 Sequence alignments of the viral *att*P sites and host tRNA genes. (A) Alignment of the sequence around the SSV RH *att*P site and the fifth leucine (L5) tRNA gene of *S. solfataricus* P2. (B) Multiple alignment of the sequence around the SSV K1 *att*P site, the aspartic acid tRNA gene (D), the first glutamic acid tRNA gene (E1), and the second glutamic acid tRNA gene (E2). Sequence variations are indicated by shading.

Table 3.2 Primers used in the PCR-based integration assay

Primer	Sequence
SSV RH	
Right side	
SSVRH (virus)	……GGATTCGTGAGGTTAAGGGG
P2 L1 (host)	……GCTTAGAGATGGAACTGCACCCC
P2 L2 (host)	……CGCATTCATCCATGTAACC
P2 L3 (host)	……CCATCTGGAACGTTGTTTCC
P2 L4 (host)	……GGCGTTAAAGAGGTTATGG
P2 L5 (host)	……GAGCTTCTTAACTCCGTTCCTTCC
P2 R1 (host)	……GCAACCGGAAAACCTTCTCC
Left side	
RH (virus)	……CACGCGTGATTTCCATGTCC
P2 L5 (host)	……CGCTTTCAGCTATTAGCGGGG
SSV K1	
Right side	
SSV K1 (virus)	…….CTCAGAGGGCGGATCTCTG
P2 D (host)	……..GGGAAACCCCGAGGTCCCTGG
P2 E1 (host)	……CTGAAGTACAAATGTCAGCG
P2 E2 (host)	……GGTTATTGTGAGGGATGTAGAGG
Left side	
SSV K1 a (virus)	…GCCTAGTTTCTATGTCGG
SSV K1 b (virus)	…GCCGTCTTCTTTCAATTTCTTTAC
P2 D (host)	……..CATTCTAACTCCTTCCTCGC
P2 E1 (host)	……CCCCACGTAAATTACATTC
P2 E2 (host)	……CCTACCTATACCTAACTTCTGTGC
P2 1632500 (host)	GGAAAGGTGGATAGCTAAATTGCGC

that share considerable sequence similarity with the putative viral *att*P sequence (Figure 3.2B). SSV K1 integration into the only aspartic acid tRNA gene in the genome results in an exact copy of the tRNA gene. However, SSV K1 genome integration into the glutamic acid

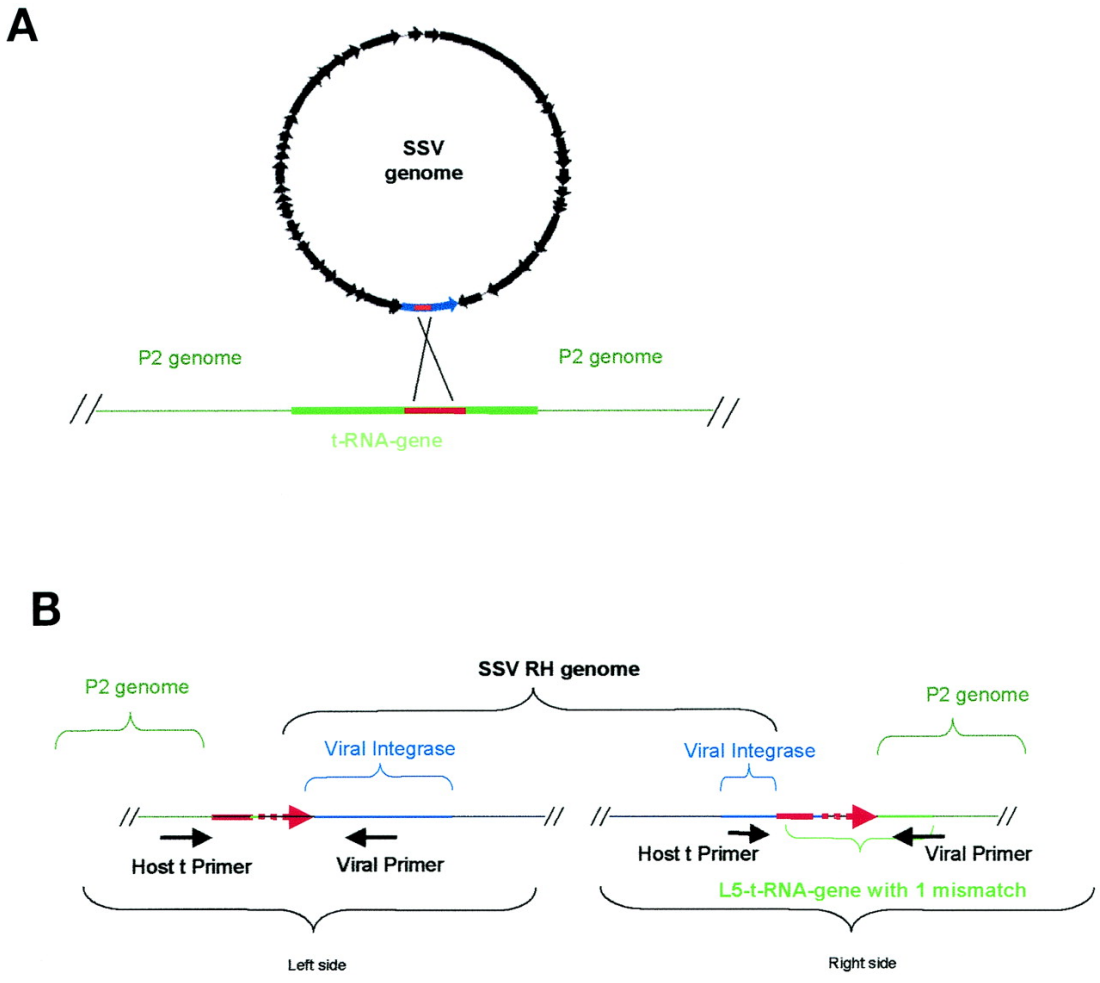


Fig. 3.3 Integration of SSV RH and SSV K1 into the *S. solfataricus* P2 genome. (A) General overview of SSV integration into a tRNA gene of *S. solfataricus* P2. The circular SSV genome is in black, the blue region represents the virus-encoded integrase, the green lines indicate host chromosome sequences, and the solid red lines indicate putative *att*P and *att*A sites in the virus and the host. (B) Schematic of SSV RH integrated into the host chromosome at the L5 tRNA gene. The general locations of viral integration and host primers that were used in the PCR-based integration assay are shown directly below the genomes. The dashed red lines indicate the region within which the recombination event occurs.

tRNA genes E1 and E2 is not as precise. Integration at both locations eliminates the 5' thymine from the tRNA genes and introduces a C to G mutation at a position corresponding to the first nucleotide of the anti-codon stem of the tRNA. These mutations may not perturb gene function, but a second mutation (G to C in E2 and A to C in E1) that we frequently observed at the first position of the anti-codon switches the anti-codon from glutamic acid to an aspartic acid. Variations in sequence at this location are a interpreted to be a result of imprecise integration of the viral genome. Apparently the exact cross over point for integration of this viral genome in either the E1 or E2 tRNA genes is not always the same. The viral and host genomes share sequences other than those identified as the putative *att*P and *att*A sites (*127, 190*). These similar sequences could potentially serve as alternative integration sites. We designed PCR primers to assay one of these alternative sites. In addition to integration at the tRNA sites we have observed integration of the SSV K1 genome into a site not annotated as a tRNA, near position 1632500 in the *S. solfataricus* (P2) genome. This region shares 39 nt with the putative viral *att*P site. This is the first report of an SSV integrating at a non-tRNA location. This discovery was surprising as bacteriophage that encode integrase proteins typically target tRNA genes and has been considered an ancient process that has been conserved during the evolution of integrating viruses (*169, 191*).

The remarkable diversity observed in the genomes of these four clearly related SSV viruses is not presently understood. The fact that there is such diversity indicates that each of these isolates should be considered different members of the *Fuselloviridae* and not as strains of the same virus. It is tempting to speculate that the observed diversity is a function of the geographic isolation of each virus, similar to that observed in their *Sulfolobus* hosts (*162*). However, first SSV viral diversity within an individual location must be established. The astonishing diversity observed in *Fuselloviruses* is reminiscent of that in the tailed mycobacteriophage (*192*). SSV diversity may be a consequence of rapid turnover rates in viral populations, differences in viral hosts and or the chemical nature of the thermal features themselves. Regardless of the cause, the genome diversity we have described should prove invaluable for guiding future experiments in understanding *Fuselloviruses* and their environments.

CHAPTER 4

NO BOUNDARIES: VIRAL MIGRATION MAINTAINS LOCAL DIVERSITY

Abstract

Viruses are the largest reservoir of genetic material on the planet, yet little is known about the population dynamics of any virus within its natural environment. Over a two year period we monitored the diversity of two archaeal viruses found in hot springs within Yellowstone National Park, USA (YNP). Both temporal phylogeny and neutral biodiversity models reveal unprecedented rates ($m \approx 1$) of viral immigration into local YNP hot spring communities. The extreme levels of viral biodiversity measured within these local hot spring communities can not be explained by mutation among members of an isolated viral community. Rather, locally diversity is maintained by viral immigration form a globally distributed metacommunity ($\Theta = 575$). These results indicate that geographic isolation does not limit viral exchange among YNP hot springs or among globally distributed thermal features. This migration model is supported by the detection of these viruses in air samples collected over YNP hot springs and by their detection in metacommunity sequencing projects conducted in the Saragossa Sea. Rapid rates of viral migration are not expected to be unique to archaeal viruses, but rather a common feature among viral metacommunities.

These findings have significant implications for understanding how this mobile genetic reservoir contributes to viral and host evolution.

Linking Phylogeny and Ecology of Viruses in Yellowstone's Hot Springs

Few studies have examined the population dynamics of an individual virus type within its natural environment. Worldwide surveillance of viral pathogens, such as influenza and HIV, reveal a global distribution where mutation is the primary driver of viral community composition through time (*193, 194*). Molecular phylogenetic approaches have been useful in characterizing these highly diverse and dynamic systems (*195, 196*). However, this study is the first time temporal phylogenetics and neutral biodiversity approaches have been combined to examine the affects of migration along with mutation to further understand viral community dynamics.

Hot springs are attractive systems for examining virus populations. High temperature (>75ºC) acidic (pH<4) hot springs found in YNP are dominated by a limited number of prokaryote hosts typically belonging to 2-10 different archaeal species (*197, 198*). In addition, these hot spring environments are small and are thought to be relatively closed systems, receiving little input from the surrounding environment.

We and others have recently described a number of unusual viruses replicating within archaeal hosts from YNP's acidic hot spring environments (*10,*

117, 122, 123). Two DNA virus types commonly found in high temperature acidic thermal features worldwide are the *Sulfolobus* spindle-shaped viruses (SSVs) (*18, 82, 173, 199*) and the *Sulfolobus islandicus* rod-shaped viruses (SIRVs) (*119*). Both viruses replicate within *Sulfolobus* and are thought to be host specific. Genome sequence comparisons of four cultured SSV isolates from Japan, Iceland, United States and Russia suggest high levels of genetic variation maybe related to biogeographical isolation (*18*).

Understanding the mechanisms that maintain diversity within a local environment and between distant habitats has broad implications for the role of viral evolution. In an effort to gain insights into viral population dynamics we monitored SSV and SIRV populations approximately every thirty days over a two year period within three different YNP hot springs. These three hot springs are separated by more than 30 km and belong to separate thermal fields (Supplemental Figure S4.1, appendix). All three hot springs are vapor-dominated systems with no obvious inputs of surface waters. At each sampling time, total environmental DNA was isolated and the pH, temperature and general water chemistry was monitored (Supplemental Figure S4.2). PCR amplification produced near full-length 16S rDNA (*200*), SSV and SIRV genome segments (Supplemental Figure S4.3) for clone libraries. Approximately 50 random clones from both viral libraries from each sampling place and time were sequenced and subjected to phylogenetic analysis.

Phylogenetic analysis of SIRV and SSV clones demonstrates high population diversity and turnover within and between sites (Figure 4.1). SIRV was detected at most sampling times for all three sites. Sequencing of 2163 independent clones revealed 722 unique sequences that formed twelve well supported clades (Figure 4.1A). Five of these clades (A-C, I and K) were detected in all three hot springs. The remaining seven clades (D-H, J and L) were only detected in one of the three hot springs. All clades demonstrated a unique temporal occurrence. Sequencing of 785 independent SSV clones revealed 320 unique sequences that formed six distinct clades (Figure 4.2). SSV-like sequences were amplified from all three monitor sites, however not at every sampling time. The three largest clades (A-C) were represented in all three host springs while three smaller clades (D-F) were found only in a single hot spring. All four previously sequenced SSV isolates from thermal features located on three different continents, fall within the YNP clades. Sequence accumulation curves for both SIRV and SSV are not asymptotic, indicating many more sequences need to be sampled before capturing the complete SSV and SIRV viral diversity (Supplemental Figure S4.4). However, for both SIRV and SSV few new clades were detected in later sampling times indicating that the majority of dominant clades had been observed.

The presence of a particular clade at a particular time is seemingly difficult to predict. SSV or SIRV clades detected in a particular hot spring are periodically replaced within a sampling interval (Figures 4.1 and 4.2). For example, the

dominant SIRV clade A in the Crater Hills hot spring was completely displaced within 167 days by clades B, H, J and I; only to reappear as the dominate constituent 55 days later. SSV populations follow a similar trend. In the Ragged Hills hot spring the dominant SSV clade A is replaced within 66 days by clades B and C. This apparent stochastic visitation of SIRV and SSV clades among each of the three hot spring environments supports an ecological model where migration dominates resident population structure. While our experimental protocol may have resulted in some unaccounted sampling bias, the same relative bias is present between the different hot springs and is consistent over the two year sampling time period. Notwithstanding possible sampling biases, it is evident that making inferences from a single time sample, as is common with many studies, would be spurious and would not be an accurate reflection of the viral community. A single sample approach would lead one to conclude that the viral communities are isolated from one another. In contrast, our long term sequential sampling demonstrates that all three hot springs have similar viral communities that are not biogeographically isolated.

Phylogenetic evidence rules out mutation as the primary influence determining local community viral dynamics. If mutation was driving clade dynamics within hot springs, we would expect to see a terminal branch ramify through time. We do not observe this pattern. Instead, we observe entire clades disappearing and being replaced (Figures 4.1 and 4.2). This suggests

migration, rather than mutation, is primarily responsible for local diversity. This interpretation is supported by pairwise temporal samples that show high viral sequence dissimilarity across a broad spectrum of sampling intervals (Figure 4.3). The high dissimilarity persists when comparing samples collected from each hot spring, indicating that the high population turnover is independent of both time and sample location.

Traditional explanations for viral biodiversity include temporal shifts in host availability, habitat heterogeneity, and mutation (*201*). These explanations do not adequately account for the temporal clade shifts that we observe in both SSV and SIRV populations. In fact, water chemistry and host community composition within a given hot spring, appeared to remain relatively stable over the two year time course, while viral clade shifts occur frequently. Each hot springs had its own unique water chemistry (Supplemental Figure S4.2) and archaeal microbial community (Supplemental Figure S4.5), with *Sulfolobus* being the only common archaeon detected at all three sites. Unmeasured environmental parameters responsible for viral dynamic population shifts are marginalized by the observation of the same dominant SSV and SIRV clades appearing in all three distinct hot springs, despite dramatic differences in their water chemistry and archaeal host composition.

The rank abundance of all subsets of SIRV and SSV sequences were fitted against distribution models. In contrast to other studies, power law functions

(i.e., Zipf and Mandelbrot) (*202*) fit less well than Etienne's sampling formal (EtSF; Figure 4.4). Unlike traditional models of rank abundance (e.g., the lognormal), the EtSF (*3*), which is a genealogical modification of Hubbell's zero sum multinomial (*203*), uses biologically realistic parameters (e.g. migration and speciation as used by Hubbell) for analyzing virus diversity within these hot springs. Central to this theory is the interplay of viral speciation (viral clade formation) and migration during the generation of metacommunity diversity. In populations where migration is negligible, the number of new viral clades produced per metacommunity generation, Θ, is a function of mutation. When the probability of resident extinction being replaced by a migrant (*m*) becomes high, migration becomes the main cause of local viral diversity. Using the EtSF, we have estimated a mean Θ of 575 and an *m* of 1.0 (Figure 4.4), which were independently validated using a Bayesian approach (*204*). However, a model estimating only Θ, shows a better fit using a likelihood ratio test (*205*) and AIC selection criteria (Figure 4.4), suggesting that these local Yellowstone hot springs are not dispersal limited (*195, 196, 202, 206, 207*) and reinforcing the expectation that a resident death will likely be replaced by an immigrant. These results are in contrast to community studies of the viral host, where migration is negligible (*162*) and further highlights how cosmopolitan viral communities may be contributing to the evolution of isolated host populations. The combined

values for virus migration and speciation reported here are possibly the highest yet observed for any community.

We have investigated possible modes of viral migration. Two obvious mechanisms are movement of viruses through the air or underground water movement between hot springs. We speculate that transport through the subsurface waters is unlikely, due to the subterranean superheating of water that feeds the Yellowstone caldera (>350C at 100m) and the unlikely prospect for globally exchanging metacommunities through the superheated aqueducts. In contrast, air transport of microbes is known to occur on a global scale and virus movement on a local scale is well establishes (i.e. movement of respiratory viruses through air and independent of any viral vector species). Preliminary results for quantitative PCR experiments have detected ~15,000 SSV genomes per liter. These findings suggest that virus transport through the air is a possible mechanism for virus migration.

The role of virus migration reported here and the potential global size of the hot springs viral metacommunity bolster the already high estimates of virus diversity (*206*). If our observation of viral migration proves to be generally applicable, then understanding the dispersal mechanisms that maintain high levels of viral diversity within and among distant habitats will be essential to understanding the evolution of these and other widely distributed viruses.

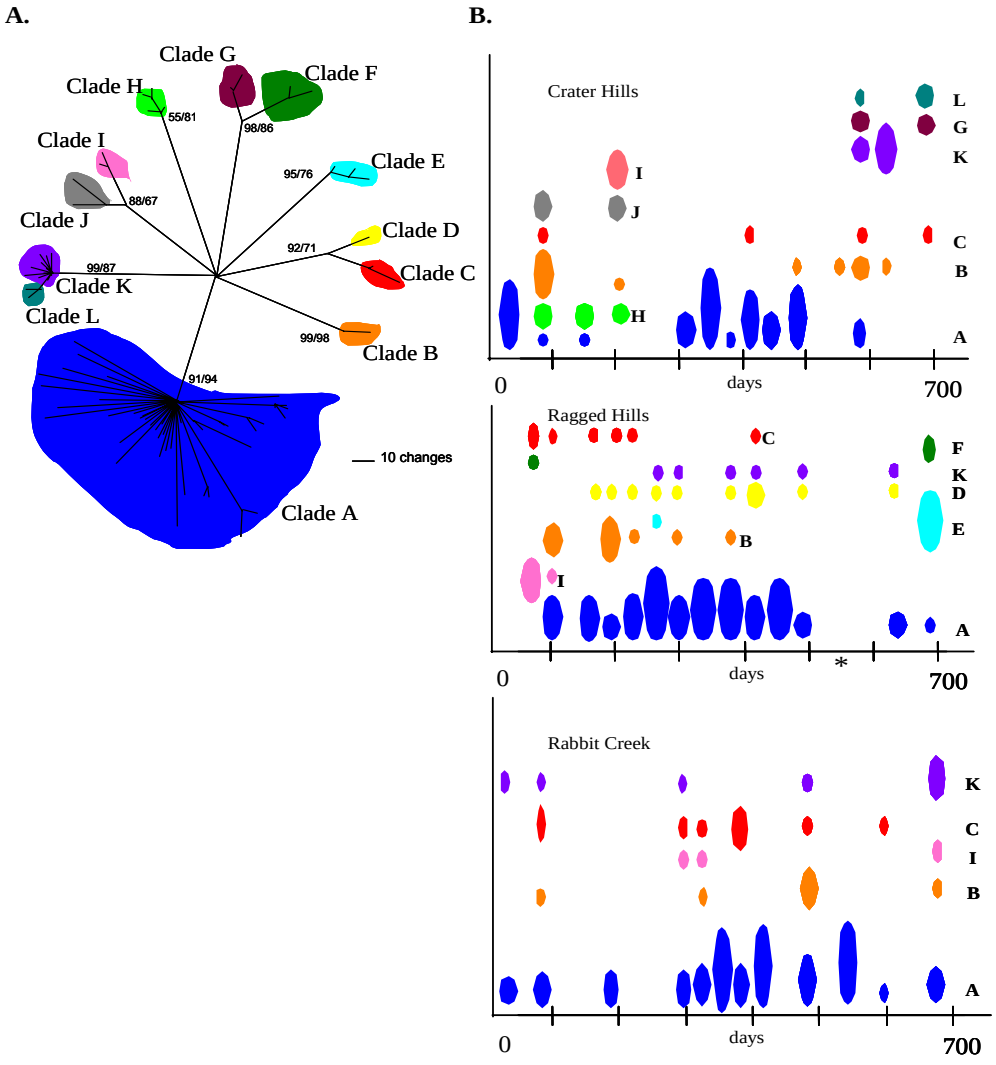


Fig. 4.1 A diverse SIRV community transiently visits geochemically distinct hot spring environments. A. SIRV phylogentic tree comprised of 2163 clones from three hot springs obtained during the 688 days of sampling. Twelve well supported clades (A-L) with both maximum parsimony bootstrap values (10,000 resampling), and maximum likelihood clade credibility values are indicated. B. Dominant SIRV clades detected over time within the Crater Hills, Ragged Hill and Rabbit Creek hot springs. The size of the oval schematically represents the relative number of sequences belonging to a particular clade detected at a particular sampling time. Over the nearly 2 year sampling period, clade representation at a given hot springs is transient and unpredictable, suggesting migration rather then mutation is dominating community dynamics.

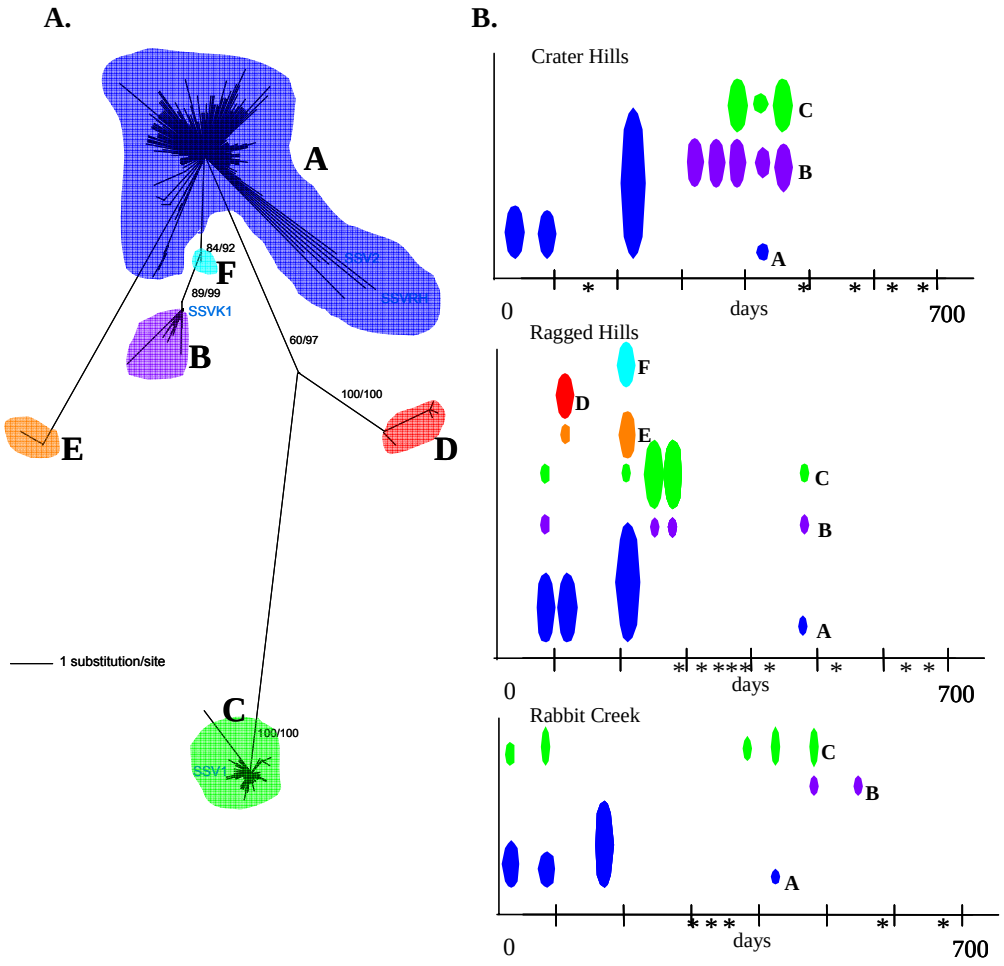


Fig. 4.2 A diverse SSV community transiently visits geochemically distinct hot spring environments. A. SSV phylogentic tree comprised of 785 clones from three hot springs obtained during the 688 days of sampling. Six well supported clades (A-F) with both maximum parsimony bootstrap values (10,000 resamplings), and maximum likelihood clade credibility values are indicated. B. Dominant SSV clades detected over time within the Crater Hills, Ragged Hill and Rabbit Creek hot springs. The size of the oval schematically represents the relative number of sequences belonging to a particular clade detected at a particular sampling time. *Indicate sampling times in which SSV sequences were not detected. Over the nearly 2 year sampling period, clade representation at a given hot springs is transient and unpredictable, suggesting migration rather then mutation is dominating community dynamics.

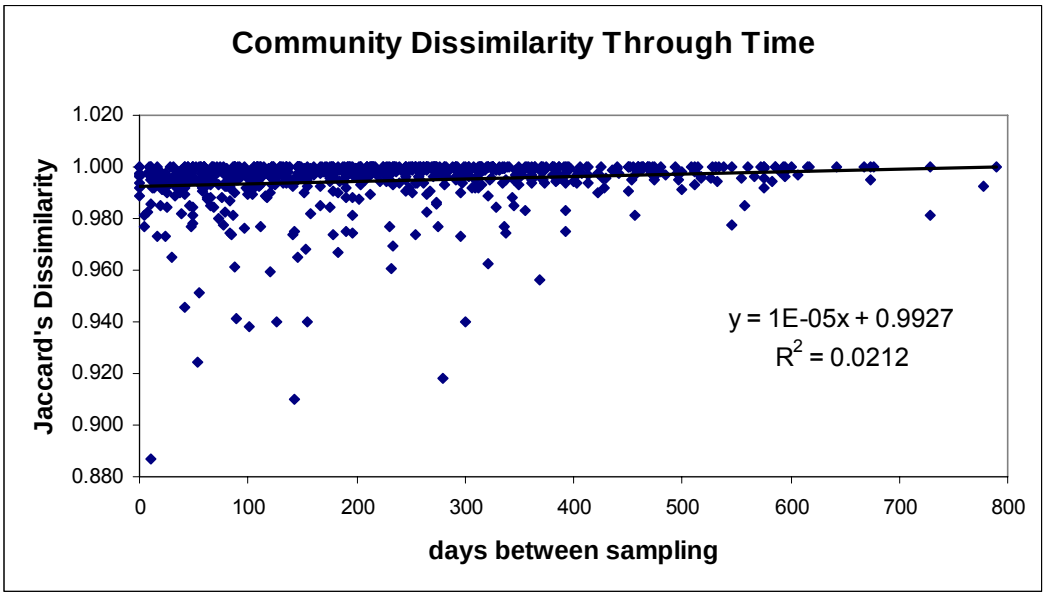


Fig. 4.3 Viral community comparisons between all pairwise temporal sampling events at Crater Hills, Rabbit Creek and Ragged Hills hot springs. Jaccard's dissimilarity was estimated using sequences shared among temporal samples. Samples collected from different hot springs on the same day, are greater than 88% different. High dissimilarity persists when comparing samples collected from any hot spring over the entire time course. This regression line illustrates high population turnover that is independent of time and sample location.

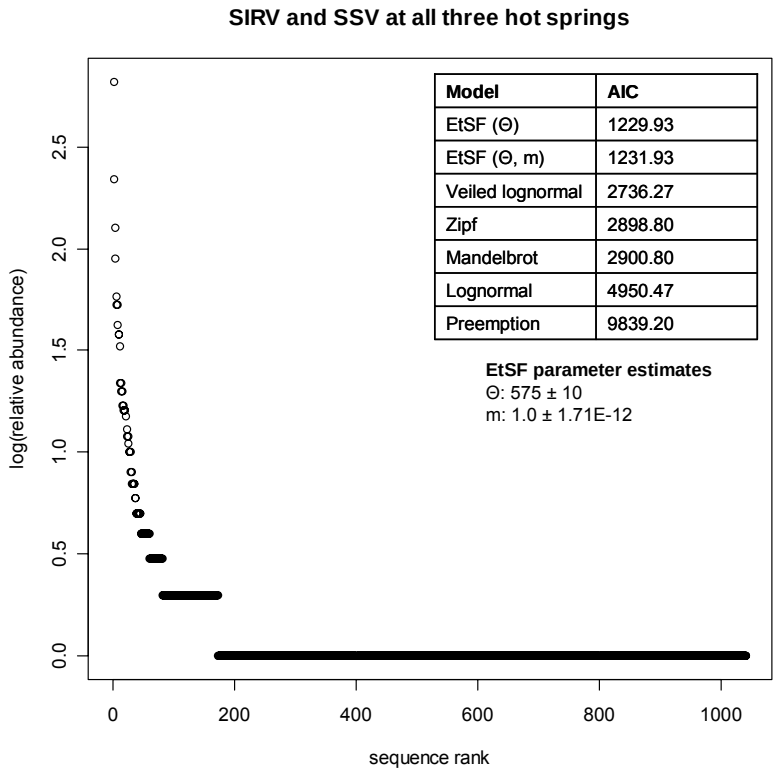


Fig 4.4 Rank abundance distribution of 2944 sequences in 1042 sequence types for SIRV and SSV sampled at all three hot springs during the nearly two year study. Etienne's sampling formula (EtSF) consistently showed the best fit to these data, even in subsets that excluded the 870 singleton sequences (leaving 2074 sequences in 172 sequence types) or when only the 11 most abundant sequences were analyzed (leaving 1408 sequences). Distributions were fit using R code (*1, 2*) and Tatame (*3*).

Material and Methods

Sample Collection and DNA Isolation

Total environmental DNA was extracted from 500 milliliter samples collected from each hot spring, at each time point, using the UltraClean™ Water DNA Extraction Kit (Mo-Bio, Solana Beach, CA).

Water Chemistry

Water hot spring samples were collected and preserved according to Energy Laboratories (Billings, MT). Each sample was analyzed for the reported (S2) inorganics (chloride and sulfate), organics (dissolved organic carbon), nutrients (ammonia, total nitrogen and phosphorus), and metals (aluminum, antimony, arsenic, barium, boron, cadmium, calcium, copper, iron, lead, magnesium, manganese, mercury, molybdenum, nickel, potassium, selenium, silicon, sodium, and zinc).

PCR, Cloning and Sequencing of Archaeal and Viral Sequences

Archaea 16S rDNA sequences were amplified from total environmental DNA using archaeal specific primers 20F (ttccggttgatcctgccgg) or 348Fa(tccaggccctacggg) paired with the universal primer UA1406R (gacgggcggtgtgtrca) (*200*). SSV and SIRV sequences were PCR amplified using primers identified by comparative viral genomic studies. Universal SSV primers UnvSSV 3F (caatcgccataggctacgg) and UnvSSV 4R(cgttta(c/t)tactataacggtac)

were designed using the complete genome sequences of four SSV isolates (*18*). These primers amplifiy ~256 base pairs of the largest conserved ORF. SIRV primers DBPF (gatattgaccaaaaatggcaaaagg) and DBPR (gcttaaatattttaacttacgtatcc) amplify the entire 405 base pairs (bps) of the coat protein gene. These primers were designed using the complete genome sequences of two Icelandic isolates (*142*) (SIRV1 and SIRV2) and partial genome sequences from two Yellowstone isolates (unpublished). PCR products were cloned into pCR2.1 using TOPO-TA cloning protocols (Invitrogen, Carlsbad, CA) and sequenced using Big Dye Termination protocols on an ABI 3700 automated capillary sequencer (Applied Biosystems, Foster City, CA). To reduce the effects of possible sequencing errors, only sequences that contained more than one base change and did not encode for stop codons when translated in the appropriate frame were used in subsequent phylogenetic and rank abundance analysis.

Phylogenetic Analysis

DNA sequences were manually edited using Sequencher 4.2.2 (Gene Codes Corporation, Ann Arbor, MI). Sequence alignments were performed using CLUSTALX (*208*). The alignments were subsequently analyzed by PAUP 4.0b10(*187*) and a maximum parsimony analysis was conducted. Bootstrap values were obtained by resampling the data 10,000 times and a bootstrap 50% majority-rule consensus tree was produced. An independent maximum likelihood

analysis was conducted using MrBayes (*188, 189*). Clade credibility values were derived using two million permutations were sampled every ten thousand generations.

Dissimilarity index, rank abundance modeling and estimations of migration and metacommunity.

Modeling of sequence abundance curves were performed using ecological models implemented in the Radfit program in the Vegan package (*1*), which is written for R statistical environment (*2*). Abundances were analyzed using the zero-sum-multinomial distribution and Bayesian methods using R code (*204*).

CHAPTER 5

KILLING THE MESSENGER – A NOVEL PROKARYOTIC IMMUNE SYSTEM BASED ON RNA INTERFERENCE

Abstract

RNA interference (RNAi) is a recently established mechanism responsible for regulating gene expression in eukaryotes. Although initially observed in plants, the process is currently considered to be an evolutionarily conserved process present in the entire eukaryotic domain where it originally evolved as a defense mechanism against viruses and other foreign nucleic acids. Here we use the thermoacidophilic archeaon *Sulfolobus solfataricus* as a model to characterize the prokaryotic immune response to viral infection. Northern blot analysis supports a an RNAi-like role for short host encoded viral-like sequences that are distributed within defined Clusters of Regularly Interspaced Short Palindromic Repeat (CRISPR) regions that are prevalent in prokaryotic genomes.

Introduction

RNA's that do not function as messenger RNA, transfer RNA or ribosomal RNA are generally referred to as non-coding RNA (ncRNA). ncRNAs range in size from ~21 nt to greater than 10,000 nt and number from the hundreds to the thousands per genome (*209-211*). The versatility of these RNAs in regulating exogenous as well as endogenous gene expression has brought new depth to

our previously naive perception of the complex regulatory mechanisms responsible for controlling gene expression. There are currently two major categories of regulatory ncRNAs; micro (miRNAs) and small interfering (siRNAs). In eukaryotes, siRNAs are generated from larger double-stranded RNA (dsRNA) typically produced during the life cycle of an invading extra-chromosomal element. dsRNA is recognized by an RNaseIII-like enzyme, termed Dicer, that randomly fragments the dsRNA into ~19-27 nt snippets (*212*). These processed small RNA duplexes are then unwound in an ATP dependent process and incorporated into the single stranded binding pocket of an Argonaute protein. Together the siRNA and the Argonaute protein are collectively referred to as the RNA induced silencing complex (RISC). The siRNA serves as a targeting oligo, duplexing with its complement and triggering the RNase (PIWI) domain of the Argonaute protein to specifically cleave a single phosphodiester bond on the target sequence across from nucleotides 10 and 11 on the siRNA (*213-215*).

miRNAs primarily differ from siRNAs in their origin and processing, rather than their function (*212*). Primary miRNA transcripts are generated from both intra- and inter-genic regions of the chromosome by RNA polymerase II, from individual promoters as well as polycystronic transcripts (*216*). In many cases both the promoter and the miRNA are evolutionarily conserved among related species (*211, 217*). Functional miRNAs are generated in a two step process. First, self-complementing sequences within the primary transcript form stem loop

structures (~70 nt), which are the hallmark of miRNAs. These structures are recognized by a dsRNA specific ribonucleases (i.e. Drosha) that cuts the stem loop from the larger transcript (*216, 218*). Liberated stem loop structures are then processed by Dicer and bound by an Argonaute protein in a manner similar to that described for siRNAs. The prototypic miRNAs, *lin4* and *let7*, were shown to target RISC to mRNAs by imperfect base pairing; an interaction that blocked translation rather than triggering mRNA degradation (*219, 220*). Although, not a hard-and-fast rule, it appears that perfect base pairing between either class of regulatory ncRNAs (siRNA or miRNA) and a target mRNA more frequently elicits targeted degradation, whereas imperfect binding results in translational blocking (*212, 216, 221, 222*).

Repeat-associated small interfering RNAs (rasiRNAs) are a third type of ncRNA that have recently been implicated in silencing transposons and other repetitive sequences in germline cells (*223-226*). Although secondary structures formed by repeat regions would seem to mimic that observed in miRNAs, rasiRNAs are processed by in a distinct, yet defined pathway that yields slightly larger (~25-29 nt) dsRNA fragments (*226*). The antisense orientation of rasiRNAs are preferentially bound by a Piwi protein that is related to but distinct from the various Argonaute proteins that bind miRNAs and siRNAs (*226*).

Repeat-associated small ncRNAs have recently been detected in total RNA extracts harvested from several archaeal species (*227-230*). Neither repetitive

DNA nor small non-coding RNAs are particularly unusual in prokaryotic genomes. However, what is remarkable is that short spacer sequences (35-40 nt) found within these distinct repetitive DNAs commonly correspond to cognate sequences found in selfish genetic elements (*231-233*). These repeats, recently coined CRISPRs (Clusters of Regularly Interspaced Short Palindromic Repeat), have been identified in all but one archaeal genome and are present in 55% of the currently sequenced bacterial genomes (*227, 234*). The broad phylogenetic distribution of CRISPR-like elements, which appear to be actively transcribed, is suggestive of an important biological function. The frequency of spacer sequences that are homologous to viral, plasmid and transposable elements suggests that CRISPRs maybe central to a novel RNAi-based prokaryotic immune system analogous to the well established antiviral defense system in plants and insects (*235-237*). Further support to this emerging hypothesis has recently been offered by genomic comparisons that have identified a variable cassette of ~25 gene families that are commonly associated with CRISPRs (*234, 238, 239*). With the benefit of these new functional clues, many of these CRISPR associated (*cas*) genes have been assigned putative functional roles analogous to those involved in the prototypic eukaryotic RNAi based gene regulation system (*239*).

The rules of evolution are universal to all of life. Like eukaryotic cells, prokaryotic life must also defend itself against parasitic infection. Historically, our appreciation for the prokaryotic 'immune system' has been limited to

restriction/modification systems (*240*). However, more recent studies suggest that prokaryotic life may have evolved a more sophisticated RNAi-like immune system. Here we use two different viruses of the thermoacidophilic archaeon *Sulfolobus solfataricus* to characterize the host response to challenge by viral infection.

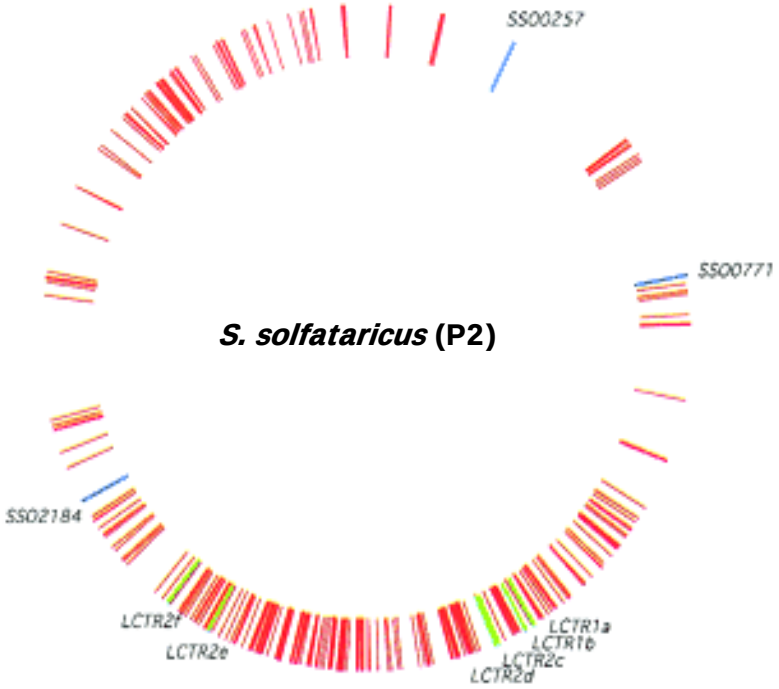
Preliminary Results

Archaea: A Hot Model for Assessing the Prokaryotic Immune Response

Sulfolobus solfataricus is a thermoacidophilic prokaryote that belongs to the archaeal domain of life (*29, 78*). *Sulfolobus* spp. were originally isolated from sulfotaric hot spring in Yellowstone National Park more than three decades ago (*76*). *Sulfolobus* spp. have since proven tenable to laboratory cultivation and have been established as versatile model organisms for studying nearly every aspect of archaeal biology. A virus free strain of *S. solfataricus* has proven valuable for thermal virology and is often used as a host for establishing primary infections and thus completing Koch's postulates. The complete genome sequences for three different *Sulfolobus* spp. reveal considerable flexibility in the CRISPR associated System (CASS). Each of the 3 genomes code for 5 to 6 CRISPRs, that range dramatically in size (~7.5 kb to ~400 nt) and are flanked by a variable set of the *cas* genes. The repeat sequence is generally conserved within a CRISPR, however sequence and length varies slightly among CRISPRs.

CRISPR repeats are generally more conserved within a genome and between closely related species but share limited conservation among phylogenetically diverse prokaryotes. The six *S. solfataricus* CRISPRs are loosely arranged in two separated clusters on the chromosome, where repeats in neighboring CRISPRs are notably more similar (Figure 5.1). As previously reported, spacer regions separating CRISPR repeats are often homologous to sequences found in selfish genetic elements (*227, 231-233*). Six different spacer regions in the *S. solfataricus* F-CRISPR are homologues to sequences in either the *Sulfolobus* spindle-shaped virus Ragged Hills (SSVrh) or the *Sulfolobus* turreted icosahedral virus (STIV) (Figure 5.2). These two viruses exhibit distinct life cycles within the *S. solfataricus* host. SSVrh is a temperate virus, with an exterior lipid envelope that may be derived from the host during budding (Supplementary S5.1). Viral transcripts are observed 2.5-8.5 hrs after infection and virus production is persistently maintained over the course of the growth curve (Figure 5.3B and C). In contrast, STIV exhibits a protracted life cycle, with viral transcripts first appearing at ~24 hrs and peaking at 36-48 hrs post infection (Figure 5.3B). A single burst of STIV occurs consistently at ~72 hrs post infection. This single burst event is consistent with a lytic life style; however the viral burst is not accompanied by a dramatic decrease in culture density. Preliminary data from samples collected over a time course of the STIV infection indicate that only a subset, ~1 in 10 cells are infected (Ortmann EFM and Brumfield TEM

A



B

A) GATTAATCCCAAAAGGAATTGAAAG
B) GATTAATCCCAAAAGGAATTGAAAG
C) GAT-AATCTCTTATAGAATTGAAAG
D) GAT-AATCTCTTATAGAATTGAAAG
E) GAT-AATCTACTATAGAATTGAAAG
F) GCT-AATCTACTATAGAATTGAAAG

Fig. 5.1 *S. solfataricus* (P2) CRISPRs are loosely arranged in two separated clusters on the chromosome. A) CRISPRs are annotated as long clusters of tandem repeats (LCTR) in the *S. solfataricus* genome and order a-f (*6*). B) *S. solfataricus* (P2) repeats are 24 nt long, with highest sequence conservation between neighboring repeats. (Figure adaped from She *et al* 2001.)

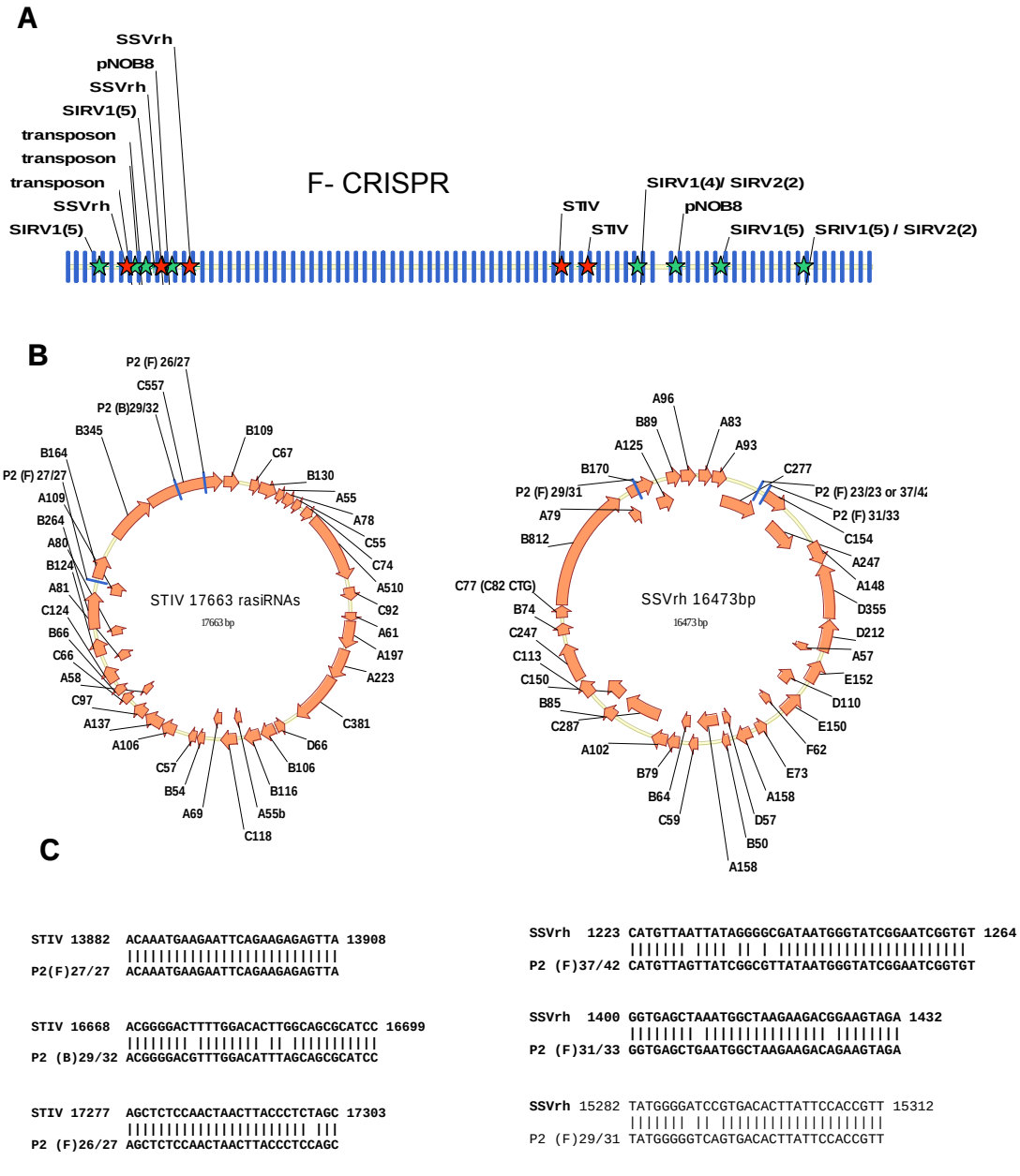


Fig. 5.2 Many of the spacer regions in the *S. solfataricus* F-CRISPRs are homologues to sequences in characterized selfish genetic elements. A) Schematic of the F-CRISPR, where vertical blue lines indicate repeat sequences. Red stars indicate the location of spacer sequences homologous to sequences in either the SSV or STIV genomes. B) Putative target sequences are indicated by blue lines in respective viral genome maps. C) Sequence alignments of viral sequences with the cognate sequence found in the P2 F-CRISPR.

unpublished data). Efforts to infect more cells in the culture, by increasing the multiplicity of infection (MOI) have been unsuccessful. This has been explained by heterogeneity within the clonal population, rendering a subset of the population unsusceptible to infection (A. Ortmann, personal communication). However, the nature of this heterogeneity has yet to be explained. One

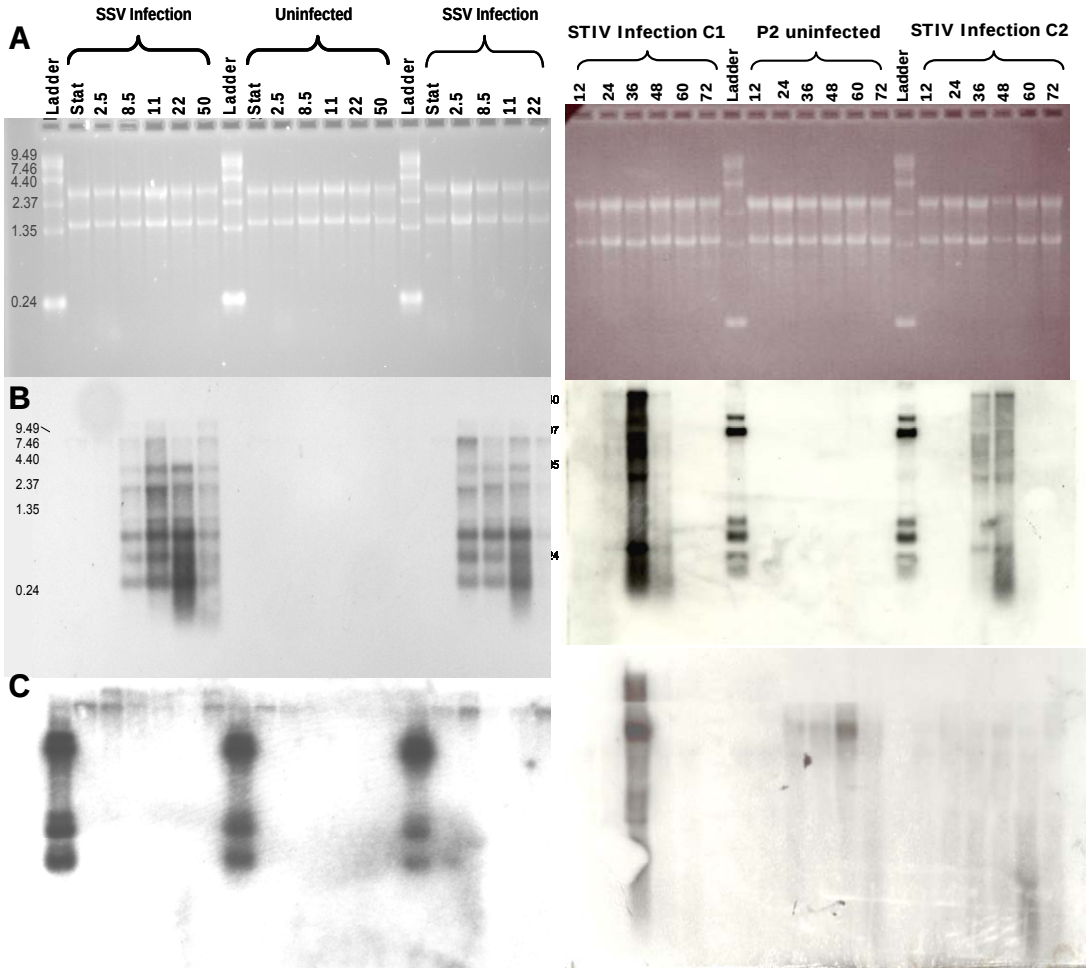


Fig 5.3 Northern analysis across a time course of both SSV and STIV infections. Row A) 1.5 μg of total RNA isolated at each time point; separated on a 1.5% agarose formaldehyde gel Row B) Northern Blot; probed with 32P-labeled SSV or STIV genomic DNA. Row C) Northern Blot; probed with 32P-labeled F-CRISPR PCR product

possibility, would be that early log phase *S. solfataricus* cells (OD~0.2) exposed to virus mount an adaptive immune response that is heritable and thus subsequent generations are no longer susceptible to infection. Preliminary evidence supporting a CRISPR-based RNAi-type immune response to STIV infection is provided by northern blots, where peak expression of STIV transcripts is concurrent with a peak in F-CRISPR expression (Figure 5.3B and C).

The co-occurrence of viral and host encoded F-CRISPR transcripts is notably different in the SSV infections. SSV transcripts are detected at the first sampling point (2.5 hrs) and are maintained over the time course. The persistence of this infection is highlighted by the absence of F-CRISPR expression (Figure 5.3B and C).

Can difference in the SSV and STIV infection cycles be explained by differences in F-CRISPR response? A definitive answer to this question awaits genetic disruption of the F-CRISPR. This is particularly important in the STIV infection where small RNA homologues in the host chromosome are expected to be involved in the targeted silencing of the viral transcripts. The absence of F-CRISPR expression in response to SSV infection is curious, especially given the presence of SSVrh-like spacer sequences. This however is not entirely unexpected; as virus and host co-evolve, host defense mechanisms are countered by virally encoded disruption strategies. The distinct induction profiles of the F-CRISPR by both viruses does correlate with the differences in virus

propagation, as well as with the apparent period of viral gene expression. A variety of RNAi suppression strategies have been identified in plant and insect viruses. Although speculative, it is possible that a virally encoded DNA binding protein could specifically recognize the palandromic signature of CRISPRs and thereby prevent its transcription.

A Collaborative Strategy; Microarray, Proteomics and Biochemistry

A *S. solfataricus* (P2) microarray that includes oligos representing both viral genomes has been constructed. Six cultures of *S. solfataricus* were grown in parallel; three of these were infected with STIV during early log phase at an MOI of 0.33, 0.4 and 2.7, respectfully (A. Ortmann, unpublished). Thus far no significant differential *cas* gene expression has been detected. This experiment is currently being repeated with an SSV infection. In addition to Microarray analysis our collaborators are complementing these studies with a proteomic approach (Bothner lab). Identifying *cas* genes that are responding to viral infection will be critical to further characterizing the prokaryotic immune response. However, understanding the precise role of each *cas* gene will require biochemical characterization. This work is being conducted in collaboration with the Van der Oost lab in Wageningen, NL.

Material and Methods

Culturing

Sulfolobus sulfataricus, strain P2, was cultured in media 182 (pH~2.8) in 2-liter long neck culturing flasks at 80ºC. Both SSV and STIV infections were preformed in triplicate. STIV infections were done at early log (OD_{650}~0.1) at an MOI of 0.33, 0.4 and 2.7. SSV infections were done at early log (OD_{650}~0.1) at an MOI of 0.03. Control cultures were handled identically without addition of virus. 200ml of each culture was sampled at the indicated sampling times. Cells were harvested by low speed centrifugation.

RNA Isolation and Quality Assessment

Total cellular RNA was extracted from *S. solfataricus* (P2) cells, according to the Qiagen's RNeasy midi protocol, with an on-column DNase step (Valencia,CA). Total RNA concentrations were estimated using a Nanodrop spectrophotometer ($OD_{260/280}$). RNA quantity and quality was independently assessed by visualization on a 1.5% agarose (wt/vol) formaldehyde gels and on an Agilent 2100 Bioanalyzer (Agilent Technologies, Palo Alto, CA).

Northern Analysis

For Northern analysis, ~1.5 µg of total RNA was separated by electrophoresis in a 1.5% agarose (wt/vol) formaldehyde gel and transferred to GeneScreen membranes as recommended by the manufacturer (NEN). RNA was membrane-

crosslinked in a UV Stratalinker (Stratagene). Blots were probed with probed with ^{32}P-labeled SSV or STIV genomic DNA or with a ^{32}P-labeled F-CRISPR PCR products (Ready-To-Go Labeling Kit, Amersham Pharmacia Bioscience).

CHAPTER 6

PERSPECTIVES AND FUTURE DIRECTIONS IN THE FIELD OF THERMAL VIROLOGY

This chapter is intended to identify emerging and existing areas of interest in thermal virology that are expected to yield high impact discoveries. Suggestions for future experimental directions are offered.

Advancing the Field of Thermal Virology

Extra-chromosomal elements of thermophilic Archaeal species were first recognized in 1982 by the late Dr. Wolfram Zillig and co-workers. Since then a rather extensive collection of extra-chromosomal elements; including cryptic and conjugative plasmids as well as viruses, have been cataloged. However, only a few of these have been characterized beyond the level of basic sequence information and because much of this sequence has proven to be largely unique, little use has been made of homology-based functional prediction tools. Comparative genomic analysis of the *Fuselloviridae* has identified a common set of 18 genes conserved among all currently sequenced SSV isolates (*7*). Two of these 18 common genes have been recognized as structural proteins by N-terminal sequencing (*171*). A third conserved gene (SSV1 D335) was predicted to be related to the type I tyrosine recombinase family of integrase proteins (*177*). Follow-up *in vitro* biochemical characterization has demonstrated both

intermolecular recombination as well as intermolecular excisive recombination of linear DNA fragments containing *att*P and *att*A sites (*147, 148*). *In vivo* assessments of how these virally encoded integrases contribute to the SSV life cycle are currently being perused in the Stedman laboratory. To my knowledge the SSV integrase is currently the only example of a sequence based functional prediction being confirmed biochemically for any archaeal virus protein.

The observed lack of conserved sequence similarity of virally encoded proteins, to proteins of known function has prompted efforts to determine crystalographic three-dimensional structures. Tertiary structure is preserved beyond that of primary sequence (*241-243*). Thus, higher order levels of homologue based comparisons offers further insights into protein evolution and function. Structural work on archaeal viruses has been done almost exclusively in the Lawrence and/or Johnson labs (Montana State University and Scripts Institute; respectively). Structural characterization of 2 SSV1 proteins have been published (*19, 20*). The SSV1 D63 structure reveals a helix-turn-helix motif that dimerizes in solution to form an antiparallel four helix bundle (Figure 6.1A) (*19*). The two monomers are related by a 2-fold axis of symmetry that is perpendicular to the long axis of the four helix bundle and runs through the center of dimer interface. Looking through the 2-fold axis presents either two antiparallel N-terminal helices (Figure 6.1B); or from the other side, two antiparallel C-terminal helices. Conserved residues identified by multiple sequence alignments of

homologous proteins form SSV2 (D57) and SSV RH (F61) are preferentially distributed on the N-terminal face of the four helix bundle. Although the

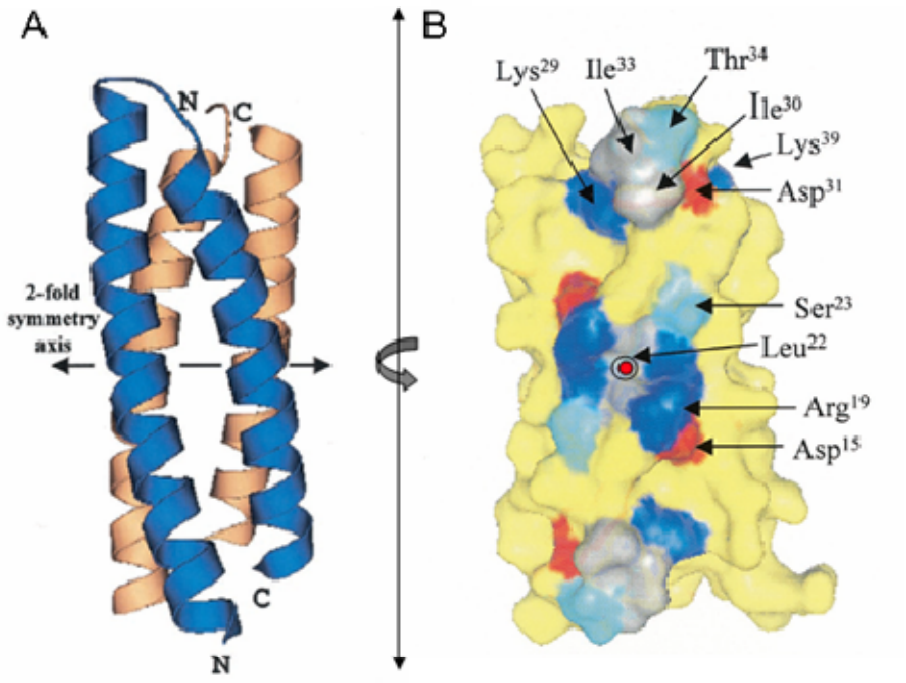


Fig 6.1 Homodimeric structure of the D63 protein from SSV1. A. Each subunit is composed of two antiparallel α-helices that are connected by a short loop (residues 35-39). The two subunits, one in blue and the other in amber, dimerize to form an antiparallel four helix bundle. B. Surface representation of conserved N-terminal residues. Basic residues are blue, acidic/red, polar/cyan, nonpolar/gray and residues not conserved among D63 homologues are yellow. Residues are symmetrically related about the 2-fold axis of symmetry (bull's eye). Conserved residues on this surface expected to be the ligand binding domain. (Figure adapted from Kraft *et al* 2004 (pdb; 1SKV))

antiparallel four helix bundle is not indicative of any particularly function, the symmetric display of these conserved residues about the two fold axis, lends itself to the hypothesis that this protein may function as an adaptor molecule that associates with two copies of a monomeric entity or with a single

macromolecule that has complementing 2-fold symmetry. One attractive hypothesis recently presented by the Lawrence group, is that D63 regulates copy number of the virus in a Rop-like (Repression of primer) manner, where the N-terminal face serves to stabilize the interaction between two complementary RNAs. Replication of the ColE1 plasmid is regulated by two RNAs, termed RNAI and RNAII (*244*). RNAII is a primer for plasmid replication that binds a complementary sequence near the plasmids origin of replication. Replication of the plasmid is regulated by RNAI, which interacts with RNAII via a complementary sequence termed the "kissing complex" (*245, 246*). In the current model, symmetrically related residues at the dimer interface of Rop/D63, bind stem loop regions of RNAI and RNAII and thus facilitates complementary base pairing (*244*). The Rop RNAI-RNAII association appears to be transient, which may indicate that the newly stabilized RNAI-RNAII duplex is released and that Rop can serve as a multi-turnover protein, facilitating RNAI-RNAII stabilizing interaction.

The genomic context of D63 also offers insights its possible functional role. D63 is coded on a polycystronic transcript (T5) that includes 10 other ORFs, one of which is the virally encoded integrase D355 (*147-149*).

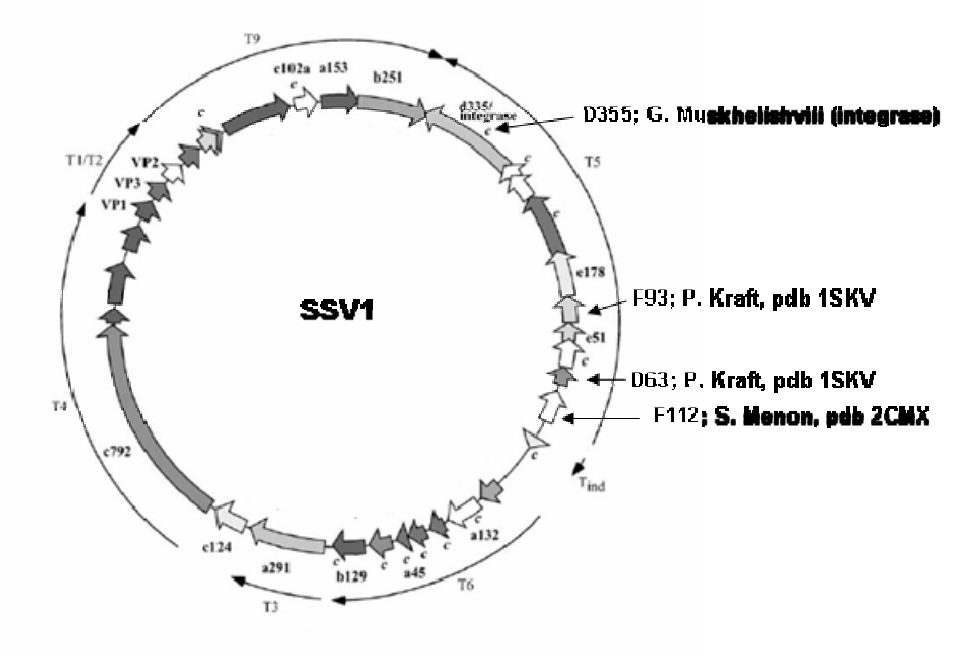


Fig 6.2 Transcript map of the SSV1 genome. Three protein structures, that are coded on transcript number 5 (T5) have been solved. Proteins coded on this transcript may participate in a common physiological process. Three structures on T5 have been solved; authors and pdb identification numbers are noted. Cysteine containing ORFs are indicated with a 'c' (Figure adapted from Stedman *et al* 2003).

Co-regulation of the D63 and the viral integrase gene maybe consistent with D63 having Rop-like function. Together, these genes and the others on this transcript may be involved in processes related to the viral latency. In addition to D63, two additional protein structures from ORFs on this transcript have been solved by the Lawrence lab (Figure 6.2). F112 (S. Menon; pdb, 2CMX) and F93 ((*20*);pdb, 1TBX) are both wHTH proteins. The ordered residues of F112 (S. Menon; pdb, 2CMX) assemble into a 3 stranded wHTH protein that is a monomer in solution. The C-terminal 40 residues of this structure are not visible.

The F93 monomer consists of four α-helixes and two β-stands (Figure 6.3; pdb; 1TBX). The first three α-helixes (H1-H3) are connected by two short loops and constitute the tri-helical core that is common among proteins belonging to the HTH superfamily (*247*). The third helix of the F93 monomer is followed by two β-strands that are connected by a reverse turn and form the flanking 'wing'. The wing and the helix-turn-helix motif (H2-H3) constitute the minimal fold of the winged-HTH subclass of DNA binding proteins. This minimal fold is commonly adorned with N or C-terminal elaborations; thus a C-terminal extension on F93, which forms a forth α-helix is not an unusual feature among members of this protein family. A VAST search for structural homologues identifies the SlyA (pdb, 1LJ9) (*150*) and MarR (pdb, 1JGS) (*151*) subfamilies as nearest structural neighbors. Both of these wHTH superclass proteins have C-terminal helices. The presumptive DNA recognition helices (H3) of each F93 monomer are separated by ~25Å and related by a 2-fold axis. This arrangement is expected to mediate the proteins association with a palindromic (or pseudo-palindromic) DNA sequence.

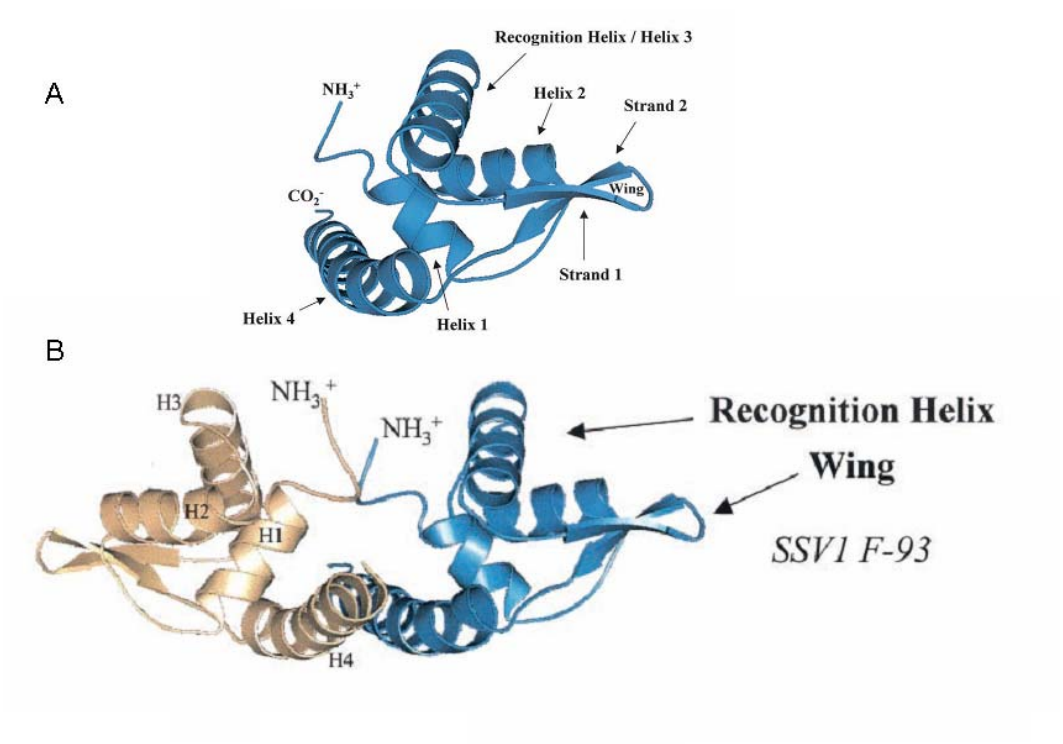


Fig 6.3 F93 is a homodimeric wHTH protein. A. The F93 monomer consists of four α-helices and two β-strands. The antiparallel β-strands from the wing, while the third helix is positioned to bind the major groove of the DNA double helix. B. Hydrophobic interactions at between N-termini and the forth α-helix of two monomers from much of the homodimeric interface. (Figure adapted from Kraft *et al* 2004 (pdb; 1TBX))

Given these new structural insights, several experiments are warranted. First; construction of an SSV1 knockout of the D63 gene. If this protein functions as a repressor of DNA replication, then we would expect the knockout to have uninhibited replications. This would not only provide substantial support for the Rop-like activity of this protein but could potentially stock the lab with limitless amounts of virus for structural and biochemical characterization. This is of course speculative and the D63 knockout could prove to be unstable or otherwise non-viable. However, the molecular tools for this type of knockout experiment are now currently in place. Kraft *et al* suggest efforts directed at identifying D63 binding partners. A combination of these two strategies will be critical in defining the precise role of D63 in the SSV life cycle.

F93 is a homodimeric wHTH protein. Accordingly, this protein is expected to bind a palindromic or pseudo-palindromic DNA sequence (*20*). The apparent absence of an appropriately spaced palindrome in the virus genome suggests that the protein may target a host encoded sequence. Although speculative it is irresistible to wonder if F93 or perhaps other virally encoded DNA binding proteins could bind to host encoded Clusters of Regularly Interspaced Short Palindromic Repeat (CRISPR). CRISPRs are anticipated to be the source of small interfering ncRNAs that target posttranscriptional gene silencing of parasitic genetic elements (*227, 231, 239*). A viral protein that binds to and represses CRISPR transcription could be a simple way of shutting down (silencing the

silencer) the presumptive RNA mediated immune response in *Sulfolobus*. Mapping the DNA footprint of F93 and other virally encoded DNA binding proteins (particularly those that from homodimers) would certainly be among my highest priorities. This is a fail-safe project. Even if none of these virally encoded DNA binding proteins turn out to be CRISPR repressors, their true targets will be of significant interest.

RNA Mediated Gene Regulation

While writing my departmental mock grant proposal, I developed an interest in RNA mediated gene regulation system and developed a strategy for identifying such a system in Archaea. This concept has recently gained traction among Archaeal virologists; where the race is on to characterize what appears to be an RNAi mediated prokaryotic immune system (above and Chapter 5). The realization that spacer regions in host encoded CRISPRs are homologous to known viral sequences has provided virologists with a tractable system for studying this response (*231-233*). In addition to characterizing host CRISPR and CASS (CRISPR associated system) response to different viral infections (SSV, STIV, ATV and SIRV) we are also interested in the structural and biochemical characterization of the CASS components. In addition to structural and biochemical studies, cas and CRISRP regions should be systematically targeted for disruption. The role of small ncRNA in Archaea may extend beyond CRISPRs,

and the host responses to viral infection. Small ncRNA libraries isolated from a variety of stressed and non-stressed conditions may provide a more comprehensive appreciation for the role of small ncRNA in Archaea.

Remarkably, one organism that has received only marginal attention in this field is the hyperthemophilic Euryarchaeote *P. furiosus*. In 2004, Song *et. al.* published the crystal structure of an Argonaut protein form *P. furiosus* (*248, 249*). Argonaute proteins are the catalytic engine of RISC (RNA Induced Silencing Complex), however despite the obvious implication of the discovery, little has been done to understand the role of this protein *in vivo*.

Viral Evolution and Ecology

Beyond providing a foundation for targeted structural and biochemical studies, the comparative genomic analysis of *Fuselloviruses* has also identified highly conserved regions used to design 'universal' SSV primers (Chapter 3 and 4). Unlike the conserved ribosomal sequences of cellular life, viruses have no universally conserved genetic signature. However within this virus family, conserved regions have been implemented as the 16S equivalent to microbial markers. A viral based molecular phylogenetic approach to SSV and SIRV (SIRV, primers developed by J. Snyder) ecology has revealed an unexpected role for viral migration in shaping viral community structures (*22-24*). Understanding the mechanism(s), frequency and migration rates between distant habitats is

essential to understanding the evolution and ecological significance of these and other widely distributed viruses. Although the stability of viruses from high temperature acidic environments may contribute to their reliable distribution, we do not expect this phenomenon to be restricted to viruses of thermoacidophilic Archaea. Rather, migration should be recognized as a critical component of viral (and perhaps microbial) evolution, with important consequences for understanding the genetic diversity of all life and how viruses contribute to genetic exchange.

Brief on Topics Outside the Lab

Several topics outside the immediate interest of our group have attracted my interest. Recent discoveries in DNA replication, largely contributed by pioneering work done in Steve Bell's lab, have caught my attention (Rev. (*250*)). Using *Sulfolobus* as a model, it has been demonstrated that the Archaeal chromosome is replicated in a manner that resembles a simplified version of that observed in Eukaryotes (*251, 252*). These similarities include many orthologous components in the replication complex (i.e. replicative helicase, sliding clamp and clamp loader) as well as the use of multiple origins of replication; a strategy previously limited to the replication of Eukaryotic chromosomes (*250*). Much less is currently published about the replication strategies of archaeal extra-chromosomal elements (*253*). Historically, our appreciation for the mechanics of

DNA replication have come form understanding viral replication (bacteriophage T4 and T7 have been of particularly importance). Although, replication strategies for archaeal viruses and plasmids remain ill defined, there have been reports of bacterial-like replication components in these parasitic DNA elements ((*145*); Prangishvili; Lawrence personal communication). These reports have been followed by recent discovery of a bacterial-like primase in the *S. solfataricus* genome (*250*). Previously the core subunits of the Eukaryal primase (PriS and PriL) had been identified in the *Sulfolobus* genome, which fit nicely into the archaeal paradigm that asserts that the archaeal information processing system is a simplified/streamlined version of that found in Eukaryotes. Although, precise *in vivo* functional role of each primase is not currently understood, it is possible that they may serve alternative roles associated with DNA repair and replication (*250*). Discovery of this bacterial-like primase adds support to the putative DnaA and DnaB elements in the genomes of extra-chromosomal elements, insomuch as providing another example of bacterial-like components to the archaeo-eukayral DNA replisome ((*145*); Prangishvili personal communication). The replicative polymerases of most bacteriophage and viruses do not utilize a clamp and clamploader, which in Archaea are Eukaryotic-like. However, these polymerases generally require accessory factors that confer processivity. It is tempting to speculate that Archaeal plasmids and viruses will use a similar strategy, bridging an apparent contradiction between the eukaryal like apparatus

for the host and the emerging bacterial like components used in for the viral replication.

Outlook

We currently know very little about the life cycle of any archaeal virus. However the current state of archaeal biology is very dynamic. Accumulating genome sequence information, the availability of archaeal microarrays and the development of facile genetic tools are changing the pace of archaeal biology. Ambitious effort to combine these tools with traditional biochemistry and insights form structural biology are certain to yield substantial insights into the fundamental underpinnings of Archaea and their viruses.

CHAPTER 7

THE MINGLED TOXICITY OF IRON AND OXYGEN

Abstract

The mechanisms responsible for managing oxidative stress are expected to be ancient, having co-evolved with the evolution of an oxygenic atmosphere ~2.8 bya. The antiquity and universal toxicity of oxidative stress has driven the evolution of a spectacularly diverse suite of protein architectures designed to mitigate damage by reactive oxygen species. In the following section I introduce several members of the ferritin-like diiron carboxylate superfamily, designed to manage the toxicity of iron and oxygen. This section is a prelude to the following chapters dedicated to the characterization of a new family of antioxidant proteins, termed DPSL.

Introduction

The evolution of oxygenic photosynthesis in the early Proterozoic Eon was a pivotal innovation to the evolution of life on Earth. The creation of a photosynthetic apparatus capable of splitting water (or bicarbonate) to generate O_2, protons and electrons, forever changed the biogeochemical environment of this planet (*254, 255*). While accumulating levels of O_2 served as the substrate for aerobic metabolism and for the subsequent development of complex life; these metabolic advantages did not arrive without biological consequence.

Oxygen is a reactive molecule, with two unpaired pi orbital electrons (σ_{2s}^2, σ_{2s}^{*2}, σ_{2p}^2, π_{2p}^4, π_{2p}^{*2}). These two pi orbital electrons are spin parallel and thus when oxygen accepts electrons from another molecule it thermodynamically favors a pair of electrons with spins in the complementing orientation; a situation resulting in no net spin. Although this constraint slows the reactivity of oxygen, it kinetically favors the univalent reduction of oxygen; a process that leaves a single unpaired pi orbital electron that is unstable and reactive. The propensity of oxygen to from free radical intermediates makes living with oxygen dangerous. In spite of this toxicity, the energetic benefit of using [$C_6H_{12}O_6 + 6O_2 \rightarrow 6CO_2 + 6H_2O$ (ΔG_0= -2870 kJ/mol)] oxygen as a terminal electron acceptor, has driven the co-evolution of aerobic metabolism and the development of an extensive antioxidant defense mechanism designed to manage the toxicity of partially reduced oxygen species (*256*).

Reactive oxygen species are generated either by energy transfer or by the reduction of O_2. The former leads to the formation of singlet oxygen, while the latter results in the sequential generation of superoxide (O_2^-), hydrogen peroxide (H_2O_2), hydroxyl radical ($HO\cdot$) and ultimately to H_2O. The four electron reduction of O_2 to water, is the energetically preferred end point of oxygen metabolism. However, reactive intermediates are an unavoidable consequence of the aerobic lifestyle. Alternatively, anaerobes forego the energetic benifits of oxygen metabolism; instead using non-oxygen substrates as terminal electors

acceptors, and minimize their exposure to oxygen. Transient O_2 exposure, or abiotic ROS generation (*257*), may account for the frequency of antioxidant proteins coded in the genomes of obligate anaerobes.

Most characterized antioxidant proteins target the degradation of either superoxide or hydrogen peroxide. Hydroxyl radicals are expected to be the most reactive oxidants encountered in biological systems. However, the volatility of this species; indiscriminately reacting with all biomolecules at nearly diffusion limiting rates, leaves little opportunity for their targeted degradation. Rather, the biological impetus appears to focus on the management of longer lived intermediates that have a larger reaction radius and thus greater opportunity for targeted elimination.

Superoxides are primarily managed by either Superoxide dismutase (SOD) (Eq 1) or Superoxide reductase (SOR) (Eq 2).

$2O_2^- + 2H^+ \rightarrow O_2 + H_2O_2$ .....(1)

$O_2^- + 2H^+ + e^- \rightarrow H_2O_2$ .....(2)

The strong anionic charge of superoxide limits its reactivity with electron-rich biomoleucules such as nucleic acids and amino acids. However metal centers within proteins are electrostatically attractive targets and thus highly susceptible to oxidation by superoxide, a fact exploited by antioxidant enzymes designed to facilitate O_2^- degradation. Inappropriately oxidized metal centers are often unstable and degrade, liberating the metal and inactivating the enzyme.

Librated metals, particularly transition metals (Fe, Cu, Co and Mn), compound the problem by reacting with H_2O_2 to generated hydroxyl radicals, via the Fenton reaction (Eq 3) (*258, 259*).

$H_2O_2 + M^{n+} \rightarrow M^{[n+1]} + HO^- + HO\cdot$ .....(3)

The importance of SOD as an antioxidant is evident in SOD-deficient strains of *Escherichia coli*, which exhibited striking growth defects when exposed to air (*260*).

Although critical components of the oxidative stress system, both SOD and SOR reactions generate H_2O_2. Hydrogen peroxide is a moderate oxidant (ε^o=+0.94V), stabilized by the oxygen-oxygen bond (*261, 262*). However, cysteinyl residues are sensitive to oxidation by H_2O_2; creating sulfenic acid adducts that can either form disulfides with other free cysteines or be further oxidized to sulfinic acid moieties (*261*). Perhaps more critically, the univalent reduction of H_2O_2 results in hydroxyl radical formation and thus H_2O_2 management is imperative. The multivalent reaction centers of oligomeric cage-like assemblies common among ferritin-like diiron carboxylate proteins, makes them ideally suited to simultaneously manage the intermingled toxicity of iron and oxygen.

Ferritin-like Diiron Carboxylate Superfamily

Although members of this broad superfamily are expected to have shared a common ancestor they have evolved a wide breadth of distinct functions

including iron regulation, mono-oxygenation, reactive radical production and hydrogen peroxide degradation. Members of this superfamily include bacterioferritin, ferritin, rubrerythrin, aromatic and alkene monooxygenase hydroxylases (AAMH), ribonucleotide reductase R2 (RNRR2), acyl-ACP-desaturases (Acyl_ACP_Desat), manganese catalases, demethoxyubiquinone hydroxylases (DMQH), ubiquinol oxidases (AOX), DNA binding proteins from nutrient starved cells (DPS) and a recently characterize subclass termed DPS-Like proteins (DPSL). In this section we focus on; ferritins, bacterioferritins, DPS, and DPSL protein subclasses. These structurally homologous proteins all oligomerize into cage-like architectures that mediate oxidation-reduction reactions typically involving iron and O_2 or H_2O_2.

Ferritin

Ferritins are spherical protein cages that serve as the primary iron repository in organisms from all three domains of life. Although sequence identity among these proteins varies (~12-98%); secondary, tertiary and quaternary structures are highly conserved. The polypeptide chain of the ferritin subunit (M_w=~20kDa) folds into a conserved antiparallel four helix bundle, with an additional short C-terminal helix capping one end. Twenty four subunits self assemble into a hollow cage like structure having 432 point symmetry. The assembled cage has an exterior diameter of ~120Å, with an internal iron storage compartment of ~80Å that is theoretically capable of banking ~4500 iron atoms. Ferritins isolated

from vertebrates are heteropolymers consisting of Heavy (H) and Light (L) chain subunits, while ferritins isolated from Bacteria, Archaea and plants are homopolymers consisting of only H-chain subunits (*263*). Although H and L chain subunits share only about 50% amino acid sequence identity, the two structures are very similar. The α-carbons of heavy and light chain subunit structures superimpose with an rmsd of 0.5Å (*264*).

Despite considerable structural similarity, the two subunit types have distinct functional roles in the iron sequestration process. H-chain subunits contain a diiron binding site, commonly referred to as the ferroxidase center (FC), that is responsible for the catalyzed oxidation of Fe^{II} to Fe^{III} (Figure 7.1) (*264, 265*). Although the ferroxidase center diiron binding motif is are not present in L-chain subunits, cluster of glutamate residues have been implicated in promoting ferrihydrite nucleation and core formation (*265*).

The iron mineralization reaction in ferritins is a multistep process involving the Fe^{II} sequestration and binding, ferroxidase catalyzed oxidation of iron (FC) (Eq 4), Fe^{III} hydrolysis and finally Fe^{III} migration and nucleation of the mineral core (Eq 5) (*265-267*).

$$2Fe^{II}_{(FC)} + O_2 + 3H_2O \rightarrow [Fe^{III}_2O(OH)_2]_{(FC)} + H_2O_2 + 2H^+ \qquad (4)$$

$$[Fe^{III}_2O(OH)_2] + H_2O \rightarrow 2FeOOH_{(core)} + 2H^+ \qquad (5)$$

The oxidation/hydrolysis reaction is performed at the H-chain ferroxidase center, which is located within a hydrophilic pocket at the center of the four-helix bundle (*264, 268*). Metal ions have direct access to the ferroxidase site via the hydrophilic three fold axis of symmetry (*269*). Metal ions at the ferroxidase center are non-symmetrically coordinated at binding sites referred to as the A and B sites. The A site metal is ligated by Histidine (His65) and a monodentate glutamate (Glu27). The B site is ligated by Glu62 and two monodentate glutamates (Glu61 and Glu107). The two non-equivalent sites are electronically coupled by a bridging glutamate (Glu62) (*5, 270*). Two additional residues Gln141 and Tyr34 are conserved in all H-chain ferritins and are implicated in ferroxidase activity. Ferroxidase center residues have been extensively studied by site directed mutagenesis. Substitution of metal center residues (E27A, Y34F, E62K, H65G, E107A and Q141E) results in the substantial attenuation of ferroxidase activity (*5, 270-273*).

Coordination of binuclear metal ions at the bacterial ferritin ferroxidase center is similar to that observed in mammalian H-chain ferritins; however the B site metal ion is coordinated by one monodentate and one bidentate glutamate, rather than two monodentate glutamates as observed in human H-chain ferritin. A more striking difference is a third metal (C site) situated on the inner surface of the protein shell, ~7 Å away from the *E.coli* ferroxidase center. The C site

metal is coordinated by three monodentate glutamates (Glu49, Glu126 and Glu129), 2 waters and a bridging glutamate (Glu130) that couples B and C site metals (*12*). Although the significance of the C site metal has not been well established, it is likely to be an important feature. In addition to the EcFtnA, a third metal binding site has been implicated in the frog ferritin (FrMF) (*13*) and a third iron ~6 Å from the nearest ferroxidase center iron (B site) has been observed in the tetrahedra ferritin from *Archaeoglobus fulgidus* (*274*).

The ferroxidase metal binding sites in most ferritin crystal structures are only partially occupied and often by metals other than iron (Mn, Tb and Zn). The physiological significance of these non-iron metals - if any - and their potential influence on the arrangement of metal binding ligands has complicated both structural and functional assessment of the ferroxidase center.

Bacterioferritin

Bacterioferritins (BFR) are a phylogenetically distinct subclass of proteins within the ferritin-like diiron carboxylate superfamily. These proteins are structural homologues of the ferritin and bacterial ferritin protein cages described above. The bacterioferritin monomer is a bundle of four long antiparallel α-helices that are capped by a short C-terminal α-helix (*15, 275, 276*). The bacterioferritin fold is superimposable onto that of ferritins and the monomer assembles into a into a 24 subunit spherical cage-like architecture having 432

point symmetry. The bacterioferritin cage has similar dimensions to that of ferritins, with an equivalent iron storage capacity.

What then distinguishes bacterioferritins from ferritins? From a structural standpoint two prominent features discriminate ferritins from bacterioferritins. First, bacterioferritins are heme containing proteins. The bacterioferritin from *E. coli* (*Ec*EFR) coordinates up to 12 protoporphyrin IX heme groups. Each heme is located at the dimer interface (2-fold axis of symmetry) and is coordinated by two symmetry-related methionine residues (Met52) on adjacent subunits (*14*). The precise role of the heme in Bacterioferritin biology has not been well established. However, site directed mutation of the heme coordinating methionine residues, indicates that the heme is not required for the aerobic oxidative uptake of iron. In fact, the iron content of the heme free mutants was ~4 times greater and proceeded 2-3 times faster in BFR (M52H) mutants than for wild-type BFR (*277*).

Like ferritins, the ferroxidase site in H-chain type bacterioferritins is located within a hydrophilic binding pocket at the center of each subunit. Most characterized bacterioferritins are H-chain type homopolymers; however heteromeric bacterioferritins (reminiscent of H and L chain type subunits) have been described (*278-280*). The bacterioferritin from *E. coli* (*Ec*EFR) is a homo H-chain type cage, with a conserved set of iron binding ligands. Bacterioferritin ferroxidase site metals are coordinated by two histidines and four gutamates

(His54, His130, Glu18, Glu51, Glu94 and Glu127; EcBFR numbering) (*15*). His130 and Glu127 are conserved iron ligands that are unique to bacterioferritin and DPSL ferroxidase centers (*17, 281*). In bacterial and mammalian ferritins these residues are substituted by a glutamate and a glutamine that are not directing involved as iron binding ligands. Differences in ferroxidase coordination chemistry may help explain important mechanistic differences between BFR and mammalian ferroxidase centers. Mammalian ferroxidase centers efficiently transfer oxidized iron to the growing mineral core. This is in stark contrast to the stability of ferric iron at the BFR ferroxidase center. Iron transfer from the oxidized mammalian ferroxidase center has been attributed to the swinging mobility of a conserved glutamate (Glu61), which is not found in bacterioferritin ferroxidase centers (Figure 7.1) (*282*). Another distinction between these ferroxidase centers is that BFR requires ferroxidase center participation throughout the process of mineral core formation, whereas in ferritins, core growth continues at the mineral surface even after ferroxidase center fowling by Zn (*282*). These authors have recently proposed a model whereby BFR bound iron is not deposited at the mineral core but rather electrons generated by iron oxidation at the core are used to reduce ferroxidase center iron. Reduced iron atoms at the ferroxidase site may be reserved for the reduction of toxic oxidants (*282*) including O_2 and H_2O_2 (*283, 284*).

DPS (*D*NA Binding *P*roteins from Nutrient *S*tarved Cells)

The prototypic DPS protein (DNA binding Protein under Starved conditions) was isolated from nutrient starved *E. coli* cells (*285*). The crystal structure of this protein (*Ec*DPS) and several DPS structures that have followed, all reveal a variation of the core antiparallel four helix bundle structure described for ferritin and bacterioferritin proteins (*286*). However, instead of the 'capping' C-terminal helix like that found in (bacterio)ferritins, the DPS structure reveals a fifth helix position approximately perpendicular to the long axis of the four helix bundle. This helix is on a long loop connecting opposite ends of the of the B and C helices of the four helix bundle. Rather than the 24 subunit assemblies characteristic of (bacterio)ferritins, DPS proteins assemble into hollow homo-dodecameric cage-like architectures having 32 symmetry. This assembly leads to two non-equivalent environments along a single 3-fold axis. One corresponds to the 3-fold interactions similar to those found in 24 subunit ferritins and involves the N-terminal end of three subunits, whereas the other 3-fold environment consists of residues on the C-terminal end of three subunits.

Another distinctive feature of DPS proteins is the location of the diiron binding motif (*286, 287*). The DPS ferroxidase center is located at the dimer interface, where iron binding ligands are contributed by neighboring subunits (Figure 7.1). Ferrous iron atoms bound at the DPS ferroxidase site are efficiently oxidized by hydrogen peroxide (Eq 6) (*288*).

$$2Fe^{II} + H_2O_2 + 2H_2O \rightarrow 2Fe(O)OH_{(core)} + 4H^+ \quad (6)$$

DPS mediated reduction of H_2O_2 leads to the formation of a mineral core, similar to that observed in ferritins. However the dodecameric assembly of DPS proteins offers a substantially smaller iron storage compartment reaching a max capacity at ~500 iron atoms per cage. Although this may at first, seem limiting, it may not be relevant. *In vivo* DPS proteins are expressed during oxidative or under nutritional stress, including metal ion starvation. The DPS protein is neither ideally designed to store iron (as compared to ferritins) nor expressed under relevant biological conditions (abundant iron). Although DPS proteins are clearly capable of iron mineralization, these observations suggest that DPS proteins are not primarily charged with iron storage but rather principally function as antioxidant proteins.

The antioxidant properties of some DPS proteins is complemented by non-specific association with the chromosome, a characteristic responsible for the DPS acronym (*285*). However, some of the subsequently characterized DPS proteins do not have DNA binding capability. This has led to a growing list of acronyms intended to highlight various attributes of these non-DNA binding dodecameric cages (i.e. NAP, Dpl, Dpr, Flp, ect.).

DPS-Like

We have recently characterized a novel class of antioxidant enzymes that respond to H_2O_2. These enzymes were originally termed DPS-Like, based on predicted structural similarities and shared expression patterns with authentic DPS proteins. However, in spite of these conspicuously DPS-type features, phylogenetic analysis unambiguously recognizes DPSL proteins as a distinct monophyletic cluster within the ferritin-like diiron carboxylate superfamily (*7, 25*). The recently published crystal structure of the prototypic DPSL protein from *Sulfolobus solfataricus* (*Ss*DPSL) reveals several important structural features that distinguish this subclass from others within this superfamily (*17*). With the exception of an additional short N-terminal helix, the overall fold of the DPSL polypeptide chain and the oligomeric state of this protein are indistinguishable from those of authentic DPS proteins. Remarkably however, the metal binding site within this DPS-type architecture is radically different than that in all authentic DPS proteins (Figure 7.1). Rather than the intersubunit metal binding site characteristic of DPS proteins, the DPSL metal binding site is located at the center of the four helix bundle; a position similar to that observed in ferritin and bacterioferritin proteins. Further, metals within each intrasubunit binding site are coordinated by two histidine residues and four acidic residues (His73, His159, Glu37, Asp70, Glu124, and Glu156), bearing strikingly similarity to the metal binding motif conserved in bacterioferritins (Figure 7.1). Clearly the location and

coordination of metals at the *Ss*DPSL binuclear metal binding site offers structural validation to the phylogenetic distinction of DPSL proteins.

What are the functional implications of a bacterioferritin-like metal binding site in the context of a dodecameric DPS-type architecture? *In vitro* hydrogen peroxide more efficiently oxidizes Fe^{II} loaded *Ss*DPSL ferroxidase centers than does oxygen (*7*). This result is similar to what has recently been reported for bacterioferritins (*283, 284*) and contrary to the efficient O_2 mediated oxidation of Fe^{II} loaded ferroxidase centers in ferritins. Thus, regardless of oligomeric context (12 subuints in the DPSL protein and 24 subunits in the bacterioferritin cage) the bacterioferritin-type metal binding site appears to be tuned for oxidation by hydrogen peroxide (*7, 283, 289*). Similarities between DPSL and the bacterioferritin protein structures go beyond the conserved metal binding site within the four helix bundle. Access to metal binding sites in both proteins is through a boot-shaped access channel extending from the outer surface of the four-helix bundle into the central metal binding cavity (*17*). Interestingly, a boot-shaped access channel with nearly identical dimensions to that observed in DPSL proteins is also observed in the heme-containing Catalase-Peroxidase (CPx) enzymes (*290*). The conservation this access channel in structurally distinct and phylogenetically diverse lineages of H_2O_2 degrading enzymes suggests that these channels may be important features in oxidant and/or metal accessibility.

Given the well established role of iron in the generation of ROS, it is not surprising that iron storage proteins like ferritins, are up-regulated in response to iron rich conditions. It is surprising however, that DPS and DPSL proteins do not respond to high iron conditions, but rather are up-regulated in iron depleted conditions. Obviously this expression pattern is not consistent with a primary role for these proteins in iron storage. However, most antioxidant enzymes use iron, manganese, or nickel as cofactors and thus antioxidants capable of recycling their active site metal center would be critical to the antioxidant defense system under metal limiting conditions. A pair of conserved cysteine residues (Cys101 and Cys126) unique to DPSL proteins are located in the boot-shaped access channel, ~9.5Å above the DPSL metal binding site (*17*). Although the precise role of these residues is not currently known, they maybe involved in the redox-recycling of active site metals, by shuttling electrons from reducing agents at the surface to the oxidized metals buried in the four helix bundle.

Intriguingly, the DPSL bacterioferritin-type metal binding site shares significant similarity to the intrasubunit binuclear manganese binding site of Mn catalases (Figure 7.1) (*17*). Given the similar size, charge and electronic properties of iron and manganese it was not surprising to find Mn rather than Fe occupying the dimetal binding site in the *E. coli* bacterioferritin crystal structure (*15*). Recently however, preliminary EPR data suggests that *in vitro* the *Ss*DPSL protein has a significantly greater affinity for manganese than it does for iron

(Douglas; Hendrich unpublished). This is in contrast to the mixed occupation of iron and zinc atoms at the dimetal binding site in the *Ss*DPSL crystal structure. Characterization of the protein directly isolated from *S. solfataricus* should help clarify the physiologically relevant metal(s).

The N-terminal region of DPS proteins has been implicated in mediating the proteins non-specific association with the prokaryotic chromosome. N-terminal DPS deletion mutants or naturally occurring DPS proteins that lack the N-terminal extension do not appear to bind DNA. However, efforts to understand the nature of the N-terminal mediated DNA association have been hampered by the absence of this region in DPS crystal structures. Recently, however Stillman *et al* reported the first observation of an N-terminal α-helix in the crystal structure of a DPS protein from *Lactococcus lactis* (*291*). The N-terminal α-helix is expected to interact with the major groove of the DNA double helix. All currently identified DPSL proteins also possess an N-terminal extension that by analogy to DPS proteins has been implicated in DNA binding. The N-terminal extension of the DPSL protein from *S. solfataricus* also form an α-helix that may mediate the protein association with the *Sulfolobus* chromosome (*17*). The non-specific association of these proteins with the prokaryotic chromosome is reminiscent of Eukaryotic histones. Lysine residues on the N-terminal helix may be posttranslationally modified to regulate DNA association. Experiments intending

to address the possible role of DPSL mediated gene regulation are currently underway.

Summary

DPSL proteins form a distinct subclass of enzymes within the ferritin-like diiron carboxylate superfamily. These proteins combine unique DPSL features with structural characteristics previously restricted to authentic DPS and bacterioferritin proteins. The bacterioferritin-like metal binding site found in DPSL proteins is accessible via a boot-shaped access channel. A similarly shaped channel is observed in bacterioferritins and also in the structures of unrelated catalase-peroxidase enzymes. The *in vitro* antioxidant activity of DPSL proteins is consistent with their expression during oxidative stress. However the precise mechanism of H_2O_2 degradation and the functional importance of unique cysteine pairs have not been well established.

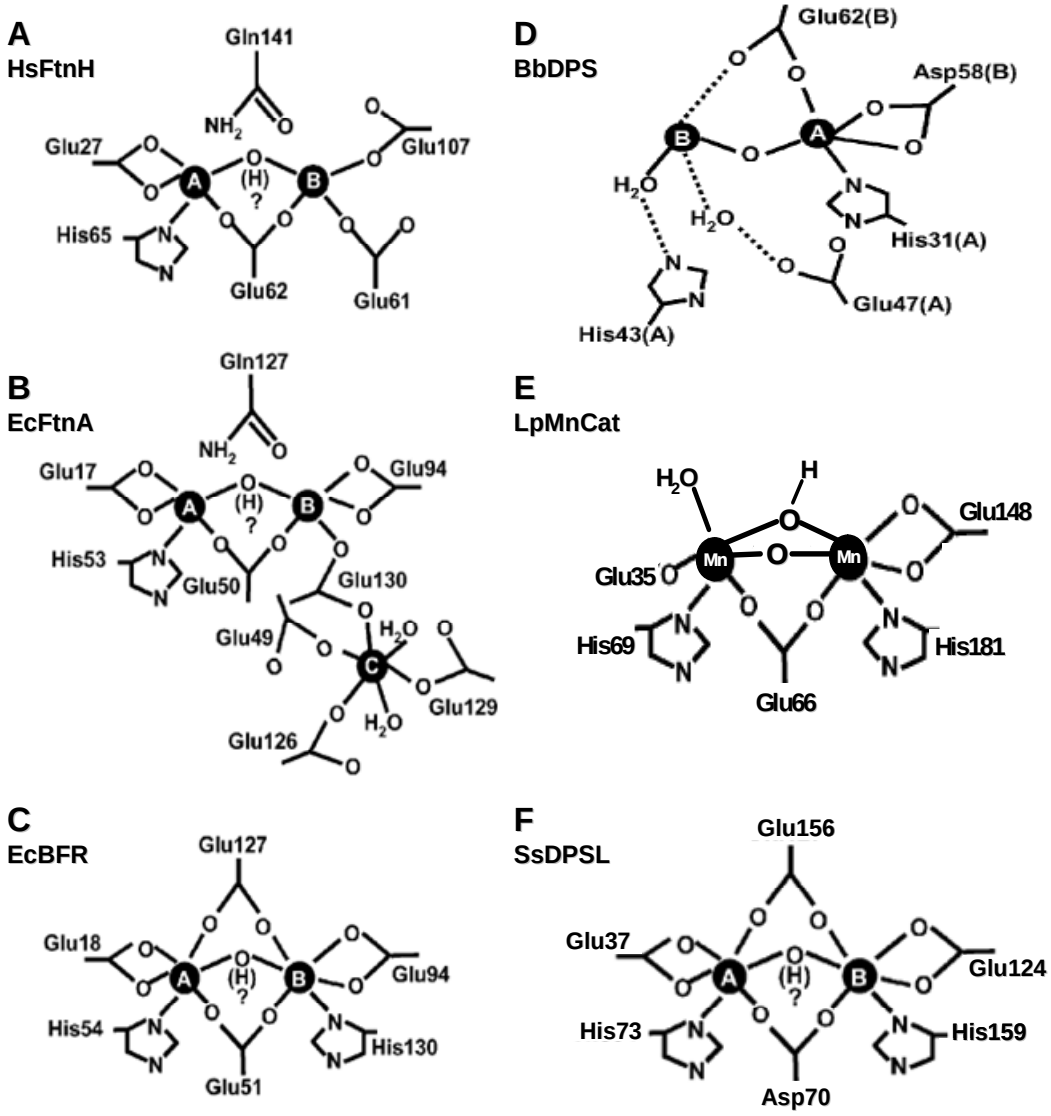


Fig 7.1 A comparison of the coordination environments of binuclear manganese and iron centers in ferritin-like 'diiron' carboxylate proteins. Coordinating metal ligands form active sites in A) Human H chain ferritin (HsFtnH) (pdb; 1FHA) (*5*), B) *E. coli* ferritin (EcFtnA) (pdb; 1EUM) (*12, 13*), C) ferroxidase centre of *E. coli* bacterioferritin (EcBFR) (pdb; 1BFR) (*14, 15*), D) *B*. *Brevis* DPS (pdb; 1N1Q) (BbDPS; intrasubunit iron binding site) E) *Lactobacillus plantarum* Mn catalase (LpMnCat) (pdb; 1JKU) (*16*) and D) S. solfataricus DPSL (SsDPSL) (pdb; 2CLB) (*17*).

CHAPTER 8

AN ARCHAEAL ANTIOXIDANT: CHARACTERIZATION OF A DPS-LIKE PROTEIN FROM *SULFOLOBUS SOLFATARICUS*

Abstract

Evolution of an oxygenic atmosphere required primordial life to accommodate the toxicity associated with reactive oxygen species. We have identified, isolated and characterized an archaeal antioxidant from the thermoacidophilic archeaon *Sulfolobus solfataricus*. The amino acid sequence of this ~22 kDa protein shares little sequence similarity with proteins of known function. However, the protein shares high sequence similarity with hypothetical proteins in other archaeal and bacterial genomes. Nine of these hypothetical proteins form a monophyletic cluster within the broad superfamily of ferritin-like diiron-carboxylate proteins. Higher order structural predictions and image reconstructions indicate that the *S. solfataricus* protein is structurally related to a class of DNA-binding protein from starved cells (DPS). The recombinant expressed *S. solfataricus* DPS-like (*Ss*DPSL) protein self assembles into a hollow dodecameric protein cage having tetrahedral (2 3) symmetry. The outer shell diameter is ~10 nm, and the interior diameter is ~5 nm. DPS proteins have been shown to protect nucleic acids by physically shielding DNA against oxidative damage and by consuming constituents involved in Fenton chemistry. *In vitro*, the assembled archaeal protein efficiently uses H_2O_2 to oxidize Fe(II) to Fe(III) and stores the oxide as a

mineral core on the interior surface of the protein cage. The *ssdpsl* gene is upregulated in *S. solfataricus* cultures grown in iron-depleted media and upon H_2O_2 stress, but is not induced by other stresses. *Ss*DPSL-mediated reduction of hydrogen peroxide and possible DNA-binding capabilities of this archaeal DPS-like protein are mechanisms by which *S. solfataricus* mitigates oxidative damage.

Introduction

The evolution of oxygenic photosynthesis marks the dawn of oxidative stress and represents one of the greatest selective pressures imposed on primordial life. The association of molecular oxygen with abundant ferrous iron pools created two major biological consequences. First, life dependent on the redox properties of Fe(II) would have to contend with its oxidation and the precipitation of Fe(III) as Fe_2O_3. Second, life would have to contend with the toxicity of reactive oxygen species generated by the partial reduction of dioxygen by ferrous iron. Iron metabolism and oxidative stress are thus intimately interwoven, with the toxicity of one being dependent on the other. The significance of these selective pressures remains evident in the genomes of modern organisms where complex and overlapping defense mechanisms protect against oxygen toxicity and iron stress (*292-294*).

Oxidative stress is a universal phenomenon experienced by both aerobic and anaerobic organisms from all three domains of life (*261, 295*). Although the

significance of oxidative stress in biology is well established, the processes by which intracellular reactive oxygen species (ROS) are generated continue to be recognized. One pathway generally recognized as a significant source of ROS is the partial reduction of molecular oxygen by the autoxidation of flavoproteins (*296, 297*). This process generates a mixture of superoxide (O_2-) and hydrogen peroxide (H_2O_2). The strong anionic charge of superoxide limits its reactivity with electron-rich molecules such as nucleic acids and amino acids (*261, 297*). However, metal clusters within proteins are highly susceptible to superoxide. For example, Fe(II) ions at the catalytic center of 4Fe-4S containing enzymes are electrostatically attractive targets (*298-300*). Oxidized clusters are unstable and degrade, resulting in the release of iron and the inactivation of the enzyme. The released iron is available to react with hydrogen peroxides through the well-characterized Fenton reaction (described below) to produce hydroxyl radicals ($HO^{\cdot}$), the most toxic of all ROS (*259*). Hydroxyl radicals indiscriminately oxidize most biomolecules at nearly diffusion-limiting rates.

The identification of superoxide dismutase (SOD) provided the first indication of active mechanisms designed to specifically minimize the effects of oxidative stress (*301*). This enzyme catalyses the disproptionation of superoxide (Eq. 1).

$$2O_2\text{-} + 2H^+ \rightarrow H_2O_2 + O_2 \qquad (1)$$

SOD genes are ubiquitous in the genomes of aerobic organisms and homologous genes are found in the genomes of some anaerobes. The apparent absence of an SOD gene in the genomes of some anaerobes maybe a reflection of this enzyme's generation of toxic oxygen (*302*). A second mechanism identified in anaerobes has been described for superoxide detoxification (*295, 302*). This mechanism utilizes a mononuclear iron-containing enzyme, SOR (super oxide reductase), for the efficient reduction of super-oxide to H_2O_2 without an O_2 byproduct (Eq. 2).

$$O_2\text{-} + 2H^+ + e^- \rightarrow H_2O_2 \qquad (2)$$

H_2O_2 is a powerful oxidant produced by both SOR and SOD reactions. Potential targets of H_2O_2 oxidation include 4Fe-4S clusters and the sulfur atoms of cysteine and methionine residues with potentially lethal consequences (*261*). More importantly, the interaction of free Fe(II) with hydrogen peroxide is known to efficiently generate HO· through the Fenton reaction (Eq. 3) (*258, 259*).

$$H_2O_2 + Fe^{2+} \rightarrow OH^- + FeO^{2+} + H^+ \rightarrow Fe^{3+} + OH^- + HO\cdot \qquad (3)$$

Peroxide detoxification and iron metabolism are important components in mitigating oxidative damage. Other enzymes, such as catalase, affecting

peroxide detoxification have been characterized. Although these enzymes efficiently remove H_2O_2 they do little to minimize the toxicity of free iron. A DNA binding protein from nutrient starved *E. coli* cells (DPS) which protects nucleic acids against oxidative damage has been described (*285*). Twelve copies of the DPS protein self-assemble into a ~10 nm spherical cage-like structure. These structures are homologous to the multimeric assemblies formed by the iron mineralizing family of ferritin proteins. In addition to their structural similarities, assembled DPS proteins also mineralize iron within the interior of the protein cage. However, DPS mineralization is distinct from that described for ferritins. While ferritins prefer O_2 as an oxidant producing H_2O_2, DPS utilizes H_2O_2 as an oxidant according to Eq. 4 (*288, 303*). However, it has been shown that bacterioferritins exibit a similar preference for H_2O_2 over O_2 (*282*).

$$2Fe^{2+} + H_2O_2 + 2H_2O \rightarrow 2Fe(O)OH_{(s)} + 4H^+ \qquad (4)$$

In this way DPS simultaneously eliminates the two components of the Fenton reaction [Fe(II) and H_2O_2] that contribute to the generation of hydroxyl radicals. This mineralization reaction, in combination with the *in vivo* association of DPS with DNA, represents a multifunctional approach to cellular protection against oxidative damage.

We are interested in evolutionarily conserved mechanisms by which iron dependant life manages the paradoxical relationship between iron and oxygen. Hyperthermophilic archaeon are deeply rooted in the tree of life and therefore may harbor ancient mechanisms for managing oxidative stress**.** *Sulfolobus solfataricus* is an aerobic thermophilic (70-90°C) archaeon that thrives in acidic terrestrial thermal features that are commonly associated with high iron (*78*). We have identified a previously unrecognized protein from *S. solfataricus* that functions as a DPS. Here we provide evidence supporting this protein's role in protection against oxidative damage.

Materials and Methods

Identification of *Ss*DPSL

The *S. solfataricus* (P2) genome was queried, using the BLAST algorithm, with ferritin and DPS-like proteins from several sources (i.e. human, horse, bullfrog, *Escherichia coli, Listeria innocua, Archeaoglobus fulgidus* and an uncultured crenarchaeote (4B7) (*181*). A putative ferritin from the uncultured crenarchaeote 4B7 was the only sequence sharing detectable similarity to any annotated protein in the *S. solfataricus* (P2) genome (*181*). The identified protein, annotated as a hypothetical protein (gi:15898865), was subjected to a threading routine and submitted to the fold recognition server 3D-pssm (http://www.sbg.bio.ic.ac.uk/servers/3dpssm/) (*304*).

Phylogenetics

Amino acid sequences of ferritin-like diiron carboxylate proteins were aligned with CLUSTALX (*186*). Parsimony analysis was performed with test version 4.0b10 of PAUP* and bootstrapped with 10,000 resamplings (*187*).

Culturing of *S. solfataricus*

Liquid cultures of *S. solfataricus* (P2), were grown aerobically in a revised version of an established salt based medium; 22.70 mM $(NH_4)_2SO_4$, 2.19 mM $K_2HPO_4 \cdot 3H_2O$, 13.42 mM KCl, 9.33 mM Glycine, 4.55 µM $MnCl_2 \cdot 4H_2O$, 9.44 µM $Na_2B_4O_7 \cdot 2H_2O$, 382.56 nM $ZnSO_4 \cdot 7H_2O$, 146.64 nM $CuCl_2 \cdot 2H_2O$, 61.99 nM $Na_2MoO_4 \cdot 2H_2O$, 59.29 nM $VOSO_4 \cdot 5H_2O$, 17.79 nM $CoSO_4 \cdot 7H_2O$, 19.02 nM $NiSO_4 \cdot 6H_2O$, supplemented with 0.2% tryptone (*159*). Iron depleted versions of this media were prepared using an iron specific chelating column (Nielsen *et. al.* unpublished). Iron depleted media was used directly or supplemented with $FeSO_4$ or $FeCl_3$. All cultures were grown in long neck Erlenmeyer flasks at 78°C in shaking oil bath incubators.

Western Analysis

Polyclonal antibodies were raised in mice against purified recombinant *Ss*DPSL protein. IgG antibodies were isolated using the ImmunoPure purification kit according to the manufacture's protocol (Pierce, Rockford, IL). Western blots were performed using the purified IgG. *S. solfataricus* cells were harvested by centrifugation and resuspended in 2X SDS gel-loading buffer (*305*). Lanes were

loaded with approximately equal cell numbers according to OD_{650}. *S. solfataricus* proteins were separated on 15% SDS-polyacrylamide gels and transferred to Hybond-ECL nitrocellulose membrane (Amersham Bioscience, Piscataway, NJ). Colorometric detection was used to identify the *Ss*DPSL protein (Bio-Rad, Hercules CA). Protein and RNA concentrations were estimated using AlfaImager imaging system (IS-2200) according to average pixel density.

Northern Analysis

Total cellular RNA was extracted from *S. solfataricus* (P2) cells, according to the TRI REAGENT protocol (Molecular Research Center, Inc. Cincinnati OH). Total RNA concentrations were estimated by agarose gel electrophoresis and by spectrophotometric measurements (OD_{260}). For northern analysis, 0.5-2 µg of total RNA was separated by electrophoresis in a 1.5% agarose (wt/vol) formaldehyde gel and transferred to GeneScreen membranes as recommended by the manufacturer (NEN Life Science Products, Boston, MA). RNA was membrane-crosslinked in a UV Stratalinker (Stratagene, San Diego, CA). Blots were probed with ^{32}P labeled, polymerase chain reaction (PCR) products (Ready-To-Go Labeling Kit, Amersham Bioscience). *S. solfataricus* DNA was isolated using previously established methods and served as the template in PCR reactions (*173*).

Cloning and Expression

PCR primers were designed to amplify the dps-like gene from *S. solfataricus* (P2) (gi:15898865). The forward primer (5' gggtgta**catatg**caagagaaacccc 3') included an *Nde*I restriction endonuclease site (in bold) directly upstream of the start codon (underlined). The reverse primer (5'ac**ggatcc**ttatttcttggaatatggagcg 3') included the stop codon for the gene (underlined) and a *Bam*HI site (in bold). The resulting PCR product was digested with *Nde*I and *Bam*HI restriction endonucleases, purified with a PCR purification kit (Qiagen Inc., Valencia, CA), ligated into pET-30a(+) (Novagen, Madison, WI) and transformed into Xl-2 blue *Escherichia coli* (Stratagene). The cloned sequence was confirmed by DNA sequencing (Applied Bio Systems, Foster City, CA). The plasmid was subsequently transformed into BL21 *E. coli* for protein expression (Novagen). For protein expression, *E. coli* were grown and induced according to the manufactures protocol (Novagen). Cells were harvested by centrifugation and screened for *Ss*DPSL expression by SDS-PAGE.

*Ss*DPSL Purification from *E.coli*

Cells were pelleted from one-liter cultures by centrification, resuspended in 30 mls of lysis buffer (50 mM MES, 100 mM NaCl, pH 6.5, 0.002 mg/ml DNAse, 0.05 mg/ml RNAse, 1 mg/ml lysozyme) and incubated for 30 minutes at room temperature. The slurry was sonicated (3 x 5 min.) and cellular debris was removed by centrifugation. The resulting supernatant was heated at 65°C for 10

minutes, cooled on ice and then centrifuged to remove denatured proteins. The recovered supernatant was passed through a 0.2 μm filter and loaded onto a Superose 6 size exclusion column (Amersham Biosciences) equilibrated with MES buffer (50 mM MES, 100 mM NaCl, pH 6.5). Elution of the protein was monitored at 260, 280 and 410 nm. Protein concentration was determined by the biuret method, and confirmed by the molar adsorbtivity at 280 nm of 3.00 x 10^4 $M^{-1} \cdot cm^{-1}$.

*Ss*DPSL Fe_2O_3 Mineralization

0.2 mg of *Ss*DPSL (7.66 x 10^{-7} mmol) was diluted with 3 mL of MES buffer (100 mM MES, 100 mM NaCl, pH 6.5) in a standard quartz cuvette. A total of 2.25 x 10^{-4} mmol of de-aerated Fe(II) was added in 6 increments (50 Fe/cage) at 10 minute intervals. A half molar equivalent of hydrogen peroxide (1 H_2O_2:2 Fe^{2+}) was delivered two minutes after each iron addition. A UV-Vis spectrum was taken every two seconds to monitor core formation by the increase in metal-to-ligand charge transfer band (350-400 nm).

*Ss*DPSL Particle Characterization

Dynamic light scattering was performed on the assembled protein before and after mineralization to determine particle size as previously described (*306*). Particle size was independently estimated by mobility on size exclusion columns (Superose 6, Amersham Biosciences) and by transmission electron microscopy (TEM, Leo 912AB operating at 120keV). Samples for TEM were imaged both

unstained and stained with 2% uranyl acetate. Mass of the *Ss*DPSL protein was determined by electrospray mass spectrometry (LC-MS).

Image Reconstruction

Aliquots of unmineralized *Ss*DPSL (~3 µl) were stained for 30 sec with 2% uranyl acetate. Images were recorded using a CM120 electron microscope (FEI/Philips) at a magnification of 60,000 ±1% and focal pairs with a range of ~0.5-2.5 µm underfocus. Negatives were digitized on a Zeiss SCAI flat-bed scanning densitometer (ZI/Zeiss) with a step size of 7 µm, which resulted in a pixel size of 1.17 Å on the object scale.

Image processing was performed with polar Fourier transform (PFT) methods, with a modification for tetrahedral symmetry by D. Belnap (*307*). About 2,230 particle pairs were manually extracted using the program X3D and the images were scaled to the mean and SD for all images (*308*). The contrast transfer function (CTF) parameters for each micrograph were determined from the computed Fourier transform of the carbon film of each micrograph, and phase corrections were then applied to each particle image. A 30 Å starting model was calculated from the x-ray structure of the DPS isolated from *Listeria innocua* (PDB code 1QGH) for the reference-based alignment. The x,y origin and rotational orientation of the particles were iteratively refined. For each cycle, particles that deviated by 0.5 from the average were rejected. In general, about one-quarter of the particles were rejected. Two independent reconstructions

were computed to estimate the resolution of the refined data by Fourier shell correlation and subsequently combined. Using a correlation coefficient cut-off value of 0.5, the resolution was estimated to be 18 Å. The final 3D map was derived from 1,402 image pairs.

Results

*Ss*DPSL Identification

Using the default parameters of BLAST, we were unable to identify any amino acid sequence in the *S. solfataricus* (P2) genome related to characterized ferritin or DPS proteins. However, a BLAST search using a putative ferritin identified in an uncultured crenarchaeote (AAK66802) matched a *S. solfataricus* (P2) sequence (gi:15898865) with 55% identity. This *S. solfataricus* (P2) open reading frame (SSO2079), annotated as a conserved hypothetical protein, codes for a predicted 188 amino acids with an estimated MW of 21,753Da. This sequence threads onto the four-helix bundle structure of the DPS protein from *Listeria innocua*, with high confidence (PSSM E=0.0171) (*287, 303, 309*). Based on this predicted secondary and tertiary structural similarity, the *S. solfataricus* (P2) protein was expected to assemble into a spherical dodecameric cage similar to DPS family proteins (*286*).

In vivo Expression Patterns of *ssdpsl*

Northern blot analysis of RNA isolated from exponential growth phase culture of *S. solfataricus* (P2) identifies a single transcript corresponding to the approximate size of the *ss*dpsl gene (~600 nt). The transcript is up regulated in a non-linear fashion in response to oxidative stress (Figure 8.1A), with a dramatic increase observed at 30 μM H_2O_2. Western blots indicate a more linear accumulation of the *Ss*DPSL protein in response to increasing concentrations of H_2O_2 (Figure 8.1B). Loading controls and densitometry are available as supporting information in the appendix (Figure S8.8).

Taking advantage of this oxidative induction, we were able to isolate approximately 0.3 mg of the assembled protein from a 1-liter culture of *S. solfataricus* (P2) grown in media supplemented with 30 μM H_2O_2. No detectable amounts of this protein could be purified from *S. solfataricus* (P2) culture grown under standard conditions following this same protocol. These results confirmed our prior observations that neither the transcript nor the protein is readily detectable under normal growth conditions (Figure 8.1).

We further examined the possibility of *Ss*DPSL being involved in a general stress response rather then being exclusively related to oxidative stress (Figure 8.2 and in the appendix as Figure S8.9). Transcription of the *ss*dpsl gene is low or not detectable in late log phase cells or in exponential growth phase cells UV irradiated (300KJ), grown in an alternative carbon source (sucrose), virus

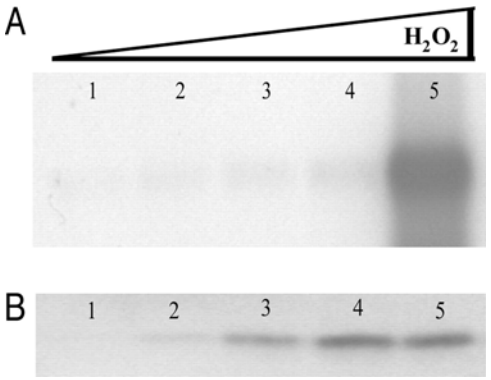


Fig. 8.1 *Ss*DPSL expression in response to oxidative stress. (A) Northern blot analysis of RNA extracted from exponential growth phase (OD_{650}~0.5) *S. solfataricus* cells. Lane 1, standard growth conditions (0 µM H_2O_2); lane 2, 5 µM H_2O_2; lane 3, 10 µM H_2O_2; lane 4, 20 µM H_2O_2; lane 5, 30 µM H_2O_2. Approximately 1.5 µg of total RNA was loaded in each lane. (B) Western blot analysis of total protein extract from logarithmic (OD_{650}~0.5) *S. solfataricus* cells, loads are normalized according to number of cells and lane assignments are as above.

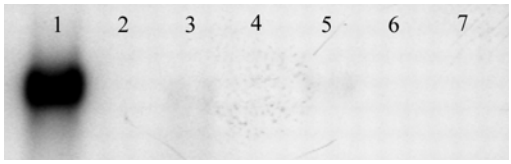


Fig. 8.2 *Ss*Dps expression is specific to oxidative stress. Northern blot analysis of RNA extracted from *S. solfataricus* cultures grown under standard conditions (lane 1), 30 µM H_2O_2 (lane 2), late log cells (OD_{650}~1.2) (lane 3), 300 KJ UV, which accounts for RNA degradation as shown in the load control (lane 4), Sucrose as the sole carbon source (lane 5), SSV RH virus infected (lane 6), 11hrs. heat shock at 90°C (lane 7), 11hrs. cold shock at 60°C. Approximately 0.75 µg of total RNA was loaded in each lane.

Fig. 8.3 *Ss*DPSL expression in response to iron. Western blot analysis of total protein extract from exponential growth phase (OD_{650}~0.5) *S. solfataricus* cells. Loads were normalized according to number of cells. Lane 1, standard growth conditions; lane 2, iron extracted media; lane 3, iron extracted media supplemented with 1.25 µM Fe_2SO_4; lane 4, iron extracted media supplemented with 1.25 mM Fe_2SO_4; lane 5, iron extracted media supplemented with 5.0 mM Fe_2SO_4; lane 6, iron extracted media supplemented with 5.0 mM Fe_2Cl_3; lane 7, molecular weight ladder; lane 8, 2 ng purified *Ss*DPSL.

infected (SSVRH), heat shocked at 90°C or cold shocked at 60°C for eleven hours. However, consistent with the characterized expression profiles of other DPS proteins, *Ss*DPSL is expressed under iron-limiting growth conditions (Figure 8.3) (*292, 293, 310*). Dps proteins use ferrous iron atoms as cofactors for the pair-wise reduction of H_2O_2. It follows then, that in the absence of iron, DPSL mediated reduction of hydrogen peroxide will be compromised, resulting in increased H_2O_2 stress.

Particle Purification and Characterization

The *Ss*DPSL protein expresses to high levels in *E. coli*. Approximately 40 mg of the recombinant protein can be purified from a 1-liter *E. coli* culture. This compares to 0.3 mg of the protein that could be purified from 1-liter cultures of *S. solfataricus* stressed with H_2O_2. Purification of the recombinant *Ss*DPSL protein was simplified by taking advantage of the protein's inherent thermal stability. A ten-minute incubation at 65°C was sufficient to denature a majority of the *E. coli* proteins, which were precipitated by a low speed centrifugation. The recombinant protein eluted as a single peak on an analytical size exclusion column (Figure 8.4). SDS PAGE of the purified protein revealed a single band with Mw of ~22 kDa (available on line as supporting information Figure 10S). This was in good agreement with the deconvoluted mass of the recombinant protein (21,768.23 Da), as determined by electrospray mass spectrometry.

The recombinant *Ss*DPSL protein self-assembles into a 12 subunit cage-like architecture. Dynamic light scattering (DLS) of the recombinant protein indicates an average diameter of 9.7 nm $\pm$ 0.4 nm, consistent with a 12 subunit cage-like assembly. This architecture is directly evident from negatively stained transmission electron micrographs (TEM) (Figure 8.4). *Ss*DPSL image reconstruction, at 18 Å resolution, reveals a dodecameric protein cage with an exterior diameter of ~100 Å and an interior cavity of ~50 Å (Figure 8.5). As a tetrahedron the structure is composed of a trimer of dimers that assembles to form two nonequivalent three-fold (3F) environments at opposing ends of the 3F axis (Figure 8.5, B and C).

*Ss*DPSL Mineralization

At pH 6.5 the recombinant *Ss*DPSL protein efficiently catalyzes the oxidation of Fe(II) to Fe(III) in the presence of H_2O_2 (Figure 8.6). Like authentic DPS proteins, *Ss*DPSL catalyzed oxidation of Fe is dependent on H_2O_2 as the oxidant. Substitution of O_2 for H_2O_2 results in significantly retarded rates of Fe oxidation. This is in contrast to Fe(II) oxidation by all known ferritins, which efficiently catalyzes the oxidation of Fe(II) with O_2. Fe oxidation reactions were monitored by the increase in O$\rightarrow$ Fe(III) by absorbance at 350 nm. Fe(II) was added to the protein in aliquots of 50 Fe atom per protein and oxidized in the presence of H_2O_2. This iron loading process (50 Fe:25 H_2O_2) could be repeated 6 times,

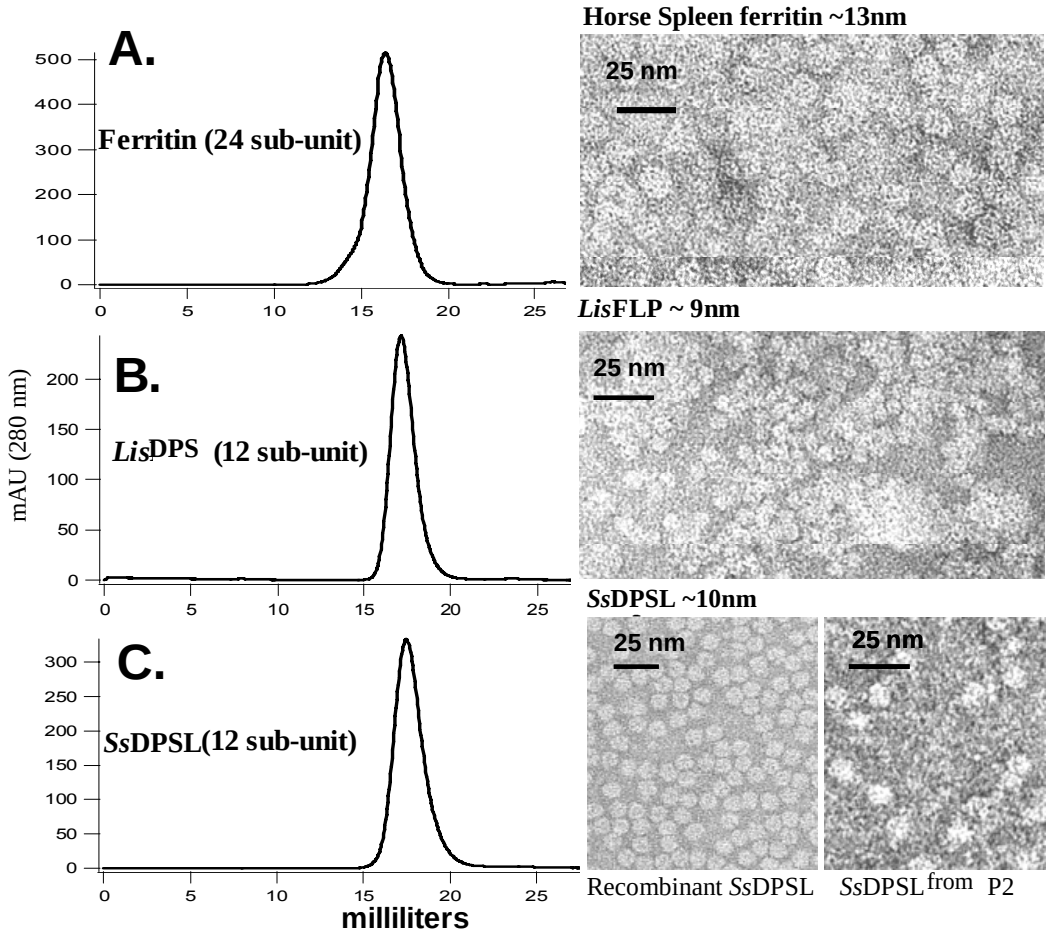


Fig. 8.4 Size exclusion liquid chromatography chromatograms and TEM images of corresponding peaks. Retention times according to size exclusion liquid chromatography are shown for (A) the 24 subunit horse spleen ferritin (B) the 12 subunit ferritin-like protein from *Listeria innocua* and (C) the 12 subunit Dps-like protein from *S. solfataricus*. The *L. innocua* FLP and the Dps-like protein from *S. solfataricus* have retention time consistent with a 12 subunit, 260 kDa protein, where as the 24 subunitferritin from horse spleen elutes earlier. Transmission electron microscopy of each peak reveals intact cage diameters of (A) ~13 nm, (B) ~9 nm and (C) ~10 nm, commiserate with each proteins retention time.

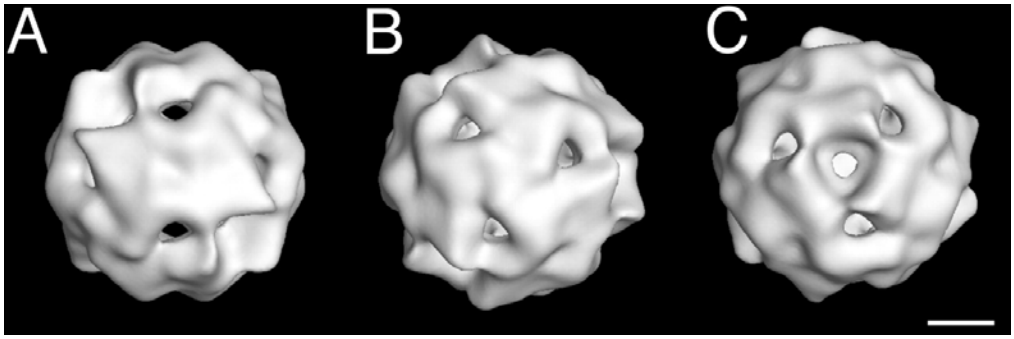


Fig. 8.5 3D image reconstruction of the assembled *Ss*DPSL cage. Surface-shaded views of reconstructed negative-stained images displayed along the twofold (2F) axis (*A*), and along the two nonequivalent environments at each end of the threefold (3F) axis (*B* and *C*). (Scale bar: 2.5 nm).

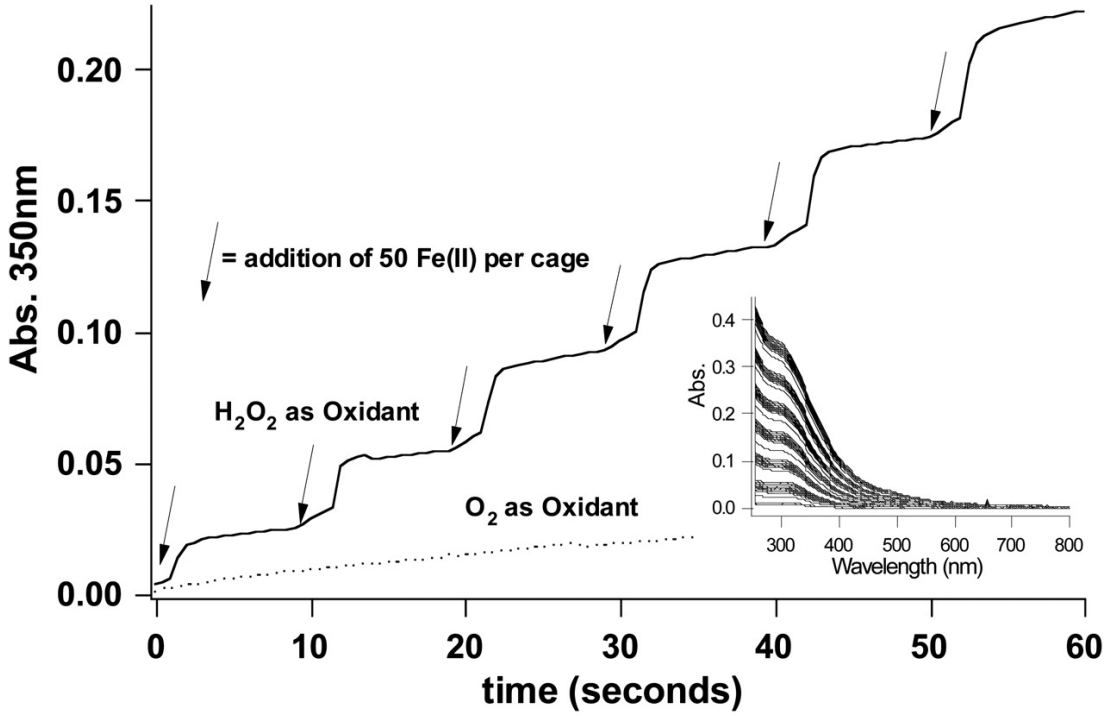


Fig. 8.6 *Ss*Dps-catalyzed mineralization of iron. *Ss*DPSL efficiently uses H_2O_2 to oxidize iron in a stepwise progression ($1H_2O_2$:2Fe) (solid line). In contrast, O_2 serves as a relatively poor oxidant of iron in this reaction (dotted line). (*Inset*) Iron confined within the *Ss*DPSL cage remains soluble as shown in the full-spectrum.

resulting in a theoretical load of 300 iron atoms per cage. Iron mineralized (Fe_2O_3) preparations of *Ss*DPSL exhibited no change in hydrodymanic the protein in aliquots of 50 Fe atom per protein and oxidized in the presence of H_2O_2. This iron loading process (50 Fe:25 H_2O_2) could be repeated 6 times, resulting in a theoretical load of 300 iron atoms per cage. Iron mineralized (Fe_2O_3) preparations of *Ss*DPSL exhibited no change in hydrodymanic diameter before and after mineralization reaction. Electron dense cores commensurate with the ~5 nm interior cavity could be visualized by TEM. This is indicative of a spatially controlled reaction in which Fe_2O_3 particles are formed within the confines of the protein cage, analogous to ferritin and DPS biomineralization.

Phylogenetic Identification of a New Subclass of Antioxidant Proteins

The *Ss*DPSL protein shares high sequence similarity with hypothetical proteins in fourteen prokaryotic genomes (Expect $\geq$ 7e-04). Nine of these sequences form a well-supported monophyletic cluster outside of the currently characterized DPS clade (Figure 8.7). These sequences are only distantly related to those in the currently characterized DPS clade. In spite of this distant relationship, a putative DPS-like metal binding motif (diiron-binding site) is identifiable in the primary structures of each of these new DPS-like sequences (sequence alignments available as supporting information in the appendix, Figure S8.11).

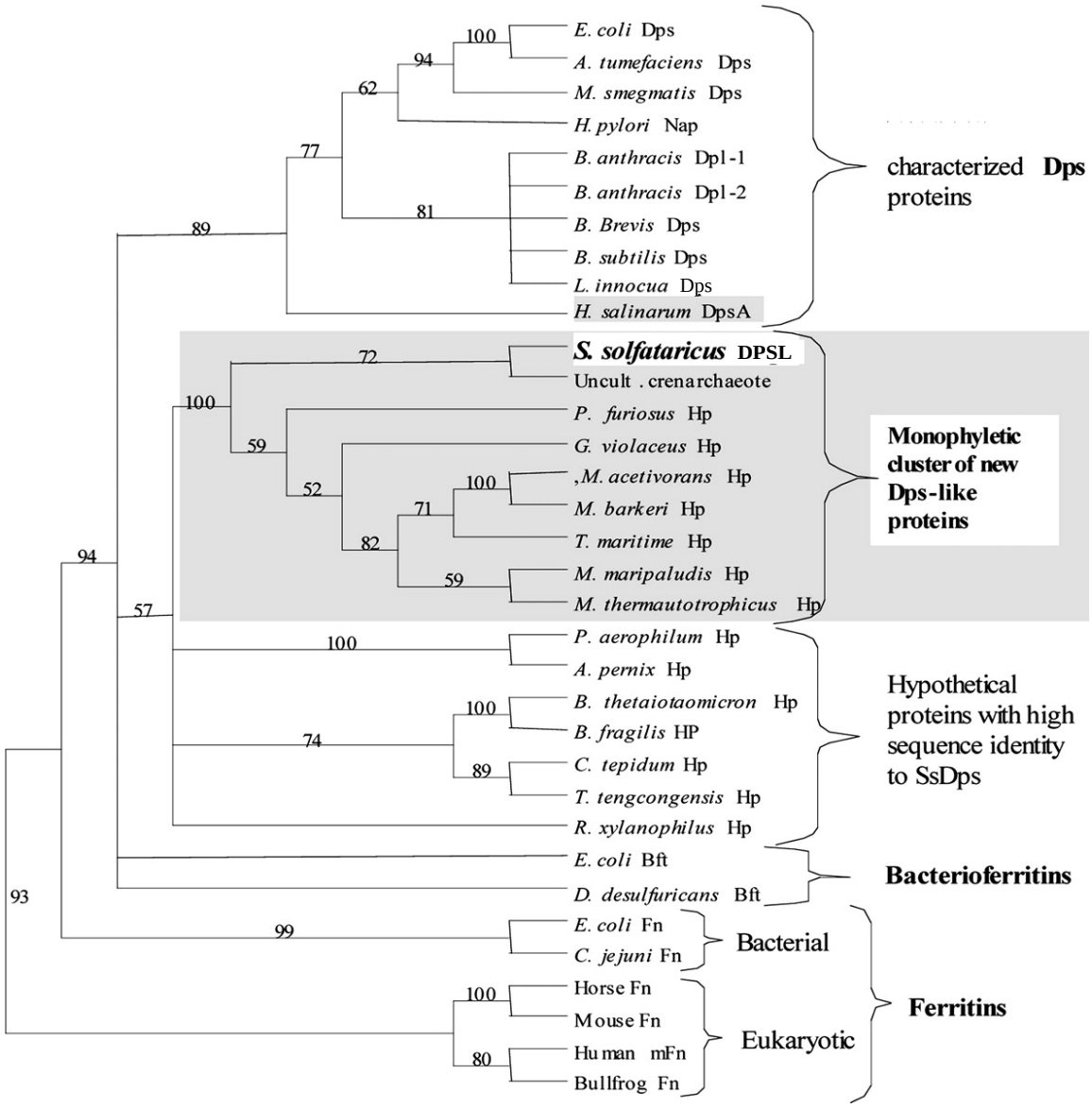

E. coli Dps
A. tumefaciens Dps
M. smegmatis Dps
H. pylori Nap
B. anthracis Dpl-1
B. anthracis Dpl-2
B. Brevis Dps
B. subtilis Dps
L. innocua Dps
H. salinarum DpsA
characterized Dps proteins
S. solfataricus DPSL
Uncult . crenarchaeote
P. furiosus Hp
G. violaceus Hp
,M. acetivorans Hp
M. barkeri Hp
T. maritime Hp
M. maripaludis Hp
M. thermautotrophicus Hp
Monophyletic cluster of new Dps-like proteins
P. aerophilum Hp
A. pernix Hp
B. thetaiotaomicron Hp
B. fragilis HP
C. tepidum Hp
T. tengcongensis Hp
R. xylanophilus Hp
Hypothetical proteins with high sequence identity to SsDps
E. coli Bft
D. desulfuricans Bft
Bacterioferritins
E. coli Fn
C. jejuni Fn
Bacterial
Horse Fn
Mouse Fn
Human mFn
Bullfrog Fn
Eukaryotic
Ferritins

Fig. 8.7 Phylogenetic analysis of the ferritin-like diiron-carboxylate superfamily. The DPSL protein from *S. solfataricus* is more closely related to a group of hypothetical proteins than it is to other characterized members of the ferritin-like diiron-carboxylate superfamily. *H. salinarum* DpsA (shaded) and *S. solfataricus* Dps are currently the only examples of archaeal Dps proteins; all other characterized Dps proteins are bacterial. Database gene identification numbers are as follows: *E. coli* Dps (*Escherichia coli*) gi:16128780, *A. tumefaciens* Dps (*Agrobacterium tumefaciens*) gi:15889746, *M. smegmatis* Dps (*Mycobacterium smegmatis*) gi:17887432, *H. pylori* Nap (*Helicobacter pylori*) gi:15611298, *B. anthracis* Dpl-1 (*Bacillus anthracis*) gi: 21730371, *B. anthracis* Dpl-2 (*Bacillus anthracis*) gi: 21730378, *B. Brevis* Dps (*Bacillus brevis*) gi:31615600, *B. subtilis* Dps (*Bacillus subtilis*) gi:16080351, *L. innocua* Dps (*Listeria innocua*) gi:16800011, *H. salinarum* DpsA (*Halobacterium salinarum* sp. NRC-1) gi:15791220, *S. solfataricus* Dps (*Sulfolobus solfataricus*) gi:15898865, Uncult. crenarchaeote (uncultured crenarchaeote non-heme iron-containing protein) gi:14548145, *P. furiosus* Hp (*Pyrococcus furiosus*) gi:18977565, *G. violaceus* Hp (*Gloeobacter violaceus*) gi:37523861, *M. acetivorans* Hp (*Methanosarcina acetivorans*) gi:20091704, *M. barkeri* Hp (*Methanosarcina barkeri*) gi:48837814, *T. maritima* Hp (*Thermotoga maritima*) gi: 15644274, *M. maripaludis* Hp (*Methanococcus maripaludis*) gi:45358735, *M. thermautotrophicus* Hp (*Methanothermobacter thermautotrophicus*) gi:7482214, *P. aerophilum* Hp (*Pyrobaculum aerophilum*) gi:18313528, *A. pernix* Hp (*Aeropyrum pernix*) gi:14601419, *B. thetaiotaomicron* Hp (*Bacteroides thetaiotaomicron*) gi:29349231, *B. fragilis* HP (*Bacteroides fragilis*) gi:53714735, *C. tepidum* Hp (*Chlorobium tepidum*) gi:21674150, *T. tengcongensis* Hp (*Thermoanaerobacter tengcongensis*) gi:20808605, *R. xylanophilus* Hp (*Rubrobacter xylanophilus*) gi:46106146, *E. coli* Bft (*Escherichia coli*) gi:16131215, *D. desulfuricans* Bft (*Desulfovibrio desulfuricans*) gi:14326006, *E. coli* Fn (*Escherichia coli*) gi:16129855, *C. jejuni* Fn (*Campylobacter jejuni*) gi:15791972, Horse Fn (*Equus caballus* L-chain ferritin) gi:406209, Mouse Fn (*Mus musculus* L-chain ferritin) gi:55154579, Human mFn (*Homo sapiens* mitochondrial ferritin) gi:29126241, and Bullfrog Fn (*Rana catesbeiana*) gi:85895. Protein abbreviations are as follows: Hp, hypothetical protein; Dps, DNA-binding protein from nutrient-starved cells; Bft, bacterioferritin; Fn, ferritin; mFn, mitochondrial ferritin; Dpl, Dps-like protein; and Nap, neutrophil-activating protein. Numbers at branching nodes are bootsrap values from 10,000 resamplings.

While this putative metal binding motif is distinct from that identified in currently recognized DPS proteins, it does share similar charge and spacing characteristics to the known DPS metal binding sites.

DPSL Surface Charge and DNA Binding Domains.

Surface charge determined by Zeta potential measurements of the *Ss*DPSL assembly indicate that the cage retains a net negative surface charge at pH values above 4.1. This overall net negative surface charge is consistent with all other characterized DPS proteins irrespective of their DNA binding character. Although the DNA binding domain in DPS proteins are not well defined, basic residues in either the N-terminal or C-terminal regions have been implicated. For example, extension of the N-terminal *E. coli* DPS subunit or in the C-terminal extension of the *Mycobacterium smegmatis* DPS subunit have been implicated in DNA binding (*286, 311, 312*). These suggestions have recently been corroborated by N-terminal deletion mutants of *E. coli* DPS that do not bind DNA (*313*). Multiple sequence alignments identify basic residues in extended N-terminal domains of both the *Ss*DPSL and in the DPS protein from *Halobacterium salinarum*, suggesting their role in DNA association (*314-316*). A short C-terminal extension of the *Ss*DPSL protein also contains three basic amino acids that may also play a role in DNA binding (available as supporting information in the appendix, Figure 8.11S). Further investigations of DNA binding by archaeal

DPS's will be important for appreciating the significance of DPS and DPSL proteins in Archaea.

Discussion

The evolution of oxygenic photosynthesis marks the dawn of oxidative stress and represents one of the greatest selective pressures imposed on primordial life. The evolution of a single protein capable of managing the paradoxical relationship between iron and oxygen may represent an important ancestral component of the antioxidant defense system. We have identified and characterized a unique DPS-like protein from the hyperthermophilic archaeon *S. solfataricus* (*Ss*DPSL). Although the primary structure of this protein is only distantly related to characterized DPS proteins, the *Ss*DPSL protein does assemble into dodecameric cage-like structures that efficiently oxidizes Fe(II) to Fe(III) using H_2O_2 as the oxidant. In all previously characterized DPS proteins, this redox reaction is mediated by a unique diiron-binding motif (Figure S 8.11) (*286-288, 310, 315*). This motif has been implicated in coordinating the pair-wise reduction of H_2O_2. This mechanism has been shown to avoid the production of toxic hydroxyl radicals generated through the Fenton reaction (Eq. 4) (*288*). A putative DPS diiron-binding motif is identifiable in the primary structure of each new DPS sequence. While this putative metal-binding motif is distinct from that described in previously recognized DPS proteins, it does share

similar characteristics to known DPS metal binding sites including charge and motif spacing.

In exponential growth, the primary role of previously characterized DPS proteins is to minimize H_2O_2 stress. *In vivo* expression patterns of *Ss*DPSL are consistent with this protein's role in mitigating oxidative damage. *Ss*DPSL appears to be exclusively up regulated in exponential growth phase cultures of *S. solfataricus* in response to H_2O_2 stress and does not function in a general stress response. However, *Ss*DPSL is up regulated in exponential growth phase cultures of *S. solfataricus* under iron-limited conditions. Expression of this protein under iron limiting conditions appears contradictory given the well-established role of iron in oxidative stress. However, DPSL proteins use two ferrous ions as cofactors for the pair-wise reduction of H_2O_2. In the absence of ferrous iron, DPSL mediated reduction of hydrogen peroxide is compromised, resulting in the continued accumulation of H_2O_2. This increase in H_2O_2 likely induces the up regulation of the *Ss*DPSL protein. Although other cellular mechanisms are present to minimize H_2O_2 stress, DPS proteins play a proven role in mitigating oxidative damage. DPS knockouts have been shown to be more sensitive to H_2O_2 mediated oxidative damage during exponential growth (Figure S10.1; in the appendix) (*317*). While DPS expression is generally considered specific to oxidative stress during exponential growth, DPS proteins have been shown to effectively protect against a broad spectrum of different stresses while in

stationary phase (*318*). *Ss*DPSL is also likely to infer protection against a broader range of environmental stress during stationary phase, however growth of *S. solfataricus* is limited by the accumulation of inorganic ions and thus complicates studies aimed at understanding *S. solfataricus* physiologies under true nutrient limitation (*319*). Nevertheless, it is clear that the oxidative stress response in *S. solfataricus* is coupled to both H_2O_2 and iron levels through the expression of *Ss*DPSL.

It is of interest to recognize that sequences related to the *Ss*DPSL protein are not limited to thermoacidophilic Archaea. In fact, BLAST analysis indicates that *Ss*DPSL sequences are well distributed across the spectrum of phylogenetically diverse prokaryotes. This suggests that *Ss*DPSL sequences are not evolutionarily maintained solely as a consequence of protracting thermal stability. A case in point is the identification of a *Ss*DPSL sequence (85% similar) in the genome of a mesophilic bacteria (*Gloeobacter violacea* PPC 7421) (*320*). Although the amino acid sequence of this DPSL protein is highly conserved, the nucleotide sequence is not. The low G+C content of the *ssdpsl* gene (40%) is typical of genes originating from hyperthermophiles and is consistent with that of *S. solfataricus* (P2) genome. In contrast, the relatively high G+C content (60%) of the Dps-like gene from *G. violaceas* is consistent with that of the overall G+C content of the genome. This suggests that distribution of this newly identified *dpsl* gene is not a result of a recent horizontal gene transfer event.

This newly identified DPSL protein fits within the broad ferritin-like diiron-carboxylate protein superfamily (Figure 8.7). Other members of this superfamily include authentic DPS proteins, ferritins and bacterioferritins, which form three distinct sub-classes within the superfamily (*321*). All proteins within these, four subclasses (including the new DPSL sub-class) assemble into multimeric cage-like structures that functionally sequester iron. However, unlike the 24-subunit ferritins and bacterioferritins that function to sequester iron when concentrations are high, DPS proteins are 12 subunit assemblies that functionally sequester iron only as a consequence of reducing H_2O_2. Identification of a unique amino acid sequence in the *S. solfataricus* genome that self-assembles into a dodecameric cage and functions to mitigate oxidative damage has allowed us to confidently assign DPS-Like function to genes from nine different organisms whose function was previously unknown. The identification of a new sub-class of *dspl* genes provides further insights into the evolution and diversity of biochemical adaptations to oxidative stress.

CHAPTER 9

DPS-LIKE PROTEIN FROM THE HYPERTHERMOPHILIC ARCHAEON *PYROCOCCUS FURIOSUS*

Abstract

Oxidative stress is a universal phenomenon experienced by organisms in all domains of life. Proteins like those in the ferritin-like di-iron carboxylate superfamily have evolved to manage this stress. Here we describe the isolation and characterization of a Dps-like protein from the hyperthermophilic archaeon *Pyrococcus furiosus* (*Pf*DPSL). Phylogenetic analysis, primary structure alignments and higher order structural predictions all suggest that the *P. furiosus* protein is related to the new characterized DPS-like subclass of proteins within the broad superfamily of ferritin-like di-iron carboxylates. Recombinant *Pf*DPSL protein self-assembles into a 12 subunit quaternary structure with an outer shell diameter of ~10nm and an interior diameter of ~5nm. DPS proteins functionally manage the toxicity of oxidative stress by sequestering intracellular ferrous iron and using it to reduce H_2O_2 in a two electron process to form water. The iron is converted to a benign form, as Fe(III) within the protein cage. This Dps-mediated reduction of hydrogen peroxide, coupled with the protein's capacity to sequester iron, contributes to its service as a multifunctional antioxidant.

Background

The oxygenation of the atmosphere, estimated at two and a half billion years ago, has been termed by some as the greatest pollution event in history. While molecular oxygen is crucially important for many biological reactions, the increase of free oxygen led to the prevalence of reactive oxygen species (ROS), which are harmful to biological systems.

One such ROS is the superoxide ion ($O_2^{\cdot -}$) which is highly reactive as a nucleophile (*322*). Metal clusters, found in many proteins, are often the point of attack by the superoxide ion. More specifically, 4Fe-4S clusters in several citric acid cycle enzymes are targets of oxidation by superoxide (*298*). Oxidation of the Fe(II) in these clusters results in enzyme inactivation. As such, life without adequate mechanisms for coping with this toxicity perished. Evidence of this prehistoric selection event is can be observed in the genomes of contemporary life, where complex defense mechanisms protect against oxygen toxicity and iron stress.

Biochemical studies have identified enzymes dedicated to the detoxification of superoxide. Several of these enzymes catalyze the reduction of superoxide to hydrogen peroxide. Superoxide dismutase (SOD) was the first identified enzyme in this category (*301*). SOD reduces two superoxides with two protons and releases hydrogen peroxide and molecular oxygen, as shown in Eq. 1. SOD enzymes are found in aerobes, but not typically in anaerobic organisms (*295*).

$$2O_2^- + 2H^+ \rightarrow H_2O_2 + O_2 \quad (1)$$

A second enzyme in this category, to date only found in anaerobes, is superoxide reductase (SOR). SOR enzymes contain a non-heme iron at their catalytic site, and catalyze the reduction of superoxide to hydrogen peroxide without an oxygen byproduct using one electron (Eq. 2) (*323*).

$$O_2^- + 2H^+ + e^- \rightarrow H_2O_2 \quad (2)$$

Another class of antioxidants, catalases, act to remove the hydrogen peroxide byproducts of the SOD and SOR reactions. Catalases are widely occurring proteins, which have been found and characterized in all three domains of life (*324-326*). Catalases drive the disproportionation of hydrogen peroxide to form water and molecular oxygen according to Eq. 3.

$$2H_2O_2 \rightarrow 2H_2O + O_2 \quad (3)$$

To date, neither superoxide dismutase nor catalase-like sequences have been identified in the genome of *Pyrococcus furiosus* (*295*). This is not particularly surprising, as the mechanisms for both reactions release molecular oxygen (Eq.

1, 3), a toxin to this anaerobic hyperthermophilic archaeon. The *P. furiosus* genome was found instead to contain a superoxide reductase, which releases only hydrogen peroxide as a byproduct (Eq. 2) (*302*). However, superoxide reductase is not the only known source of peroxide production in *P. furiosus*. For example, the flavoprotein NADH oxidase (NOX1) from *P. furiosus* has been shown to evolve H_2O_2 during oxidation of NADH to NAD^+ (*327*). Although hydrogen peroxide is only considered a moderate oxidant itself, its interaction with Fe(II) and subsequent generation of hydroxyl radicals ($HO^{\cdot}$), through the Fenton reaction (Eq. 4), represents a lethal combination for life.

$$H_2O_2 + Fe^{2+} \rightarrow Fe^{3+} + OH^{-} + HO^{\cdot} \qquad (4)$$

Hydroxyl radicals are the most toxic of all reactive oxygen species and thus their generation is of significant biological consequence. Protection against the toxicity of hydroxyl radicals is reflected in the genomes of contemporary life, where multiple mechanisms have evolved to manage both components of the Fenton reaction (*265, 288, 292, 293, 328*).

DPS proteins, named after the DNA binding characteristics of the first characterized member (DNA binding protein from nutrient starved *E. coli*), are well suited for managing both iron and H_2O_2. Twelve copies of the DPS subunit protein self-assemble into a roughly 10 nm spherical cage-like quaternary

structure. The secondary and tertiary structures of DPS proteins are homologous to the multimeric assemblies formed by the ferritin family of proteins. Subunits from both proteins are composed of five alpha helices, four of which fold into a structurally conserved four-helix bundle. The fifth helix is found within a loop structure and defines the subunit dimer interface in structurally characterized DPS proteins. Although both DPS and ferritins serve to protect the organism from oxidative damage by sequestering intracellular ferrous iron, the active sites at which iron is bound and oxidized differ between the two protein classes. In ferritins, iron is bound in a carboxylate di-iron binding motif, located within a four-helix bundle. The location and general characteristics of this metal binding motif are conserved across ferritins and bacterioferritins. Strikingly, the structurally characterized DPS proteins contain a distinct di-iron metal binding site, located at the interface between subunits at the two-fold axis of symmetry (*286, 316*). While both O_2 and H_2O_2 can serves as oxidant for Fe(II) oxidation in ferritin, we and others have shown that H_2O_2 is the preferred oxidant for Fe(II) oxidation in DPS and bacterioferritins according to Eq. 5 (*7, 283, 288, 303*). Through this reaction, the activity of the DPS proteins simultaneously mitigates the H_2O_2 toxicity through a two electron reduction of H_2O_2, to form water, while oxidizing reactive Fe(II) to Fe(III). In this way generation of the toxic hydroxyl radical created during the Fenton reaction is avoided.

$$2Fe^{2+} + H_2O_2 + 2H^+ \rightarrow 2Fe^{3+} + 2H_2O \quad (5)$$

We have previously reported the isolation and characterization of a DPS-like protein from *Sulfolobus solfataricus* (*Ss*DPSL). Based on phylogenetic analysis, this *Ss*DPSL sequence, together with nine other hypothetical proteins form a monophyletic cluster that is distinct from all other protein subclasses within the di-iron carboxylate superfamily (*7*). We have termed this subclass of proteins "Dps-like" based on the dodecameric assembly and antioxidant properties of the prototypic *Ss*DPS-like protein. Another member of this newly identified monophyletic cluster is the Dps-like sequence from *Pyrococcus furiosus* (*Pf*DPSL). *P. furiosus* is a strictly anaerobic hyperthermophilic (70-103ºC, optimal 100ºC) archaeon, which thrives in slightly acidic (pH 6.4-6.8) marine thermal vents. These environments are commonly associated with high concentrations of iron and strongly reducing conditions (*329*). Here we provide biochemical evidence supporting the classification of this *P. furiosus* protein as a new member of the Dps-like subclass and describe its role as a functioning antioxidant *in vitro*.

Oxidative stress and organismal responses are operative in all domains of life. Many human diseases, such as Parkinson's, Huntington's, and Alzheimer's have established connections to oxidative stress. Therefore, understanding the biochemical responses to oxidative stress has far wider implications than

contributing to the understanding of archaeal biochemistry and may provide insights into the oxidative stress responses across all domains of life.

Material and Methods

Identification of *Pf*DPSL

The *P. furiosus* genome was queried for ferritin-like proteins through the Superfamily 1.67 HMM Library and Genome Assignments Server (http://supfam.mrc-lmb.cam.ac.uk/SUPERFAMILY/cgi-bin/genome.cgi?sf=47240;cgi_pu=yes) (*330*). Ten proteins were found to have ferritin-like domains. Sequences of the ten were individually compared to the *Ss*DPSL using the BLAST algorithm (blastp) to find similarities (*181*). The identified protein, annotated as a hypothetical protein (Pf1193), was subjected to a threading routine and submitted to the fold recognition server 3D-pssm (http://www.sbg.bio.ic.ac.uk/~3dpssm) (*304*).

Sequence Alignment

Amino acid sequences of *Pf*DPSL and five other ferritin-like diiron carboxylate proteins were aligned with CLUSTAL W 1.81 (*185*). Subsequent adjustments of these alignments were done manually.

Phylogenetic Analysis

The phylogenetic tree was generated using STRING (http://string.embl.de/) (*4*). The analysis is based on orthology information from the COG database (*331*) with a confidence score of 0.40.

Culturing of *P. furiosus*

Liquid cultures of *P. furiosus* (generously provided by Dr. Michael Adams, University of Georgia) were grown anaerobically for downstream DNA purification in a previously described complex salt based medium; 479.10mM NaCl, 15.20mM $MgSO_4 \cdot 7H_2O$, 13.29mM $MgCl_2 \cdot 6H_2O$, 4.43mM KCl, 4.68mM NH_4Cl, 955µM $CaCl_2$, 1.20µM $Na_4EDTA \cdot 2H_2O$, 12µM HCl, 12.32µM $FeCl_3 \cdot 6H_2O$, 809nM H_3BO_3, 367nM $ZnCl_2$, 176nM $CuCl_2 \cdot 2H_2O$, 253nM $MnCl_2 \cdot 4H_2O$, 255nM $(NH_4)_2MoO_4$, 169nM $AlK(SO_4)_2 \cdot 12H_2O$, 210nM $CoCl_2 \cdot 6H_2O$, 210nM $NiCl_2 \cdot 6H_2O$, 10µM $(NH_4)_2WO_4$, 4.13mM L-Cysteine free base, 6.41mM $Na_2S \cdot 9H_2O$, 11.90mM $NaHCO_3$, 450µM KH_2PO_4, 550µM K_2HPO_4, supplemented with 0.5% D-(+)-Maltose, 0.5% yeast extract and 0.5% casein hydrolysate at pH 6.8 (*332*). All cultures were grown under ultra high purity argon in sealed 200mL serum bottles at 95°C in a circulating oil bath.

Cloning and Expression

P. furiosus DNA was isolated from lab grown cultures (see above) using previously established methods and served as the template in PCR reactions (*305*). PCR primers were designed to amplify the Dps-like gene from *P. furiosus*

(Pf1193). The forward primer (5' at**tctaga**aataatttt*aaggag*atatacatatgccagagcataataggagattag 3') included an *Xba*I restriction endonuclease site (in bold) upstream of the ribosome binding site (in italics) and the start codon (underlined). The reverse primer (5' at**gaattc**ctatctagttttcaaaaactttgag 3') included the stop codon for the gene (underlined) and an *Eco*RI restriction site (in bold). The resulting PCR product was digested with *Xba*I and *Eco*RI restriction endonucleases, purified with a PCR purification kit (Qiagen Inc., Valencia, CA), ligated into pET-30a(+) (Novagen, Madison, WI) and transformed into XL-2 blue ultracompetent *Escherichia coli* (Stratagene). The cloned sequence was obtained by DNA sequencing (Applied Bio Systems, Foster City, CA). The plasmid was subsequently transformed into *E. coli* BL21(DE3) for protein expression (Novagen). For protein expression, *E. coli* were grown in LB broth with 30mg/mL kanamycin and protein was expressed through leaky expression. Cells were harvested by centrifugation and screened for *Pf*Dps expression by SDS-PAGE.

*Pf*DPSL Purification from *E. coli*

Cells were pelleted from one-liter cultures by low speed centrifugation and resuspended in 30mL of lysis buffer (50mM MES, 100mM NaCl, pH 6.5, 0.002mg/ml DNaseA, 0.05mg/ml RNaseI, 1mg/ml lysozyme). The slurry was French pressed, sonicated (3 x 5 min.), and centrifuged to remove cellular debris. The resulting supernatant was heated at 85ºC for 10 minutes, cooled on

ice, and then centrifuged to remove denatured proteins. The recovered supernatant was passed through a 0.2µm filter and loaded onto a Superose 6 size exclusion column (Amersham Biosciences) equilibrated with MES buffer (50mM MES, 100mM NaCl, pH 6.5). Elution of the protein was monitored at 260, 280, and 350nm.

Western Analysis

Approximately 8µg of protein was loaded in each lane and electrophoretically separated on a 15% SDS-polyacrylamide gel. Proteins were transferred to a Hybond-ECL nitrocellulose membrane (Amersham Pharmacia Bioscience) and stained with 5µM copper phthalocyanine tetrasulfonic acid tetrasodium salt in 12mM HCl (Kodak, Rochester, NY). The membrane was destained in 250mM $NaHCO_3$ and probed with polyclonal antibodies raised against purified recombinant *Ss*Dps-like protein (*7*). Colorimetric detection was used to identify the DPSL protein (Bio-Rad).

*Pf*DPSL Particle Characterization

Dynamic light scattering was performed on the assembled protein to determine particle size as previously described (*306*). Particle size was independently estimated by mobility on size exclusion columns (Superose 6, Amersham Biosciences) and by transmission electron microscopy (TEM, Leo 912AB operating at 80keV), stained with 2% uranyl acetate. Mass of the *Pf*Dps-

like protein was determined by electrospray mass spectrometry (Bruker Daltronics, MicroTOF LC).

*Pf*DPSL Fe(II) Oxidation with H_2O_2 and O_2.

0.2mg of *Pf*DPSL (9.43×10^{-6}mmol) was diluted into 3mL of MOPS buffer (25mM MOPS, 25mM NaCl, pH 6.5) in a quartz cuvette. A total of 3.9×10^{-5}mmol of de-aerated Fe(II) was added in 6 increments (50 Fe/cage) at 10 minute intervals. A half molar equivalent of hydrogen peroxide (1 H_2O_2: 2 Fe^{2+}) was delivered two minutes after each iron addition. A UV-Vis spectrum was taken every ten seconds to monitor Fe oxidation by the increase in absorbance in the region from 350 to 400nm. Light scattering due to increase in particle size (and bulk precipitation of Fe-oxyhydroxide precipitates) was monitored at 800nm. The same protocol was also repeated using air oxidation instead of H_2O_2.

Results

*Pf*DPSL Identification

The *P. furiosus* DPS-Like (*Pf*DPSL) sequence was found using the default parameters of BLAST (blastp) (*181, 323*). This 185 amino acid protein (Pf1193), annotated as a conserved hypothetical protein, has a predicted molecular weight of 21,357Da. The sequence threads, with high confidence (4.5e-13), onto the four-helix bundle structure of proteins belonging to the ferritin-like di-iron

carboxylate family (*304*). Due to the similarity in predicted secondary and tertiary structures of authentic DPS proteins, the *P. furiosus* DPS-Like protein was also expected to assemble into a dodecameric and roughly spherical cage, consistent with other members of the DPS and DPSL families of proteins (*287*).

Phylogenetics

The *Pf*DPSL sequence belongs to a new monophyletic cluster within the ferritin-like di-iron carboxylate superfamily (*7, 321*). This new subclass has been termed "DPS-Like," due to their apparent structural and functional similarities with authentic DPS proteins. Sequences related to the *Pf*DPSL sequence are conserved in organisms spanning both prokaryotic domains of life and is present in organisms that occupy diverse ecological niches (Figure 9.1). This broad distribution includes both mesophilic and hyperthermophilic organisms from both the Archaeal and Bacterial domains of life, suggesting that the evolutionary conservation of this sequence is not related to thermal stability but rather to attributes which extend beyond those of traditional DPS family proteins.

Recombinant *Pf*DPSL

The recombinant *Pf*DPSL protein expresses to high levels in the heterologous *E. coli* expression system. Approximately 25mg of the recombinant protein can be purified from a 1L *E. coli* culture in LB broth. Purification of the recombinant *Pf*Dps-like protein was simplified by taking

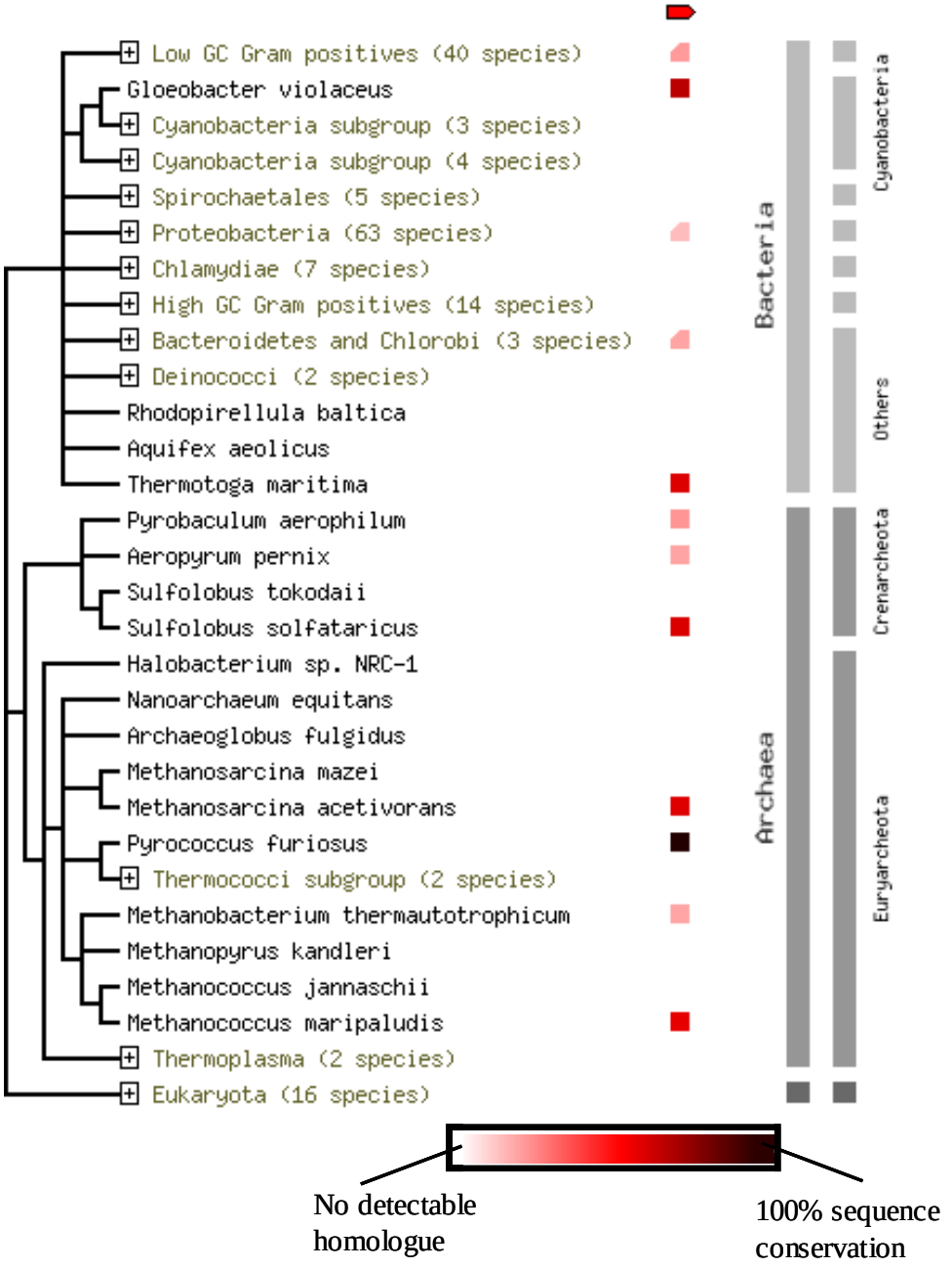


Fig. 9.1 Phylogenetic distribution of PfDPSL related proteins. The PfDPSL sequence is conserved in phylogenetically diverse prokaryotes spanning both the Bacterial and Archaeal domains of life. Organisms encoding PfDpsl related sequences occupy diverse ecotypes, including both microaerophilic and obligate anaerobic organisms. Additionally, both mesophiles and hyperthermophiles are represented. The phylogenetic tree was generated at the STRING site (http://string.embl.de/) (*4*) using orthology information from the COG database (*7, 11*).

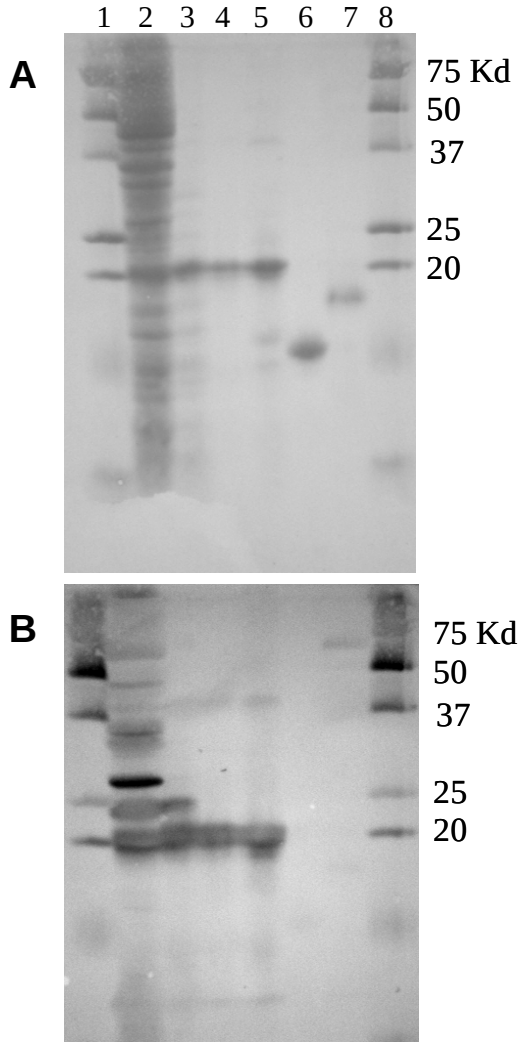


Fig. 9.2 Heterologous expression and purification of *Pf*DPSL. (A) A 15% SDS-PAGE, transferred to a nitrocellulose membrane and stained with copper phthalocyanine tetrasulfonic acid. Lane 1, molecular mass standard; lane 2, lysed *Escherichia coli* (BL21) DE3 cells transformed with the pET-30a (+) expression plasmid carrying the *Pf*Dps gene; lane 3, after heat denaturation at 85ºC for 10 min.; lane 4, after size exclusion liquid chromatography. Lanes 5, 6, and 7 are the *S. solfataricus* Dps-like protein, *L. innocua* Dps, and Human H-chain ferritin, respectively, for molecular weight comparison. (B) Western blot of the above membrane. The *Pf*DPSL and *Ss*DPSL proteins are 66% identical at the amino acid level and polyclonal antibodies generated against the *Ss*DPSL protein cross react with the *Pf*DPSL protein (lanes 2-5), whereas *Li*DPS and Human H-chain ferritin are not detected by the *Ss*DPSL antibody.

advantage of the protein's thermophilic origin (Figure 9.2A). A ten-minute incubation at 85˚C was sufficient to denature a majority of the *E. coli* proteins, which were precipitated by a 12000xg centrifugation. The recombinant protein eluted as a single peak on a pre-packed Superose 6 size exclusion column (Figure 9.3). SDS-PAGE of the purified recombinant protein reveals a dominant band with an estimated molecular mass of 22kDa (Figure 9.2A). This is consistent with the deconvoluted mass of 21220.7Da, as determined by electrospray mass spectrometry (Bruker Daltronics, MicroTOF LS), and is in agreement with the theoretical mass of the recombinant PfDPSL protein. A second, faint band running at the approximate size of a dimer (~40kDa) is visible in both the SDS-PAGE and in the Western (Figure 9.2B). Dimer stability maybe a consequence of covalent cross-linking between subunits. Polyclonal antibodies raised against the recombinant *Ss*DPSL protein cross-react with the recombinant *Pf*DPSL protein. This is not surprising given the high similarity between the two sequences (66% identity). DNA sequencing of multiply *pfdpsl* clones consistently reveals a two nucleotide difference, between cloned versions of the gene and the published genome sequence. These nucleotide differences are predicted alter the amino acid sequence at two locations (T12A and K162R). Although the identification of these residues may be important they are not in highly conserved regions of the DPSL family, do not interfere with cage assembly,

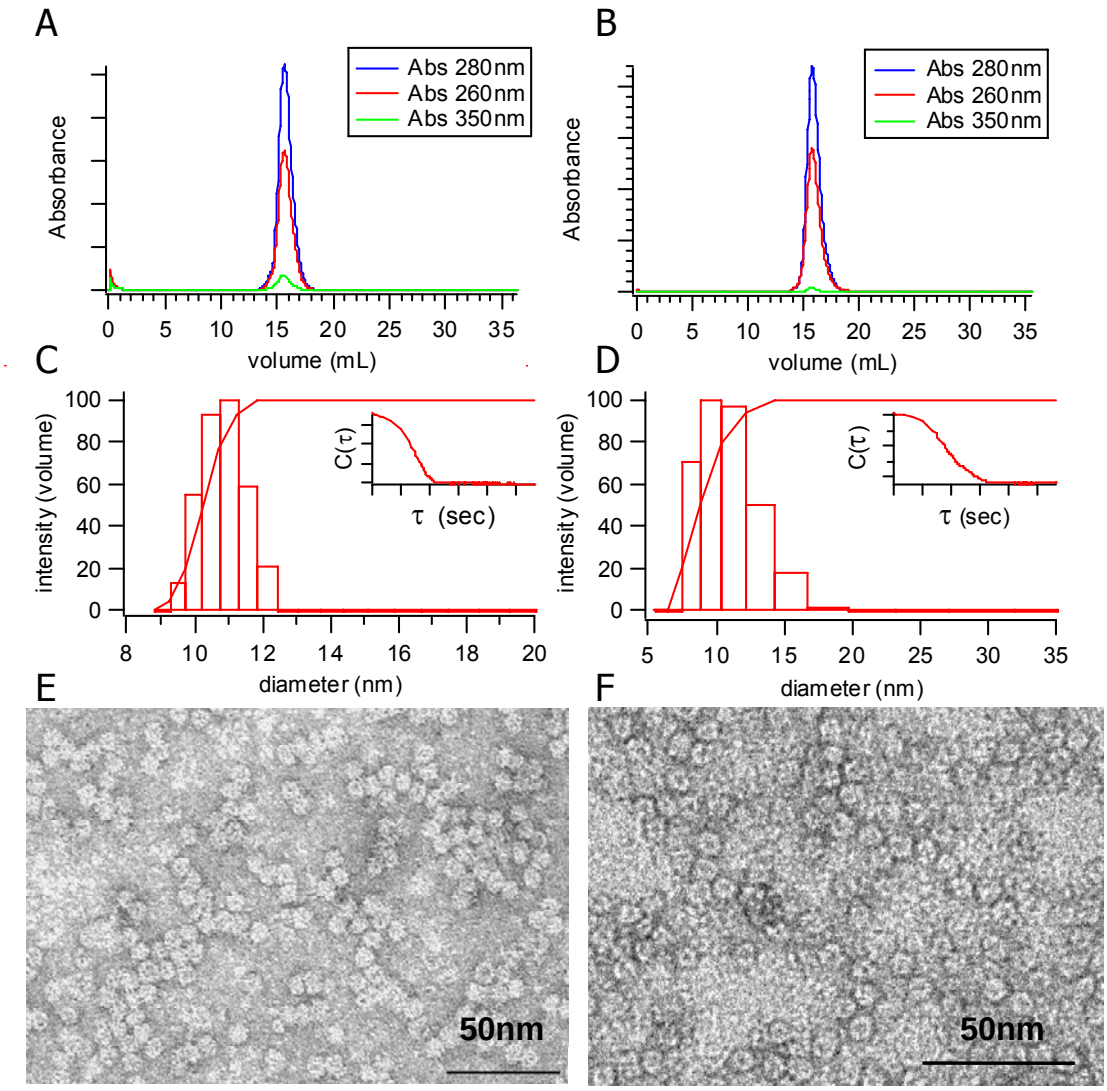


Fig. 9.3 Size exclusion liquid chromatography chromatograms, dynamic light scattering, and TEM images spectra of corresponding peaks. Elution volumes from size exclusion liquid chromatography are shown for (A) the 12 subunit DPSL protein from *P. furiosus* and (B) the 12 subunit *S. solfataricus* (P2) DPSL. The *S. solfataricus* (P2) DPSL and the DPSL protein from *P. furiosus* both have retention times consistent with a 12 subunit, 260kDa protein Dynamic light scattering reveals a mean diameter of 10.54 ± 0.02 nm for PfDps (C) and 9.71 ± 0.24 nm for SsDps (D). Transmission electron microscopy at 31,500· magnification reveals intact cages with diameters of ~10 nm for both PfDPSL (E) and SsDPSL (F).

and are not in regions likely to be critical for metal binding, and thus are not expected to compromise the functional properties of this protein. Dynamic light scattering (DLS) of the recombinant protein indicates an average diameter of 10.54nm $\pm$ 0.02nm, consistent with a 12 subunit cage-like assembly (Figure 9.3C). This architecture is directly evident from negatively stained transmission electron micrographs (TEM) (Figure 9.3E) in which the *Pf*DPSL cage has a measured exterior diameter of ~10nm and an interior cavity of ~5nm, similar to other previously characterized DPS proteins. This implies that the quaternary structure will be based on tetrahedral (23) symmetry as predicted (*286*).

*Pf*DPSL Catalyzed Fe(II) Oxidation

At pH 6.5 the recombinant *Pf*DPSL protein efficiently catalyzes the oxidation of Fe(II) to Fe(III) in the presence of H_2O_2 (Figure 9.4). *Pf*DPSL mediated oxidation of Fe(II) is dependent on H_2O_2 as the oxidant, and substitution of O_2 for H_2O_2 results in no measurable activity for the Fe(II) oxidation. This is in contrast to Fe(II) oxidation by all known ferritins, which efficiently catalyze the oxidation of Fe(II) with either O_2 or H_2O_2 as an oxidant. Iron oxidation reactions were monitored by the increase in absorbance at 350 nm. Fe(II) was added to the protein in aliquots of 50 Fe atom per protein cage and oxidized in the presence of H_2O_2 or O_2. This iron loading process (50 Fe:25 H_2O_2) was repeated 6 times resulting in total loading of 300 Fe per *Pf*DPSL cage. Reactions were also monitored at 800nm where no electronic transitions for Fe-O are expected. No

increase in absorbance (scattering) was observed, which is an indication that there was no increase in particle size due to bulk precipitation of Fe(III)-oxide during the course of the reaction. In contrast, control reactions in the absence of the DPSL protein resulted in rapid accumulation of an insoluble Fe-oxyhydroxide-like material. This suggests a role for the *Pf*DPSL in Fe biomineralization, as has been suggested for other DPS proteins, although we have no evidence at present for iron oxide mineral core formation in *Pf*DPSL.

Iron Binding Motifs

Recombinant versions of the *Pf*DPSL and *Ss*DPSL proteins appear to form dodecameric quaternary structures similar to those of the prototypic DPS family. In addition to these structural similarities, *in vitro* Fe(II) oxidation reactions indicate that these proteins, like DPS proteins, prefer H_2O_2 as the oxidant. Given these similarities we were previously inclined to look for an iron biding motif similar to that recognized in authentic DPS proteins (*7*). Although, a motif related to the intrasubunit di-iron binding motif of authentic DPS proteins can be identified, it has not been experimentally validated. Additional sequence alignments, comparing DPSL sequences to three different 24-subunit ferritins has identified a di-iron binding motif that is homologous to the ferroxidase site in (bacterio)ferritins (Figure 9.5). Although a bacterioferritin-like di-iron binding motif is not expected in these proteins, given their similarity to DPS proteins, the di-iron binding motif of bacterioferritins is remarkably well conserved in DPSL

proteins. High resolution structural analysis will be required to confidently identify the true iron binding motif in this new protein family.

Discussion

Oxidative stress represents a universal selective pressure imposed on all life. The evolution of a single protein capable of managing the relationship between iron and oxygen may represent an important ancestral component of the antioxidant defense system. We have biochemically characterized an antioxidant protein from the hyperthermophilic archaeon *Pyrococcus furiosus* (*Pf*DPSL) that may function to manage two aspects of oxidative stress (Fe and H_2O_2). The recombinant *Pf*DPSL protein self assembles into a dodecameric cage-like structure, that efficiently oxidizes Fe(II) to Fe(III) using H_2O_2 as a preferred oxidant. Although the mineralization mechanism for this protein has not been elucidated, multiple sequence alignments comparing DPSL sequences to authentic ferritins has identified a di-iron binding motif that is homologous to the ferroxidase site in both ferritins and bacterioferritins. The highly conserved nature of this motif strongly suggests that this protein will employ a ferritin-like mineralization mechanism.This protein belongs to a newly identified subclass of DPS-Like protein that fit within the broad ferritin-like di-iron-carboxylate superfamily. Other members of this superfamily include the prototypic DPS proteins, ferritins and bacterioferritins, which form three distinct subclasses

within the superfamily (*321*). We have designated this new subclass of proteins as "DPS-Like." This designation is based on characteristics that these proteins share with authentic DPS proteins; namely, cellular expression in response to H_2O_2 stress and their quaternary assembly into dodecameric protein cages. This designation may however prove misleading, as we do not yet know why this sequence is preferentially maintained over the traditional DPS sequences,

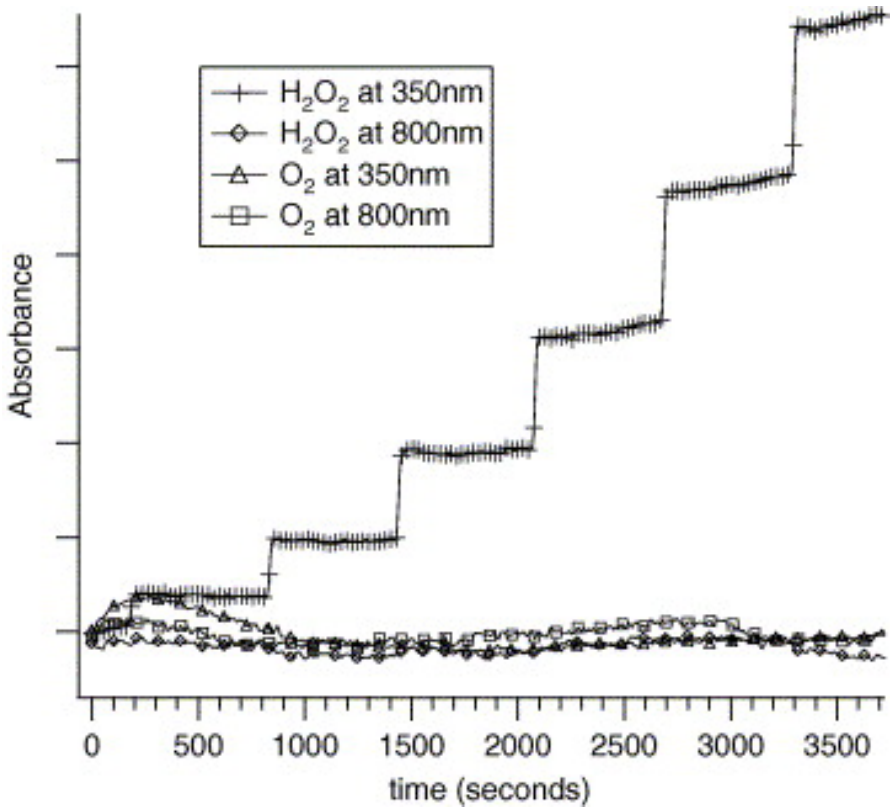


Fig. 9.4 *Pf* DPSL catalyzed oxidation of Fe(II). *Pf* DPSL uses H_2O_2 to efficiently oxidize Fe(II) to Fe(III) in vitro (1 H_2 O_2 :2 Fe). In contrast, O_2 serves as a relatively poor oxidant of iron in this reaction. The iron oxidation reactions done using H_2O_2 as the oxidant were monitored at 350 nm (+) and H_2 O_2 at 800 nm (◇). Oxidation reactions done using O_2 as the oxidant were monitored at the same wavelengths, labeled Δ and □, respectively.

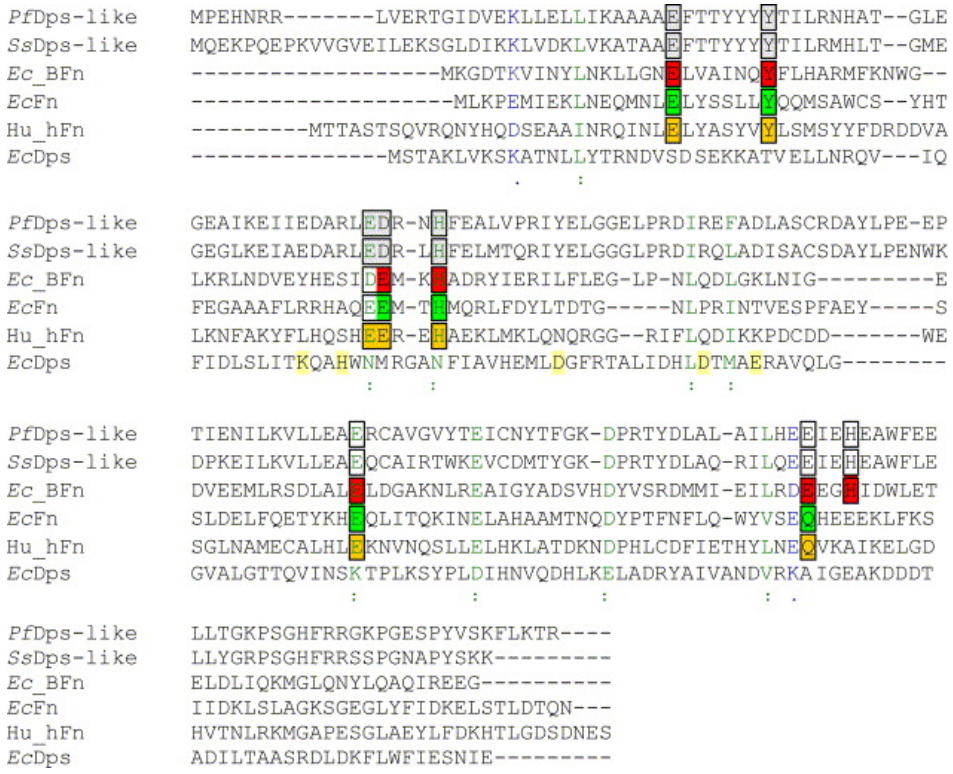


Fig. 9.5 Sequences representing each of the four subclasses within the broad superfamily of ferritin-like di-iron-carboxylate proteins were aligned using CLUSTALW. Experimentally identified iron binding motifs are indicated for *E. coli* bacterioferritin (*Ec* _BFn; red), *E. coli* ferritin (*Ec* Fn; green), *Homo sapiens* H-Chain ferritin (Hu_hFn; gold) and *E. coli* DPS (*Ec* Dps; yellow). The DPSL proteins from both *P. furiosus* (*Pf* Dps-like) and *Sulfolobus solfataricus* (*Ss* Dps-like) contain a di-iron binding motif that is similar to that in (bacterio)ferritins (boxed). This motif is not present in authentic DPS proteins.

which also function as antioxidants. The primary structure of proteins within this new "DPS-Like" subclass are as different from ferritin sequences as they are from the sequences of the prototypic DPS subclass, and thus, labeling these proteins as "DPS-Like" may prove to be inappropriate. This new subclass is expected to have unique functional characteristics which distinguish them from all other protein subclasses within this superfamily.

The *Pf*DPSL protein has been identified as a new member of the DPS-Like subclass. Here we have biochemically validated this protein's role as an antioxidant and confirmed its earlier phylogenetic placement within this new subclass. It is of interest to note that the *Pyrococcus furiosus* genome codes for both a DPS-Like protein (PF1193) and ferritin (PF0742) and multiple ferritin-like proteins (PF1190,1195,1196, 1199), while there is no clearly identifiable ferritin gene encoded in the *Sulfolobus solfataricus* genome. Future investigations of proteins within the new DPS-Like subclass will provide insights into the evolutionary division of proteins in this superfamily and the roles they play in biochemical adaptations to iron and oxidative stress.

Interestingly, both the *Ss*DPSL and the newly characterized *Pf*DPSL proteins appear to incorporate a bacterioferritin ferroxidase-like di-iron binding motif, which is clearly absent from authentic DPS proteins. Thus, these new DPS-Like proteins represent the first dodecameric assemblies within the ferritin-like di-iron carboxylate superfamily that contain ferritin-like ferroxidase centers. This new monophyletic cluster may represent a deeply rooted branch within the ferritin-DPS family, containing characteristics of both ferritin-like ferroxidase metal binding as well as DPS metal binding capacity. Higher resolution structures and detailed mechanistic studies should reveal the functional purpose for the evolutionary conservation of distinct DPS-Like sequences in the genomes of such distantly related organisms.

CHAPTER 10

GLOBAL EVALUATION OF THE OXIDATIVE STRESS RESPONSE IN THE THERMOACIDOPHILIC ARCHEAON *SULFOLOBUS SOLFATARICUS*.

Introduction

The evolution of oxygenic photosynthesis marks the dawn of oxidative stress and represents one of the greatest selective pressures experienced by primordial life. The antiquity and universal toxicity of oxidative stress has driven the evolution of diverse antioxidant defense networks that are evident in the genomes of contemporary life. Although the oxidative stress response has been assessed in both bacterial and eukaryal life; the archaeal defense system has yet to be evaluated. Here we present the first comprehensive genomic analysis of the oxidative stress response in the archaeal domain of life.

Sulfolobus solfataricus is an aerobic Crenarchaeota that thrives in high temperature (~80ºC) acidic environments (pH2-4). The P2 stain of *S. solfataricus* has a single circular chromosome of 2,992,245 bps that codes for an estimated 3,000 ORFs. Approximately 40% of these putative genes have orthologs found exclusively in Archaea, ~12% in bacterial genomes and ~3% are shared exclusively with eukaryotes. Although *Sulfolobus* spp. are among the most widely studied organisms in the archaeal domain, ~45% of annotated genes in the *S. solfataricus* genome have not been formally assigned function.

In this study we use two different microarray platforms to identify genes associated with the temporal oxidative stress response of *S. solfataricus*.

Materials and Methods

Culturing of *S. solfataricus*

Liquid cultures of *S. solfataricus* (P2), were grown aerobically in DSMZ media 182 (22.78 mM KH_2PO_4 + 18.90 mM $(NH_4)_2$ SO_4 + 0.81 mM $MgSO_4$ + 1.7 mM $CaCl_2$ + 0.2% Yeast Extract) pH adjusted to 2.8 with 6N H_2SO_4. All cultures were grown in long neck Erlenmeyer flasks. Shaker temperature was maintained at 86°C; however measured culture temperatures remained steady at 75°C.

Growth Curves

Three liters of *Sulfolobus sulfataricus* media 182 (pH~2.8) were inoculated with 15 mls of a late-log phase (OD_{650} 0.52) *S. solfataricus* culture and then divided evenly between three, 2-liter long neck culturing flasks. Culture progress was monitored by optical density at 650 nanometers (OD_{650}). At 62.5 hrs (D_{650} ~0.3) 20 mls of each 1 liter culture was removed (pink line) and grown in a separate 50 ml flask as a growth control. An additional 50 ml aliquot was collected from each culture and used for RNA isolation (t=0). The three cultures (~930 mls) were each stressed with hydrogen peroxide, at a finial concentration of 30μM (yellow line). At 30 mins post H_2O_2 exposure an additional 50 ml aliquot, also used for RNA isolation, was collected. This protocol was repeated,

again preformed in triplicate; however in addition to samples collected at t=0 and t=30 an additional sample was collected at and again at 60 mins post exposure (t=60).

RNA Isolation and Quality Assessment

Total cellular RNA was extracted from *S. solfataricus* (P2) cells, according to the Qiagen's RNeasy midi protocol, with an on-column DNase step (Valencia,CA). Total RNA concentrations were estimated using a Nanodrop spectrophotometer ($OD_{260/280}$). RNA quantity and quality was independently assessed by visualization on a 1.5% agarose (wt/vol) formaldehyde gels and on an Agilent 2100 Bioanalyzer (Agilent Technologies, Palo Alto, CA).

In Vivo Expression Patterns of the *ssdpsI* Gene

For Northern analysis, ~1.2 µg of total RNA was separated by electrophoresis in a 1.5% agarose (wt/vol) formaldehyde gel and transferred to GeneScreen membranes as recommended by the manufacturer (NEN, Wellesley, MA). RNA was membrane-crosslinked in a UV Stratalinker (Stratagene, La Jolla, CA). Blots were probed with ^{32}P-labeled PCR products (Ready-To-Go Labeling Kit, Amersham Pharmacia Bioscience). *S. solfataricus* DNA was isolated by using previously established methods and served as the PCR template.

Microarray

Two different *S. solfataricus* (P2) microarray platforms were independently used to assess the organism's transcriptional response to 30µM H_2O_2. The

Nimblegen oligo expression array includes 15 probe pairs (24mers) per target for a total of 3082 genes. Hybridizations, array scans, data extraction and preliminary data analysis were performed by NimbleGen (Madison, WI).

The Isogen Lifescience oligo array includes a total of 3318 oligo probes, ranging from 50-70 nt in length (melting temperature 78ºC $\pm$ 5ºC). Oligonucleotide synthesis and probe spotting was preformed by Isogen Lifescience. Each probe is represented twice on every chip in spatially distinct locations. cDNA libraries from each sample were constructed in reactions including 1.5µg of total RNA, 5µg random hexamer primers, 5xRT buffer, 0.5mM dNTP mix (4:1 amino allyl dUTP/dTTP) and Supperscript II and dithiolthreitol (DTT). All products except aadNTPs (Sigma) and dNTPs (Promega) were purchased from Invitrogen. Primers and RNA template were incubated for 10 min. at 70ºC and cooled on ice for 2 min. prior to primer extension; performed at 42ºC for 2hrs. RNA template was subsequently degraded in 20mM EDTA and 50mM NaOH. cDNAs larger then ~70nt were purified using MinElute (Enzymatic Clean up Kit) filtration columns (Qiagen). cDNA libraries were fluorescently labeled at room temperature, in the dark with either Cy3 or Cy5 (Amersham), for 1.5hr. Excess dye was removed using the MinElute kit. Slides were prehybidized in 50 mls of 5xSSC, 0.1% SDS, 0.5g BSA for 40 mins. at 42ºC. Labeled cDNAs were combined in a hybridization mixture (27µl formamide, 15µl 20xSSC and 0.6µl 10% SDS), applied to the array, covered with a lifter slip (Erie Scientific,

NH) in a hybyridization chamber (Arrayit, CA) and incubated at 42^{o} C overnight (16-20 hrs). After hybridization, each microarray was washed with 2xSSC, 0.1%SDS for 5 minutes at 42^{o} C, then with 0.1xSSC, 0.1%SDS for 20 minutes at 42^{o} C and rinsed in 5 times in 0.1xSSC at room temperature. After hybridization, arrays were scanned at 10 µm resolution using an Agilent scanner, Model G2565B (Agilent Technologies). Primary data collection and analysis were carried out using GenePix Pro 6.0 (Axon Instruments).

Analysis

Most of the microarray data collected in this experiment has not been analyzed. Comments in the section below are based only on preliminary analysis comparing RNA isolated from two independent replicates. Preliminary analysis is limited to t=0 RNA vs. RNA isolated at 30 mins post H_2O_2 exposure. GenePix Pro 6.0 was used to align the grid and evaluate spot quality. Low-quality spots were flagged and exclude from further analysis based on previously described criteria (*252*). Normalization was done by averaging the Cy5/Cy3 ratio to 1, with no background subtraction. Intensities from duplicate spots on the array were averaged and transformed by the Log_2. Each Log_2 value represents two hybridization experiments performed in duplicate, (t=0Cy3 vs. t=30Cy5) and a dye-swap (t=0Cy5 vs. t=30Cy3) performed using cDNA derived from the same two *S. solfataricus* cultures. Statistical significance for observed ratios was

calculated using SAM (significance analysis of microarrays) and is reported at the 90% confidence interval (*333*).

Data from the Nimblegen array has been normalized (by Nimblegen) using some tools available through the Bioconductor project (http://www.bioconductor.org) using quantile normalization (*334*) and a preliminary gene list has been provided by Dr. C. Martin Lawrence.

Results

We have recently characterized a novel antioxidant protein form *S. solfataricus* (SsDPSL) (*7, 17*). The amino acid sequences of this protein and orthologous sequences in the genomes of phylogenetically diverse prokaryotes form a monophyletic subclass within the broad superfamily of ferritin-like diiron carboxylate proteins (*7, 17, 25*). Similar to other members in this superfamily, the SsDPS-like (DPSL) protein self-assembles into a hollow dodecameric cage-like structure. Northern and Western blot analyses have demonstrated the up-regulation of these enzymes in response to hydrogen peroxide (*7, 25*). Although the precise mechanism(s) of antioxidation have not been clearly elucidated a dimetal binding site, located in the center of each subunit (4-helix bundle) has been strongly implicated in the two electron reduction of H_2O_2 (*7, 17, 25*). Without *a priori* information on the response time of the *S. solfataricus* antioxidant defense network, we performed a high resolution time course to

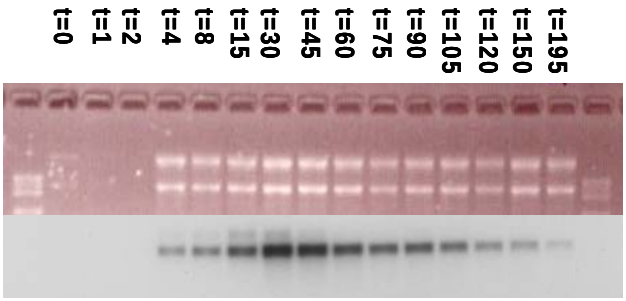


Fig 10.1 The *ssdpsI* transcript increases in abundance until the signal peaks at 30 minutes post H_2O_2exposure. Northern blot analysis of total RNA isolated from exponential growth phase culture of *S. solfataricus* identifies a dominate transcript corresponding to the approximate size of the *ssdpsI* gene (~600 nt), at 30mins post H_2O_2 exposure. A second, larger transcript (~900 nt), is transiently detectable in between 8-45 mins. post exposure. The larger band may suggest transcriptional read-through of the *ssdpsI* gene.

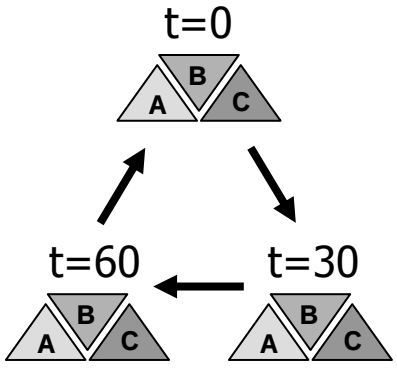


Fig 10.2 Schematic of the microarray experimental design. *S. solfataricus* samples were collected at three different time points (t=0, t=30 and t=60 mins.), from three independent cultures (A, B and C). RNA isolated for the Nimblegen array was collected from three additional cultures sampled at t=0 and t=30 mins.

assess *ssdps1* expression following H_2O_2 exposure (Figure 10.1). According to Northern blot analysis, the *ssdps1* transcript peaks at ~30 mins. post H_2O_2 exposure. Using the *ssdps1* gene as a hallmark for the oxidative stress response we designed a microarray experiment to assess gene expression differences between samples collected at t=0, t=30 and t=60 (Figure 10.2). Preliminary analysis of microarray data on the *S. solfataricus* oxidative stress response is corroborated by biological replicates performed on two different microarray platforms (Figure 10.3). Ten of the top 15 genes up-regulated on the NimbleGen array are also up-regulated according to the custom array ($p < 0.1$). Data from the NimbleGen array indicates that the *ssdps1* gene is the most up-regulated gene in the *S. solfataricus* transcriptome at 30 minutes post H_2O_2 exposure. Although we anticipated the up-regulation of this gene upon H_2O_2 exposure, we could not have predicted how this gene would be regulated relative to other oxidative stress response genes in *S. solfataricus*. Many of the genes tentatively identified as oxidative stress response genes are conserved in the other two *Sulfolobus* genomes (Figure 10.4). Eight of the ten genes up-regulated on both *S. solfataricus* arrays are present in *S. acidocaldarius* genome. These genes are often oriented in the same direction and are located in a similar genomic context. The *S. tokodaii* genome is contains nine genes related to the ten *S. solfataricus* oxidative stress genes. However, many of the *S. tokodaii* genes do not share a similar genomic context and are often coded on the opposite strand, with respect

to the other genomes. Perhaps most notable, is the absence of a recognizable *dpsl* gene in the *S. tokodaii* genome. The *S. solfataricus* oxidative stress genes detected on both arrays fall into 3 major categories; antioxidant proteins (SSO2079, 2121, 2642 and 2643), metal transporters (SSO0928, 2078, 2080 and 2568) and hypothetical proteins, with no recognizable domains (SSO2644 and 2645).

A

Custom Slide vs. NimbleGen

	Gene ID	q-value(%)		Fold-Up
1	**SSO2079**	0	**SSO2079**	52.68151
2	**SSO2080**	0	**SSO2080**	36.98263
3	**SSO2078**	0	**SSO2078**	27.56076
4	**SSO2121**	0	**SSO2121**	13.85793
5	**SSO2568**	0	**SSO2568**	12.14592
6	**SSO2644**	0	**SSO2644**	10.60498
7	**SSO2643**	0	**SSO2643**	9.109113
8	**SSO2645**	0	**SSO2645**	8.622496
9	**SSO2621**	0	**SSO2621**	5.074682
10	SSO2642	0	**SSO0928**	4.15569
11	SSO11867	0	SSO2244	3.946349
12	SSO1054	0	SSO0923	4.217787
13	SSO0366	0	SSO0245	3.974587
14	SSO0927	3.233688	SSO1933	4.723718
15	**SSO0928**	9.054326	SSO2096	4.273313

B

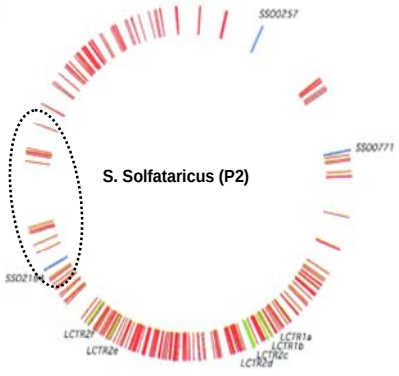


Fig 10.3 Preliminary microarray data on the *S. solfataricus* oxidative stress response is corroborated by biological replicates preformed on two different microarray platforms. A. Bold type indicates genes that are found to be up-regulated on both array platforms and gray boxes highlight neighboring genes. Reported q-values were calculated using SAM and are based on the dye-swap experiment. The top 15 up-regulated genes are listed for the NimbleGen custom array. These are based on relative intensities calculated from two different slides. B. Genomic location of genes up-regulated in response to hydrogen peroxide. This Region of Oxidative Stress Response has tentatively been coined (ROSTR), to reflect the functional and spatial relationship of genes in this area.

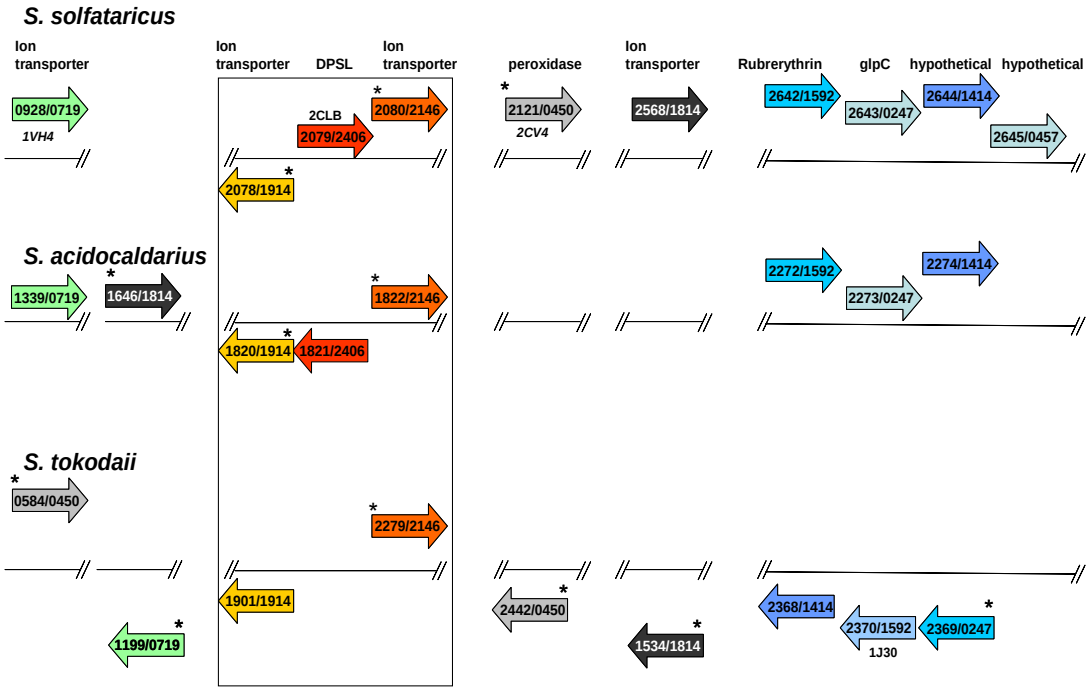


Fig. 10.4 Comparative genomics of genes involved in the *S. solfataricus* oxidative stress response. The top ten genes up-regulated at 30 mins. post H_2O_2 exposure, according to both array platforms, are schematically mapped on each of the three sequenced Sulfolobus chromosomes. Each gene is represented as a color coded arrow containing the gene identification number and the COG number (Cluster of Orthologous Genes), respectively. Orthologous genes share a common color. The presence of highly related sequences within a particular genome, are denoted with an asterisk (*). Protein data bank (pdb) identification numbers are indicated when available. Italicized pdb numbers are used when only the structure of a related sequence is available.

When we consider 10 additional genes implicated by one or the other arrays, but not on both, the trend continues to include iron transporters (SSO0927), the ferric uptake regulator (SSO2244), phosphate and ammonium transporters (SSO0923 and 1054 respectively), a putative glutamine synthetase (SSO0366) and another hypothetical protein with no recognizable domain (SSO1186). Many of these genes are highly conserved in other genomes.

Perspectives and Publication Strategies

This project was originally motivated by the fortuitous discovery and subsequent characterization of a novel antioxidant protein from *S. solfataricus,* termed *Ss*DPSL (*7, 17, 25*). Though we were originally interested in finding a ferritin in *S. solfataricus*, the cage-like super-structure and iron binding activity of this new DPSL protein captured our attention. In many respects the discovery of this protein has broadened our perspective on the many roles of iron in biology. In fact the antioxidant activity of this protein has challenged all of us all, at one point or another, to reconsider the biological importance of this protein and its role in managing the mingled toxicity of iron and oxygen. Preliminary analysis of the microarray data punctuates the iron/oxygen relationship by implicating numerous metal transport proteins in the oxidative stress response. This relationship between oxidative stress and metal transport systems has recently been noted in *Rhodobacter sphaeroides* (*335*).

Processing and analyzing the remaining microarray data is of course, the first step toward publication. Although microarray papers are currently a dime-a-dozen; at this point in time there are only a few published reports of arrays done on archaeal organism and none on the archaeal oxidative stress response. Given the novelty of this biological system, combined with the use of the two different array platforms; I anticipate results from this project to be publishable without additional support from the proteomic analysis. Perhaps preliminary proteomic

analysis could be incorporated in this manuscript, with a more thorough analysis that includes the identification of posttranslational modifications and there possible biological role, being the theme of a second manuscript.

The *LacS* insertion disruption of the *dpsl* gene in the 98/2 strain of *S. solfataricus* is significantly more sensitive to H_2O_2 than are wild type 98/2 cells (Supplemental Figure 10.1). RNA isolated from the 98/2 strain has recently been analyzed on the *S. solfataricus* (strain P2) microarray. Preliminary indications suggest high similarity between the 98/2 and P2 strains of Sulfolobus (personal communication Brouns and Blum). The 98/2 *dpsl* gene was disrupted in collaboration with Dr. Paul Blum's lab (Supplemental Figure 10.2). The mutant has been confirmed by polymerase chain reaction and by Western blot analysis. A microarray experiment assessing differences in the oxidative stress response of 98/2 as compared to that of the *dpsl* mutant may provide further insights into the biological role of DPSL family of proteins.

REFERENCES CITED

(1) Oksanen, J. (2005), University of Oulu, Oulu, Finland.

(2) (http://www.R-project.org), R. D. C. T. V. (2005).

(3) Chave, J., and Jabot, F. (2006).

(4) von Mering, C., Jensen, L. J., Snel, B., Hooper, S. D., Krupp, M., Foglierini, M., Jouffre, N., Huynen, M. A., and Bork, P. (2005) STRING: known and predicted protein-protein associations, integrated and transferred across organisms. *Nucleic Acids Research 33*, D433-D437.

(5) Lawson, D. M., Artymiuk, P. J., Yewdall, S. J., Smith, J. M. A., Livingstone, J. C., Treffry, A., Luzzago, A., Levi, S., Arosio, P., Cesareni, G., Thomas, C. D., Shaw, W. V., and Harrison, P. M. (1991) Solving the Structure of Human H-Ferritin by Genetically Engineering Intermolecular Crystal Contacts. *Nature 349*, 541-544.

(6) 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.

(7) Wiedenheft, B., Mosolf, J., Willits, D., Yeager, M., Dryden, K. A., Young, M., and Douglas, T. (2005) An archaeal antioxidant: Characterization of a Dps-like protein from Sulfolobus solfataricus. *Proceedings of the National Academy of Sciences of the United States of America 102*, 10551-10556.

(8) Khayat, R., Tang, L., Larson, E. T., Lawrence, C. M., Young, M., and Johnson, J. E. (2005) Structure of an archaeal virus capsid protein reveals a common ancestry to eukaryotic and bacterial viruses. *Proceedings of the National Academy of Sciences of the United States of America 102*, 18944-18949.

(9) Maaty, W. S. A., Ortmann, A. C., Dlakic, M., Schulstad, K., Hilmer, J. K., Liepold, L., Weidenheft, B., Khayat, R., Douglas, T., Young, M. J., and Bothner, B. (2006) Characterization of the archaeal thermophile Sulfolobus turreted icosahedral virus validates an evolutionary link among double-stranded DNA viruses from all domains of life. *Journal of Virology 80*, 7625-7635.

(10) Rice, G., Stedman, K., Snyder, J., Wiedenheft, B., Willits, D., Brumfield, S., McDermott, T., and Young, M. J. (2001) Viruses from extreme thermal environments. *Proc. Natl. Acad. Sci. USA 98*, 13341-13345.

(11) 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 Research 29*, 22-28.

(12) Hempstead, P. D., Hudson, A. J., Artymiuk, P. J., Andrews, S. C., Banfield, M. J., Guest, J. R., and Harrison, P. M. (1994) Direct Observation of the Iron-Binding Sites in a Ferritin. *Febs Letters 350*, 258-262.

(13) Stillman, T. J., Hempstead, P. D., Artymiuk, P. J., Andrews, S. C., Hudson, A. J., Treffry, A., Guest, J. R., and Harrison, P. M. (2001) The high-resolution X-ray crystallographic structure of the ferritin (EcFtnA) of Escherichia coli; Comparison with human H ferritin (HuHF) and the structures of the Fe3+ and Zn2+ derivatives. *Journal of Molecular Biology 307*, 587-603.

(14) Dautant, A., Meyer, J. B., Yariv, J., Precigoux, G., Sweet, R. M., Kalb, A. J., and Frolow, F. (1998) Structure of a monoclinic crystal form of cytochrome b1 (bacterioferritin) from E-coli. *Acta Crystallographica Section D-Biological Crystallography 54*, 16-24.

(15) Frolow, F., Kalb, A. J., and Yariv, J. (1994) Structure of a Unique Twofold Symmetrical Heme-Binding Site. *Nature Structural Biology 1*, 453-460.

(16) Barynin, V. V., Whittaker, M. M., Antonyuk, S. V., Lamzin, V. S., Harrison, P. M., Artymiuk, P. J., and Whittaker, J. W. (2001) Crystal structure of manganese catalase from Lactobacillus plantarum. *Structure 9*, 725-738.

(17) Gauss, G. H., Benas, P., Wiedenheft, B., Young, M., Douglas, T., and Lawrence, M. (2006) The Dps-like protein from Sulfolobus solfataricus combines properties from two distinct protein sub-classes into a single protein architecture: a bacterioferritin-like di-iron binding site within a Dps-like dodecameric assembly. *Biochemistry*.

(18) Wiedenheft, B., Stedman, K., Roberto, F., Willits, D., Gleske, A. K., Zoeller, L., Snyder, J., Douglas, T., and Young, M. (2004) Comparative genomic analysis of hyperthermophilic archaeal Fuselloviridae viruses. *Journal of Virology 78*, 1954-1961.

(19) Kraft, P., Kummel, D., Oeckinghaus, A., Gauss, G. H., Wiedenheft, B., Young, M., and Lawrence, C. M. (2004) Structure of D-63 from sulfolobus spindle-shaped virus 1: Surface properties of the dimeric four-helix bundle suggest an adaptor protein function. *Journal of Virology 78*, 7438-7442.

(20) Kraft, P., Oeckinghaus, A., Kummel, D., Gauss, G. H., Gilmore, J., Wiedenheft, B., Young, M., and Lawrence, C. M. (2004) Crystal structure of F-93 from Sulfolobus spindle-shaped virus 1, a winged-helix DNA binding protein. *Journal of Virology 78*, 11544-11550.

(21) Ortmann, A., Wiedenheft, B., Douglas, T., and Young, M. (2006) Hot Archaeal Viruses Reveal Deep Connections *Nature Reviews Microbiology 4*, 520-528.

(22) Snyder, J. C., Spuhler, J., Wiedenheft, B., Roberto, F. F., Douglas, T., and Young, M. J. (2004) Effects of culturing on the population structure of a hyperthermophilic virus. *Microbial Ecology 48*, 561-566.

(23) Snyder, J. C., Wiedenheft, B., Lavin, M., Roberto, F. F., Spuhler, J., Douglas, T., and Young, M. (2006) Linking phylogeny and ecology of viruses in Yellowstone hot springs reveals high rates of migration and an intercontinental metacommunity. *Submitted*.

(24) Young, M., Wiedenheft, B., Snyder, J., Spuhler, J., Roberto, F., and Douglas, T. (2005) *Geothermal Biology and Geochemsitry in Yellowstone National Park.*, Montana State University Extension Press.

(25) Ramsay, B., Wiedenheft, B., Allen, M., Gauss, G. H., Lawrence, C. M., Young, M., and Douglas, T. (2006) Dps-like protein from the hyperthermophilic archaeon Pyrococcus furiosus. *Journal of Biological Inorganic Chemistry 100*, 1061–1068.

(26) Balch, W. E., Magrum, L. J., Fox, G. E., Wolfe, R. S., and Woese, C. R. (1977) Ancient Divergence among Bacteria. *Journal of Molecular Evolution 9*, 305-311.

(27) Fox, G. E., Stackebrandt, E., Hespell, R. B., Gibson, J., Maniloff, J., Dyer, T. A., Wolfe, R. S., Balch, W. E., Tanner, R. S., Magrum, L. J., Zablen, L. B., Blakemore, R., Gupta, R., Bonen, L., Lewis, B. J., Stahl, D. A., Luehrsen, K. R., Chen, K. N., and Woese, C. R. (1980) The Phylogeny of Prokaryotes. *Science 209*, 457-463.

(28) Woese, C. R. (1987) Bacterial Evolution. *Microbiological Reviews 51*, 221-271.

(29) Woese, C. R., and Fox, G. E. (1977) Phylogenetic Structure of Prokaryotic Domain - Primary Kingdoms. *Proceedings of the National Academy of Sciences of the United States of America 74*, 5088-5090.

(30) Woese, C. R., Kandler, O., and Wheelis, M. L. (1990) Towards a natural system of organisms: Proposal for the domains Archaea, Bacteria, and Eucarya. *Proceedings of the National Academy of Science USA 87*, 4576-4579.

(31) Ochsenreiter, T., Pfeifer, F., and Schleper, C. (2002) Diversity of Archaea in hypersaline environments characterized by molecular-phylogenetic and cultivation studies. *Extremophiles 6*, 267-274.

(32) Barns, S. M., Fundyga, R. E., Jeffries, M. W., and Pace, N. R. (1994) Remarkable archaeal diversity detected in a Yellowstone National Park hot spring environment. *Proc Natl Acad Sci U S A 91*, 1609-13.

(33) Reysenbach, A. L., Ehringer, H., and Hershberger, K. (2000) Microbial diversity at 83 degrees C in Calcite Springs, Yellowstone National Park: another environment where the Aquificales and "Korarchaeota" coexist. *Extremophiles 4*, 61-67.

(34) Hershberger, K. L., Barns, S. M., Reysenbach, A. L., Dawson, S. C., and Pace, N. R. (1996) Wide diversity of Crenarchaeota. *Nature 384*, 420-420.

(35) Futterer, O., Angelov, A., Liesegang, H., Gottschalk, G., Schleper, C., Schepers, B., Dock, C., Antranikian, G., and Liebl, W. (2004) Genome sequence of Picrophilus torridus and its implications for life around pH 0. *Proceedings of the National Academy of Sciences of the United States of America 101*, 9091-9096.

(36) Bintrim, S. B., Donohue, T. J., Handelsman, J., Roberts, G. P., and Goodman, R. M. (1997) Molecular phylogeny of archaea from soil. *Proceedings of the National Academy of Sciences of the United States of America 94*, 277-282.

(37) Buckley, D. H., Graber, J. R., and Schmidt, T. M. (1998) Phylogenetic analysis of nonthermophilic members of the kingdom Crenarchaeota and their diversity and abundance in soils. *Applied and Environmental Microbiology 64*, 4333-4339.

(38) Jurgens, G., Lindstrom, K., and Saano, A. (1997) Novel group within the kingdom Crenarchaeota from boreal forest soil. *Applied and Environmental Microbiology 63*, 803-805.

(39) Ochsenreiter, T., Selezi, D., Quaiser, A., Bonch-Osmolovskaya, L., and Schleper, C. (2003) Diversity and abundance of Crenarchaeota in terrestrial habitats studied by 16S RNA surveys and real time PCR. *Environmental Microbiology 5*, 787-797.

(40) Lepp, P. W., Brinig, M. M., Ouverney, C. C., Palm, K., Armitage, G. C., and Relman, D. A. (2004) Methanogenic Archaea and human periodontal disease. *Proceedings of the National Academy of Sciences of the United States of America 101*, 6176-6181.

(41) Church, M. J., DeLong, E. F., Ducklow, H. W., Karner, M. B., Preston, C. M., and Karl, D. M. (2003) Abundance and distribution of planktonic Archaea and Bacteria in the waters west of the Antarctic Peninsula. *Limnology and Oceanography 48*, 1893-1902.

(42) Karner, M. B., DeLong, E. F., and Karl, D. M. (2001) Archaeal dominance in the mesopelagic zone of the Pacific Ocean. *Nature 409*, 507-510.

(43) Schleper, C., Jurgens, G., and Jonuscheit, M. (2005) Genomic studies of uncultivated archaea. *Nature Reviews Microbiology 3*, 479-488.

(44) Tyson, G. W., Chapman, J., Hugenholtz, P., Allen, E. E., Ram, R. J., Richardson, P. M., Solovyev, V. V., Rubin, E. M., Rokhsar, D. S., and Banfield, J. F. (2004) Community structure and metabolism through reconstruction of microbial genomes from the environment. *Nature 428*, 37-43.

(45) Venter, J. C., Remington, K., Heidelberg, J. F., Halpern, A. L., Rusch, D., Eisen, J. A., Wu, D. Y., Paulsen, I., Nelson, K. E., Nelson, W., Fouts, D. E., Levy, S., Knap, A. H., Lomas, M. W., Nealson, K., White, O., Peterson, J., Hoffman, J., Parsons, R., Baden-Tillson, H., Pfannkoch, C., Rogers, Y. H., and Smith, H. O. (2004) Environmental genome shotgun sequencing of the Sargasso Sea. *Science 304*, 66-74.

(46) Barns, S. M., Delwiche, C. F., Palmer, J. D., and Pace, N. R. (1996) Perspectives on archaeal diversity, thermophily, and monophyly from environmental rRNA sequences. *Proceedings of the National Academy of Science USA 93*, 9188-9193.

(47) Burggraf, S., Heyder, P., and Eis, N. (1997) A pivotal Archaea group. *Nature 385*, 780-780.

(48) K. O. Stetter, J. G. E., M. Keller, and the JGI-DOE [http://www.jgi.doe.gov/sequencing/why/CSP2006/korarchaeota.html].

(49) Robertson, C. E., Harris, J. K., Spear, J. R., and Pace, N. R. (2005) Phylogenetic diversity and ecology of environmental Archaea. *Current Opinion in Microbiology 8*, 638-642.

(50) Woese, C. R., and Olsen, G. J. (1986) Archaebacterial phylogeny: perspectives on the urkingdoms. *Systematic and Applied Microbiology 7*, 161-177.

(51) 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.

(52) ATCC (http://www.atcc.org/)

(53) DSMZ (http://www.atcc.org/)

(54) Allers, T., and Mevarech, M. (2005) Archaeal genetics - The third way. *Nature Reviews Genetics 6*, 58-73.

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

(56) Makarova, K. S., and Koonin, E. V. (2003) Comparative genomics of archaea: how much have we learned in six years, and what's next? *Genome Biology 4*, -.

(57) Makarova, K. S., and Koonin, E. V. (2005) Evolutionary and functional genomics of the Archaea. *Current Opinion in Microbiology 8*, 586-594.

(58) Snel, B., Lehmann, G., Bork, P., and Huynen, M. A. (2000) STRING: a web-server to retrieve and display the repeatedly occurring neighbourhood of a gene. *Nucleic Acids Research 28*, 3442-3444.

(59) von Mering, C., Huynen, M., Jaeggi, D., Schmidt, S., Bork, P., and Snel, B. (2003) STRING: a database of predicted functional associations between proteins. *Nucleic Acids Research 31*, 258-261.

(60) Koonin, E. V., Mushegian, A. R., and Bork, P. (1996) Non-orthologous gene displacement. *Trends in Genetics 12*, 334-336.

(61) Cline, S. W., and Doolittle, W. F. (1987) Efficient Transfection of the Archaebacterium Halobacterium-Halobium. *Journal of Bacteriology 169*, 1341-1344.

(62) Gernhardt, P., Possot, O., Foglino, M., Sibold, L., and Klein, A. (1990) Construction of an Integration Vector for Use in the Archaebacterium Methanococcus-Voltae and Expression of a Eubacterial Resistance Gene. *Molecular & General Genetics 221*, 273-279.

(63) Klein, A., and Horner, K. (1995) (Robb, F., Place, A., Sowers, K., HJ, S., DasSarma, S., and EM, F., Eds.), Cold Spring Harbor Laboratory Press.

(64) Bardy, S., Ng, S. Y. M., and Jarrell, K. F. (2004) Recent advances in the structure and assembly of the archaeal flagellum. *Journal of Molecular Microbiology and Biotechnology 7*, 41-51.

(65) Heinicke, I., Muller, J., Pittelkow, M., and Klein, A. (2004) Mutational analysis of genes encoding chromatin proteins in the archaeon Methanococcus voltae indicates their

involvement in the regulation of gene expression. *Molecular Genetics and Genomics 272*, 76-87.

(66) Long, S. W., and Faguy, D. M. (2004) Anucleate and titan cell phenotypes caused by insertional inactivation of the structural maintenance of chromosomes (smc) gene in the archaeon Methanococcus voltae. *Molecular Microbiology 52*, 1567-1577.

(67) Pritchett, M. A., and Metcalf, W. W. (2005) Genetic, physiological and biochemical characterization of multiple methanol methyltransferase isozymes in Methanosarcina acetivorans C2A. *Molecular Microbiology 56*, 1183-1194.

(68) Welander, P. V., and Metcalf, W. W. (2005) Loss of the mtr operon in Methanosarcina blocks growth on methanol, but not methanogenesis, and reveals an unknown methanogenic pathway. *Proceedings of the National Academy of Sciences of the United States of America 102*, 10664-10669.

(69) Pritchett, M. A., Zhang, J. K., and Metcalf, W. W. (2004) Development of a markerless genetic exchange method for Methanosarcina acetivorans C2A and its use in construction of new genetic tools for methanogenic archaea. *Applied and Environmental Microbiology 70*, 1425-1433.

(70) Stetter, K. O. (1996) Hyperthermophilic procaryotes. *Fems Microbiology Reviews 18*, 149-158.

(71) Adams, M. W. W., and Kelly, R. M. (1998) Finding and using hyperthermophilic enzymes. *Trends in Biotechnology 16*, 329-332.

(72) Vieille, C., and Zeikus, G. J. (2001) Hyperthermophilic enzymes: Sources, uses, and molecular mechanisms for thermostability. *Microbiology and Molecular Biology Reviews 65*, 1-+.

(73) Fukui, T., Atomi, H., Kanai, T., Matsumi, R., Fujiwara, S., and Imanaka, T. (2005) Complete genome sequence of the hyperthermophilic archaeon Thermococcus kodakaraensis KOD1 and comparison with Pyrococcus genomes. *Genome Research 15*, 352-363.

(74) Sato, T., Fukui, T., Atomi, H., and Imanaka, T. (2005) Improved and versatile transformation system allowing multiple genetic manipulations of the hyperthermophilic archaeon Thermococcus kodakaraensis. *Applied and Environmental Microbiology 71*, 3889-3899.

(75) Sato, T., Fukui, T., Atomi, H., and Imanaka, T. (2003) Targeted gene disruption by homologous recombination in the hyperthermophilic archaeon Thermococcus kodakaraensis KOD1. *Journal of Bacteriology 185*, 210-220.

(76) Brock, T. D., Brock, K. M., Belly, R. T., and Weiss, R. L. (1972) *Sulfolobus*: A New Genus of Sulfur-Oxidizing Bacteria Living at Low pH and High Temperature. *Archives in Microbiology 84*, 54-68.

(77) Grogan, D. W. (1989) Phenotypic Characterization of the Archaebacterial Genus Sulfolobus - Comparison of 5 Wild-Type Strains. *Journal of Bacteriology 171*, 6710-6719.

(78) She, Q., Singh, R. K., Confalonieri, F., Zivanovic, Y., Allard, G., Awayez, M. J., Chan-Weiher, C. C. Y., Clausen, I. G., Curtis, B. A., De Moors, A., Erauso, G., Fletcher, C., Gordon, P. M. K., 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. *Proceedings of the National Academy of Sciences of the United States of America 98*, 7835-7840.

(79) Kawarabayasi, Y., Hino, Y., Horikawa, H., Jin-no, Takahashi, Sekine, Baba, Ankai, Kosugi, Hosoyama, Fukui, Nagai, Nishijima, Otsuka, Nakazawa, Takamiya, Kato, Yoshizawa, Tanaka, Kudoh, Yamazaki, J., Kushida, Oguchi, Aoki, Masuda, Yanagii, Nishimura, Yamagishi, Oshima, and Kikuchi. (2001) Complete genome sequence of an aerobic thermoacidophilic crenarchaeon *Sulfolobus tokodaii* strain 7. *DNA Research 8*, 123-140.

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

(81) Snijders, A. P. L., Walther, J., Peter, S., Kinnman, I., de Vos, M. G. J., van de Werken, H. J. G., Brouns, S. J. J., van der Oost, J., and Wright, P. C. (2006) Reconstruction of central carbon metabolism in Sulfolobus solfataricus using a two-dimensional gel electrophoresis map, stable isotope labelling and DNA microarray analysis. *Proteomics 6*, 1518-1529.

(82) Palm, P., Schleper, C., Grampp, B., Yeats, S., McWilliam, P., Reiter, W.-D., and Zillig, W. (1991) Complete nucleotide sequence of the virus SSV1 of the archaebacterium *Sulfolobus shibatae*. *Virology 185*, 242-250.

(83) Aravalli, R. N., and Garrett, R. A. (1997) Shuttle vectors for hyperthermophilic archaea. *Extremophiles 1*, 183-191.

(84) Brouns, S. J. J., Wu, H., Akerboom, J., Turnbull, A. P., de Vos, W. M., and van der Oost, J. (2005) Engineering a selectable marker for hyperthermophiles. *Journal of Biological Chemistry 280*, 11422-11431.

(85) Cannio, R., Contursi, P., Rossi, M., and Bartolucci, S. (1998) An autonomously replicating transforming vector for *Sulfolobus solfataricus*. *J. Bacteriol. 180*, 3237-3240.

(86) Jonuscheit, M., Martusewitsch, E., Stedman, K. M., and Schleper, C. (2003) A reporter gene system for the hyperthermophilic archaeon *Sulfolobus solfataricus* based on a selectable and integrative shuttle vector. *Mol. Microbiol. 48*, 1241-1252.

(87) Albers, S. V., Jonuscheit, M., Dinkelaker, S., Urich, T., Kletzin, A., Tampe, R., Driessen, A. J. M., and Schleper, C. (2006) Production of recombinant and tagged proteins in the hyperthermophilic archaeon Sulfolobus solfataricus. *Applied and Environmental Microbiology 72*, 102-111.

(88) Dixit, V., Bini, E., Drozda, M., and Blum, P. (2004) Mercury inactivates transcription and the generalized transcription factor TFB in the Archaeon Sulfolobus solfataricus. *Antimicrobial Agents and Chemotherapy 48*, 1993-1999.

(89) Worthington, P., Hoang, V., Perez-Pomares, F., and Blum, P. (2003) Targeted disruption of the alpha-amylase gene in the hyperthermophilic archaeon Sulfolobus solfataficus. *Journal of Bacteriology 185*, 482-488.

(90) Miller, T. L., Wolin, M. J., Demacario, E. C., and Macario, A. J. L. (1982) Isolation of Methanobrevibacter-Smithii from Human Feces. *Applied and Environmental Microbiology 43*, 227-232.

(91) Belay, N., Mukhopadhyay, B., Demacario, E. C., Galask, R., and Daniels, L. (1990) Methanogenic Bacteria in Human Vaginal Samples. *Journal of Clinical Microbiology 28*, 1666-1668.

(92) Belay, N., Johnson, R., Rajagopal, B. S., Demacario, E. C., and Daniels, L. (1988) Methanogenic Bacteria from Human Dental Plaque. *Applied and Environmental Microbiology 54*, 600-603.

(93) Johnson, R., Belay, N., Harvey, B., Daniels, L., and Demacario, E. C. (1986) Presence of Methanogenic Bacteria in Human Subgingival Dental Plaque. *Journal of Dental Research 65*, 348-348.

(94) Saunders, N. F. W., Curmi, P. M. G., and Cavicchioli, R. (2004) An online database for the detection of novel archaeal sequences in human ESTs. *Bioinformatics 20*, 2361-2362.

(95) Krishnan, L., Sad, S., Patel, G. B., and Sprott, G. D. (2001) The potent adjuvant activity of archaeosomes correlates to the recruitment and activation of macrophages and dendritic cells in vivo. *Journal of Immunology 166*, 1885-1893.

(96) Waters, E., Hohn, M. J., Ahel, I., Graham, D. E., Adams, M. D., Barnstead, M., Beeson, K. Y., Bibbs, L., Bolanos, R., Keller, M., Kretz, K., Lin, X. Y., Mathur, E., Ni, J. W., Podar, M., Richardson, T., Sutton, G. G., Simon, M., Soll, D., Stetter, K. O., Short, J. M., and Noordewier, M. (2003) The genome of Nanoarchaeum equitans: Insights into early archaeal evolution and derived parasitism. *Proceedings of the National Academy of Sciences of the United States of America 100*, 12984-12988.

(97) Snel, B., Bork, P., and Huynen, M. (2000) Genome evolution - gene fusion versus gene fission. *Trends in Genetics 16*, 9-11.

(98) Randau, L., Munch, R., Hohn, M. J., Jahn, D., and Soll, D. (2005) Nanoarchaeum equitans creates functional tRNAs from separate genes for their 5 '- and 3 '-halves. *Nature 433*, 537-541.

(99) Randau, L., Pearson, M., and Soll, D. (2005) The complete set of tRNA species in Nanoarchaeum equitans. *Febs Letters 579*, 2945-2947.

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

(101) Forterre, P. (2002) A hot story from comparative genomics: reverse gyrase is the only hyperthermophile-specific protein. *Trends in Genetics 18*, 236-238.

(102) Meseguer, I., and Rodriguezvalera, F. (1985) Production and Purification of Halocin H-4. *Fems Microbiology Letters 28*, 177-182.

(103) Prangishvili, D., Holz, I., Stieger, E., Nickell, S., Kristjansson, J. K., and Zillig, W. (2000) Sulfolobicins, specific proteinaceous toxins produced by strains of the extremely thermophilic archaeal genus *Sulfolobus*. *Journal of Bacteriology 182*, 2985-2988.

(104) Meseguer, I., and Rodriguezvalera, F. (1986) Effect of Halocin H-4 on Cells of Halobacterium-Halobium. *Journal of General Microbiology 132*, 3061-3068.

(105) Rdest, U., Sturm, M., and Goebel, W. (1987) Bacteriocins from Halobacteria. *Journal of Cellular Biochemistry*, 183-183.

(106) Torreblanca, M., Meseguer, I., and Rodriguezvalera, F. (1989) Halocin H6, a Bacteriocin from Haloferax-Gibbonsii. *Journal of General Microbiology 135*, 2655-2661.

(107) Torsvik, T., and Dundas, I. D. (1974) Bacteriophage of Halobacterium-Salinarium. *Nature 248*, 680-681.

(108) Gropp, F., Palm, P., and Zillig, W. (1989) Expression and regulation of Halobacterium halobium phage phi H genes. *Can J Microbiol 35*, 182-8.

(109) Schnabel, H., and Zillig, W. (1982) Structural Variations in the DNA of Halobacterium-Halobium Phage-Phi-H. *Zentralblatt Fur Bakteriologie Mikrobiologie Und Hygiene Series a-Medical Microbiology Infectious Diseases Virology Parasitology 253*, 35-36.

(110) Schnabel, H., Zillig, W., Pfaffle, M., Schnabel, R., Michel, H., and Delius, H. (1982) Halobacterium-Halobium Phage Phi-H. *Embo Journal 1*, 87-92.

(111) Stolt, P., Grampp, B., and Zillig, W. (1994) Genes for DNA cytosine methyltransferases and structural proteins, expressed during lytic growth by the phage phi H of the archaebacterium Halobacterium salinarium. *Biol Chem Hoppe Seyler 375*, 747-57.

(112) Stolt, P., and Zillig, W. (1993) Antisense RNA mediates transcriptional processing in an archaebacterium, indicating a novel kind of RNase activity. *Mol Microbiol 7*, 875-82.

(113) Oren, A., Bratbak, G., and Heldal, M. (1997) Occurrence of virus-like particles in the Dead Sea. *Extremophiles 1*, 143-149.

(114) Bath, C., and Dyall-Smith, M. L. (1998) His1, an archaeal virus of the Fuselloviridae family that infects *Haloarcula hispanica*. *J. Virology 72*, 9392-9395.

(115) Reiter, W. D., Palm, P., and Zillig, W. (1988) Analysis of transcription in the archaebacterium Sulfolobus indicates that archaebacterial promoters are homologous to eukaryotic pol II promoters. *Nucleic Acids Res 16*, 1-19.

(116) Huet, J., Schnabel, R., Sentenac, A., and Zillig, W. (1983) Archaebacteria and eukaryotes possess DNA-dependent RNA polymerases of a common type. *Embo J 2*, 1291-4.

(117) Prangishvili, D., and Garrett, R. A. (2005) Viruses of hyperthermophilic Crenarchaea. *Trends in Microbiology 13*, 535-542.

(118) Geslin, C., Le Romancer, M., Erauso, G., Gaillard, M., Perrot, G., and Prieur, D. (2003) PAV1, the first virus-like particle isolated from a hyperthermophilic euryarchaeote, "Pyrococcus abyssi". *Journal of Bacteriology 185*, 3888-3894.

(119) Prangishvili, D., Arnold, H. P., Gotz, D., Ziese, U., Holz, I., Kristjansson, J. K., and Zillig, W. (1999) A novel virus family, the *Rudiviridae*: Structure, virus-host interactions and genome variability of the *Sulfolobus* viruses SIRV1 and SIRV2. *Genetics 152*, 1387-1396.

(120) Arnold, H. P., Ziese, U., and Zillig, W. (2000) SNDV, a novel virus of the extremely thermophilic and acidophilic archaeon *Sulfolobus*. *Virology 272*, 409-416.

(121) Haring, M., Rachel, R., Peng, X., Garrett, R. A., and Prangishvili, D. (2005) Viral diversity in hot springs of Pozzuoli, Italy, and characterization of a unique archaeal virus, acidianus bottle-shaped virus, from a new family, the Ampullaviridae. *Journal of Virology 79*, 9904-9911.

(122) Prangishvili, D., and Garrett, R. A. (2004) Exceptionally diverse morphotypes and genomes of crenarchaeal hyperthermophilic viruses. *Biochemical Society Transactions 32*, 204-208.

(123) Snyder, J. C., Stedman, K., Rice, G., Wiedenheft, B., Spuhler, J., and Young, M. J. (2003) Viruses of hyperthermophilic Archaea. *Research in Microbiology 154*, 474-482.

(124) Haring, M., Vestergaard, G., Rachel, R., Chen, L. M., Garrett, R. A., and Prangishvili, D. (2005) Independent virus development outside a host. *Nature 436*, 1101-1102.

(125) Martin, A., Yeats, S., Janekovic, D., Reiter, W.-D., Aicher, W., and Zillig, W. (1984) SAV1, a temperate U.V. inducible DNA virus like particle from the archaebacterium *Sulfolobus acidocaldarius* isolate B12. *EMBO J. 3*, 2165-2168.

(126) Rice, G., Tang, L., Stedman, K., Roberto, F., Spuhler, J., Gillitzer, E., Johnson, J. E., Douglas, T., and Young, M. (2004) The structure of a thermophilic archaeal virus shows a double-stranded DNA viral capsid type that spans all domains of life. *Proceedings of the National Academy of Sciences of the United States of America 101*, 7716-7720.

(127) Stedman, K. M., She, Q., Phan, H., Arnold, H. P., Holz, I., Garrett, R. A., and Zillig, W. (2003) Relationships between fuselloviruses infecting the extremely thermophilic archaeon Sulfolobus: SSV1 and SSV2. *Res. Microbiol. 154*, 295-302.

(128) Xiang, X. Y., Chen, L. M., Huang, X. X., Luo, Y. M., She, Q. X., and Huang, L. (2005) Sulfolobus tengchongensis spindle-shaped virus STSV1: Virus-host interactions and genomic features. *Journal of Virology 79*, 8677-8686.

(129) Wood, A. G., Whitman, W. B., and Konisky, J. (1989) Isolation and characterization of an archaebacterial viruslike particle from *Methanococcus voltae* A3. *J. Bacteriol. 171*, 93-98.

(130) Porter, K., Kukkaro, P., Bamford, J. K. H., Bath, C., Kivela, H. M., Dyall-Smith, M. L., and Bamford, D. H. (2005) SHI: A novel, spherical halovirus isolated from an Australian hypersaline lake. *Virology 335*, 22-33.

(131) Bamford, D. H., Ravantti, J. J., Ronnholm, G., Laurinavicius, S., Kukkaro, P., Dyall-Smith, M., Somerharju, P., Kalkkinen, N., and Bamford, J. K. H. (2005) Constituents of SH1, a novel lipid-containing virus infecting the halophilic euryarchaeon Haloarcula hispanica. *Journal of Virology 79*, 9097-9107.

(132) Baker, M. L., Jiang, W., Rixon, F. J., and Chiu, W. (2005) Common ancestry of herpesviruses and tailed DNA bacteriophages. *Journal of Virology 79*, 14967-14970.

(133) Duda, R. L., Hendrix, R. W., Huang, W. M., and Conway, J. F. (2006) Shared architecture of bacteriophage SPO1 and herpesvirus capsids (vol 16, pg R11, 2006). *Current Biology 16*, 440-440.

(134) Mallick, P., Boutz, D. R., Eisenberg, D., and Yeates, T. O. (2002) Genomic evidence that the intracellular proteins of archaeal microbes contain disulfide bonds. *Proceedings of the National Academy of Sciences of the United States of America 99*, 9679-9684.

(135) Haring, M., Peng, X., Brugger, K., Rachel, R., Stetter, K. O., Garrett, R. A., and Prangishvili, D. (2004) Morphology and genome organization of the virus PSV of the hyperthermophilic archaeal genera Pyrobaculum and Thermoproteus: a novel virus family, the Globuloviridae. *Virology 323*, 233-242.

(136) Peng, X., Blum, H., She, Q. X., Mallok, S., Brugger, K., Garrett, R. A., Zillig, W., and Prangishvili, D. (2001) Sequences and replication of genomes of the archaeal rudiviruses SIRV1 and SIRV2: Relationships to the archaeal lipothrixvirus SIFV and some eukaryal viruses. *Virology 291*, 226-234.

(137) Hendrix, R. W., Smith, M. C. M., Burns, R. N., Ford, M. E., and Hatfull, G. F. (1999) Evolutionary relationships among diverse bacteriophages and prophages: All the world's a phage. *Proceedings of the National Academy of Sciences of the United States of America 96*, 2192-2197.

(138) Bettstetter, M., Peng, X., Garrett, R. A., and Prangishvili, D. (2003) AFV1, a novel virus infecting hyperthermophilic archaea of the genus Acidianus. *Virology 315*, 68-79.

(139) Haring, M., Vestergaard, G., Brugger, K., Rachel, R., Garrett, R. A., and Prangishvili, D. (2005) Structure and genome organization of AFV2, a novel archaeal lipothrixvirus with unusual terminal and core structures. *Journal of Bacteriology 187*, 3855-3858.

(140) Vestergaard, G., Haring, M., Peng, X., Rachel, R., Garrett, R. A., and Prangishvili, D. (2005) A novel rudivirus, ARV1, of the hyperthermophilic archaeal genus Acidianus. *Virology 336*, 83-92.

(141) Arnold, H. P., Zillig, W., Ziese, U., Holz, I., Crosby, M., Utterback, T., Weidmann, J. F., Kristjanson, J. K., Klenk, H. P., Nelson, K. E., and Fraser, C. M. (2000) A novel

lipothrixvirus, SIFV, of the extremely thermophilic crenarchaeon *Sulfolobus*. *Virology 267*, 252-266.

(142) Peng, X., Blum, H., She, Q., Mallok, S., Brugger, K., Garrett, R. A., Zillig, W., and Prangishvili, D. (2001) Sequences and replication of genomes of the archaeal rudiviruses SIRV1 and SIRV2: relationships to the archaeal lipothrixvirus SIFV and some eukaryal viruses. *Virology 291*, 226-234.

(143) Peng, X., Kessler, A., Phan, H., Garrett, R. A., and Prangishvili, D. (2004) Multiple variants of the archaeal DNA rudivirus SIRV1 in a single host and a novel mechanism of genomic variation. *Molecular Microbiology 54*, 366-375.

(144) Larson, E. T., Reiter, D., Young, M., and Lawrence, C. M. (2006) Structure of A197 from Sulfolobus turreted icosahedral virus: a crenarchaeal viral glycosyltransferase exhibiting the GT-A fold. *Journal of Virology 80*, 7636-7644.

(145) Koonin, E. V. (1992) Archaebacterial Virus Ssv1 Encodes a Putative Dnaa-Like Protein. *Nucleic Acids Research 20*, 1143-1143.

(146) Grogan, D. W. (1998) Hyperthermophiles and the problem of DNA instability. *Molecular Microbiology 28*, 1043-1049.

(147) Muskhelishvili, G. (1994) The archaeal SSV integrase promotes intermolecular excisive recombination in-vitro. *Syst. Appl. Microbiol. 16*, 605-608.

(148) Muskhelishvili, G., Palm, P., and Zillig, W. (1993) SSV1-encoded site-specific recombination system in *Sulfolobus shibatae*. *Mol. Gen. Genet. 237*, 334-342.

(149) Reiter, W.-D., Palm, P., Yeats, S., and Zillig, W. (1987) Gene expression in archaebacteria: Physical mapping of constitutive and UV-inducible transcripts from the *Sulfolobus* virus-like particle SSV1. *Mol. Gen. Genet. 209*, 270-275.

(150) Wu, R. Y., Zhang, R. G., Zagnitko, O., Dementieva, I., Maltzev, N., Watson, J. D., Laskowski, R., Gornicki, P., and Joachimiak, A. (2003) Crystal structure of Enterococcus faecalis SlyA-like transcriptional factor. *Journal of Biological Chemistry 278*, 20240-20244.

(151) Alekshun, M. N., Levy, S. B., Mealy, T. R., Seaton, B. A., and Head, J. F. (2001) The crystal structure of MarR, a regulator of multiple antibiotic resistance, at 2.3 angstrom resolution. *Nature Structural Biology 8*, 710-714.

(152) She, Q., Chen, B., and Chen, L. (2004) Archaeal integrases and mechanisms of gene capture. *Biochemical Society Transactions 32*, 222-226.

(153) Yeats, S., McWilliam, P., and Zillig, W. (1982) A plasmid in the archaebacterium *Sulfolobus acidocaldarius*. *EMBO J. 1*, 1035-1038.

(154) She, Q., Peng, X., Zillig, W., and Garrett, R. A. (2001) Gene capture in archaeal chromosomes. *Nature 409*, 478.

(155) Swalla, B. M., Gumport, R. I., and Gardner, J. F. (2003) Conservation of structure and function among tyrosine recombinases: homology-based modeling of the lambda integrase core-binding domain. *Nucleic Acids Research 31*, 805-818.

(156) Kessler, A., Brinkman, A. B., van der Oost, J., and Prangishvili, D. (2004) Transcription of the rod-shaped viruses SIRV1 and SIRV2 of the hyperthermophilic archaeon Sulfolobus. *Journal of Bacteriology 186*, 7745-7753.

(157) Schleper, C., Kubo, K., and Zillig, W. (1992) The particle SSV1 from the extremely thermophilic archaeon *Sulfolobus* is a virus: demonstration of infectivity and of transfection with viral DNA. *Proc. Natl. Acad. Sci. USA 89*, 7645-7649.

(158) Liu, D. X., and Huang, L. (2002) Induction of the Sulfolobus shibatae virus SSV1 DNA replication by mitomycin C. *Chinese Science Bulletin 47*, 923-927.

(159) Zillig, W., Kletzin, A., Schleper, C., Holz, I., Janekovic, D., Hain, J., Lanzendorfer, M., and Kristjansson, J. K. (1994) Screening for Sulfolobales, Their Plasmids and Their Viruses in Icelandic Solfataras. *Systematic and Applied Microbiology 16*, 609-628.

(160) Blum, H., Zillig, W., Mallok, S., Domdey, H., and Prangishvili, D. (2001) The genome of the archaeal virus SIRV1 has features in common with genomes of eukaryal viruses. *Virology 281*, 6-9.

(161) Ackermann, H. W. (2003) Bacteriophage observations and evolution. *Research in Microbiology 154*, 245-251.

(162) Whitaker, R. J., Grogan, D. W., and Taylor, J. W. (2003) Geographic barriers isolate endemic populations of hyperthermophilic archaea. *Science 301*, 976-978.

(163) Snyder, J. C. (2005) in *Dept. Microbiology*, Montana State University Bozeman.

(164) Suttle, C. A. (2005) Viruses in the sea. *Nature 437*, 356-361.

(165) Zillig, W., Arnold, H. P., Holz, I., Prangishvili, D., Schweier, A., Stedman, K., She, Q., Phan, H., Garrett, R., and Kristjansson, J. K. (1998) Genetic elements in the extremely thermophilic archaeon *Sulfolobus*. *Extremophiles 2*, 131-140.

(166) Grogan, D., Palm, P., and Zillig, W. (1990) Isolate B12, which harbours a virus-like element, represents a new species of the archaebacterial genus *Sulfolobus*, *Sulfolobus shibatae*, sp. nov. *Arch. Microbiol. 154*, 594-599.

(167) Zillig, W., Stetter, K. O., Wunderl, S., Schulz, W., Preiss, H., and Scholz, H. (1980) The Sulfolobus-"Caldariella" Group: Taxonomy on the Basis of the Structure of DNA-Dependent RNA Polymerases. *Arch. Microbiol. 125*, 259-269.

(168) Nadal, M., Mirambeau, G., Forterre, P., Reiter, W.-D., and Duguet, M. (1986) Positively supercoiled DNA in a virus-like particle of an archaebacterium. *Nature 321*, 256-258.

(169) Reiter, W. D., Palm, P., and Yeats, S. (1989) Transfer RNA genes frequently serve as integration sites for prokaryotic genetic elements. *Nucleic Acids Res. 17*, 1907-1914.

(170) Reiter, W.-D., and Palm, P. (1990) Identification and characterization of a defective SSV1 genome integrated into a tRNA gene in the archaebacterium *Sulfolobus* sp. B12. *Mol. Gen. Genet. 221*, 65-71.

(171) Reiter, W.-D., Palm, P., Henschen, A., Lottspeich, F., Zillig, W., and Grampp, B. (1987) Identification and characterization of the genes encoding three structural proteins of the *Sulfolobus* virus-like particle SSV1. *Mol. Gen. Genet. 206*, 144-153.

(172) Serre, M. C., Letzelter, C., Garel, J. R., and Duguet, M. (2002) Cleavage properties of an archaeal site-specific recombinase, the SSV1 integrase. *J. Biol. Chem. 277*, 16758-16767.

(173) Stedman, K. M., Schleper, C., Rumpf, E., and Zillig, W. (1999) Genetic requirements for the function of the archaeal virus SSV1 in Sulfolobus solfataricus: Construction and testing of viral shuttle vectors. *Genetics 152*, 1397-1405.

(174) Arnold, H. P., She, Q., Phan, H., Stedman, K., Prangishvili, D., Holz, I., Kristjansson, J. K., Garrett, R., and Zillig, W. (1999) The genetic element pSSVx of the extremely thermophilic crenarchaeon Sulfolobus is a hybrid between a plasmid and a virus. *Mol. Microbiol. 34*, 217-226.

(175) Kletzin, A., Lieke, A., Urich, T., Charlebois, R. L., and Sensen, C. W. (1999) Molecular analysis of pDL10 from *Acidianus ambivalens* reveals a family of related plasmids from extremely thermophilic and acidophilic archaea. *Genetics 152*, 1307-1314.

(176) Keeling, P. J., Klenk, H. P., Singh, R. K., Feeley, O., Schleper, C., Zillig, W., Doolittle, W. F., and Sensen, C. W. (1996) Complete nucleotide sequence of the *Sulfolobus islandicus* multicopy plasmid pRN1. *Plasmid 35*, 141-144.

(177) Argos, P., Landy, A., Abremski, K., Egan, J. B., Haggard-Ljungquist, E., Hoess, R. H., Kahn, M. L., Kalionis, B., Narayana, S. V. L., III, L. S. P., Sternberg, N., and Leong, J.

M. (1986) The integrase family of site-specific recombinases: regional similarities and global diversity. *EMBO J. 5*, 433-440.

(178) Nunes-Duby, S. E., Kwon, H. J., Tirumalai, R. S., Ellenberger, T., and Landy, A. (1998) Similarities and differences among 105 members of the Int family of site-specific recombinases. *Nucleic Acids Res. 26*, 391-406.

(179) She, Q., Phan, H., Garrett, R. A., Albers, S. V., Stedman, K. M., and Zillig, W. (1998) Genetic profile of pNOB8 from *Sulfolobus*: the first conjugative plasmid from an archaeon. *Extremophiles 2*, 417-425.

(180) Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) *Molecular cloning: a laboratory manual, 2nd ed.*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

(181) Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J. H., Zhang, Z., Miller, W., and Lipman, D. J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research 25*, 3389-3402.

(182) Delcher, A. L., Harmon, D., Kasif, S., White, O., and Salzberg, S. L. (1999) Improved microbial gene identification with GLIMMER. *Nucleic Acids Res. 27*, 4636-4641.

(183) Suzek, B. E., Ermolaeva, M. D., Schreiber, M., and Salzberg, S. L. (2001) A probabilistic method for identifying start codons in bacterial genomes. *Bioinformatics 17*, 1123-1130.

(184) Benson, G. (1999) Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res. 27*, 573-580.

(185) 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.

(186) Jeanmougin, F., Thompson, J. D., Gouy, M., Higgins, D. G., and Gibson, T. J. (1998) Multiple sequence alignment with Clustal x. *Trends in Biochemical Sciences 23*, 403-405.

(187) Swofford. (2002), Sinauer Associates, Sunderlund, MA.

(188) Huelsenbeck, J. P., and Ronquist, F. (2001) MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics 17*, 754-755.

(189) Huelsenbeck, J. P., Ronquist, F., Nielsen, R., and Bollback, J. P. (2001) Bayesian inference of phylogeny and its impact on evolutionary biology. *Science 294*, 2310-2314.

(190) Peng, X., Holz, I., Zillig, W., Garrett, R. A., and She, Q. (2000) Evolution of the family of pRN plasmids and their integrase-mediated insertion into the chromosome of the crenarchaeon *Sulfolobus solfataricus*. *J. Mol. Biol. 303*, 449-454.

(191) Campbell, A. M. (1992) Chromosomal insertion sites for phages and plasmids. *J. Bacteriol. 174*, 7495-7499.

(192) Pedulla, M. L., Ford, M. E., Houtz, J. M., Karthikeyan, T., Wadsworth, C., Lewis, J. A., Jacobs-Sera, D., Falbo, J., Gross, J., Pannunzio, N. R., Brucker, W., Kumar, V., Kandasamy, J., Keenan, L., Bardarov, S., Kriakov, J., Lawrence, J. G., Jacobs, W. R., Jr., Hendrix, R. W., and Hatfull, G. F. (2003) Origins of highly mosaic mycobacteriophage genomes. *Cell 113*, 171-182.

(193) Ghedin, E., Sengamalay, N. A., Shumway, M., Zaborsky, J., Feldblyum, T., Subbu, V., Spiro, D. J., Sitz, J., Koo, H., Bolotov, P., Dernovoy, D., Tatusova, T., Bao, Y. M., St George, K., Taylor, J., Lipman, D. J., Fraser, C. M., Taubenberger, J. K., and Salzberg, S. L. (2005) Large-scale sequencing of human influenza reveals the dynamic nature of viral genome evolution. *Nature 437*, 1162-1166.

(194) Moore, C. B., John, M., James, I. R., Christiansen, F. T., Witt, C. S., and Mallal, S. A. (2002) Evidence of HIV-1 adaptation to HLA-restricted immune responses at a population level. *Science 296*, 1439-1443.

(195) Breitbart, M., Miyake, J. H., and Rohwer, F. (2004) Global distribution of nearly identical phage-encoded DNA sequences. *Fems Microbiology Letters 236*, 249-256.

(196) Breitbart, M., Wegley, L., Leeds, S., Schoenfeld, T., and Rohwer, F. (2004) Phage community dynamics in hot springs. *Applied and Environmental Microbiology 70*, 1633-1640.

(197) Siering, P. L., Clarke, J. M., and Wilson, M. S. (2006) Geochemical and biological diversity of acidic, hot springs in Lassen Volcanic National Park. *Geomicrobiology Journal 23*, 129-141.

(198) Takai, K., and Horikoshi, K. (2000) Rapid detection and quantification of members of the archaeal community by quantitative PCR using fluorogenic probes. *Applied and Environmental Microbiology 66*, 5066-+.

(199) Martin, A., Yeats, S., Janekovic, D., Reiter, W. D., Aicher, W., and Zillig, W. (1984) Sav-1, a Temperate Uv-Inducible DNA Virus-Like Particle from the Archaebacterium Sulfolobus-Acidocaldarius Isolate B-12. *Embo Journal 3*, 2165-2168.

(200) Baker, G. C., Smith, J. J., and Cowan, D. A. (2003) Review and re-analysis of domain-specific 16S primers. *Journal of Microbiological Methods 55*, 541-555.

(201) Chase, J. M. (2005) Towards a really unified theory for metacommunities. *Functional Ecology 19*, 182-186.

(202) Angly, F., Rodriguez-Brito, B., Bangor, D., McNairnie, P., Breitbart, M., Salamon, P., Felts, B., Nulton, J., Mahaffy, J., and Rohwer, F. (2005) PHACCS, an online tool for estimating the structure and diversity of uncultured viral communities using metagenomic information. *Bmc Bioinformatics 6*, -.

(203) Hubbell (2001) *The Unified Neutral Theory of Biodiversity and Biogeography*, Princeton University Press.

(204) Latimer, A. M., Silander, J. A., and Cowling, R. M. (2005) Neutral ecological theory reveals isolation and rapid speciation in a biodiversity hot spot. *Science 309*, 1722-1725.

(205) Alonso, D., and McKane, A. J. (2004) Sampling Hubbell's neutral theory of biodiversity. *Ecology Letters 7*, 901-910.

(206) Breitbart, M., and Rohwer, F. (2005) Here a virus, there a virus, everywhere the same virus? *Trends in Microbiology 13*, 278-284.

(207) Sano, D., and Omura, T. (2005) Construction of a cloning system for the mass production of a virus-binding protein specific for poliovirus type 1. *Applied and Environmental Microbiology 71*, 2608-2615.

(208) Higgins, D. G., Thompson, J. D., and Gibson, T. J. (1996) Using CLUSTAL for multiple sequence alignments. *Methods Enzymol 266*, 383-402.

(209) Lagos-Quintana, M., Rauhut, R., Lendeckel, W., and Tuschl, T. (2001) Identification of novel genes coding for small expressed RNAs. *Science 294*, 853-858.

(210) Lee, Y., Jeon, K., Lee, J. T., Kim, S., and Kim, V. N. (2002) MicroRNA maturation: stepwise processing and subcellular localization. *Embo Journal 21*, 4663-4670.

(211) Lim, L. P., Lau, N. C., Weinstein, E. G., Abdelhakim, A., Yekta, S., Rhoades, M. W., Burge, C. B., and Bartel, D. P. (2003) The microRNAs of Caenorhabditis elegans. *Genes & Development 17*, 991-1008.

(212) Zamore, P. D., and Haley, B. (2005) Ribo-gnome: The big world of small RNAs. *Science 309*, 1519-1524.

(213) Elbashir, S. M., Lendeckel, W., and Tuschl, T. (2001) RNA interference is mediated by 21-and 22-nucleotide RNAs. *Genes & Development 15*, 188-200.

(214) Elbashir, S. M., Martinez, J., Patkaniowska, A., Lendeckel, W., and Tuschl, T. (2001) Functional anatomy of siRNAs for mediating efficient RNAi in Drosophila melanogaster embryo lysate. *Embo J 20*, 6877-88.

(215) Hammond, S. M., Bernstein, E., Beach, D., and Hannon, G. J. (2000) An RNA-directed nuclease mediates post-transcriptional gene silencing in Drosophila cells. *Nature 404*, 293-296.

(216) Sarnow, P., Jopling, C. L., Norman, K. L., Schutz, S., and Wehner, K. A. (2006) MicroRNAs: expression, avoidance and subversion by vertebrate viruses. *Nature Reviews Microbiology 4*, 651-659.

(217) Pasquinelli, A. E., Reinhart, B. J., Slack, F., Martindale, M. Q., Kuroda, M. I., Maller, B., Hayward, D. C., Ball, E. E., Degnan, B., Muller, P., Spring, J., Srinivasan, A., Fishman, M., Finnerty, J., Corbo, J., Levine, M., Leahy, P., Davidson, E., and Ruvkun, G. (2000) Conservation of the sequence and temporal expression of let-7 heterochronic regulatory RNA. *Nature 408*, 86-89.

(218) Kim, V. N. (2005) MicroRNA biogenesis: Coordinated cropping and dicing. *Nature Reviews Molecular Cell Biology 6*, 376-385.

(219) Olsen, P. H., and Ambros, V. (1999) The lin-4 regulatory RNA controls developmental timing in Caenorhabditis elegans by blocking LIN-14 protein synthesis after the initiation of translation. *Developmental Biology 216*, 671-680.

(220) Reinhart, B. J., Slack, F. J., Basson, M., Pasquinelli, A. E., Bettinger, J. C., Rougvie, A. E., Horvitz, H. R., and Ruvkun, G. (2000) The 21-nucleotide let-7 RNA regulates developmental timing in Caenorhabditis elegans. *Nature 403*, 901-906.

(221) Yekta, S., Shih, I. H., and Bartel, D. P. (2004) MicroRNA-directed cleavage of HOXB8 mRNA. *Science 304*, 594-596.

(222) Doench, J. G., Petersen, C. P., and Sharp, P. A. (2003) siRNAs can function as miRNAs. *Genes & Development 17*, 438-442.

(223) Xie, Z. X., Johansen, L. K., Gustafson, A. M., Kasschau, K. D., Lellis, A. D., Zilberman, D., Jacobsen, S. E., and Carrington, J. C. (2004) Genetic and functional diversification of small RNA pathways in plants. *Plos Biology 2*, 642-652.

(224) Aravin, A. A., Klenov, M. S., Vagin, V. V., Bantignies, F., Cavalli, G., and Gvozdev, V. A. (2004) Dissection of a natural RNA silencing process in the Drosophila melanogaster germ line. *Molecular and Cellular Biology 24*, 6742-6750.

(225) Saito, K., Nishida, K. M., Mori, T., Kawamura, Y., Miyoshi, K., Nagami, T., Siomi, H., and Siomi, M. C. (2006) Specific association of Piwi with rasiRNAs derived from retrotransposon and heterochromatic regions in the Drosophila genome. *Genes & Development 20*, 2214-2222.

(226) Vagin, V. V., Sigova, A., Li, C. J., Seitz, H., Gvozdev, V., and Zamore, P. D. (2006) A distinct small RNA pathway silences selfish genetic elements in the germline. *Science 313*, 320-324.

(227) LILLESTØL, R. K., REDDER, P., GARRETT, R. A., and BRÜGGER, K. (2006) A putative viral defence mechanism in archaeal cells. *Archaea 2*, 59–72.

(228) Zago, M. A., Dennis, P. P., and Omer, A. D. (2005) The expanding world of small RNAs in the hyperthermophilic archaeon Sulfolobus solfataricus. *Molecular Microbiology 55*, 1812-1828.

(229) Tang, T. H., Bachellerie, J. P., Rozhdestvensky, T., Bortolin, M. L., Huber, H., Drungowski, M., Elge, T., Brosius, J., and Huttenhofer, A. (2002) Identification of 86 candidates for small non-messenger RNAs from the archaeon Archaeoglobus fulgidus. *Proceedings of the National Academy of Sciences of the United States of America 99*, 7536-7541.

(230) Tang, T. H., Polacek, N., Zywicki, M., Huber, H., Brugger, K., Garrett, R., Bachellerie, J. P., and Huttenhofer, A. (2005) Identification of novel non-coding RNAs as potential antisense regulators in the archaeon Sulfolobus solfataricus. *Molecular Microbiology 55*, 469-481.

(231) Mojica, F. J. M., Diez-Villasenor, C., Garcia-Martinez, J., and Soria, E. (2005) Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *Journal of Molecular Evolution 60*, 174-182.

(232) Pourcel, C., Salvignol, G., and Vergnaud, G. (2005) CRISPR elements in Yersinia pestis acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. *Microbiology-Sgm 151*, 653-663.

(233) Bolotin, A., Ouinquis, B., Sorokin, A., and Ehrlich, S. D. (2005) Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology-Sgm 151*, 2551-2561.

(234) Jansen, R., van Embden, J. D. A., Gaastra, W., and Schouls, L. M. (2002) Identification of genes that are associated with DNA repeats in prokaryotes. *Molecular Microbiology 43*, 1565-1575.

(235) Baulcombe, D. (2004) RNA silencing in plants. *Nature 431*, 356-363.

(236) Van Rij, R. P., and Andino, R. (2006) The silent treatment: RNAi as a defense against virus infection in mammals. *Trends in Biotechnology 24*, 186-193.

(237) Voinnet, O. (2005) Induction and suppression of RNA silencing: Insights from viral infections. *Nature Reviews Genetics 6*, 206-U1.

(238) Haft, D. H., Selengut, J., Mongodin, E. F., and Nelson, K. E. (2005) A guild of 45 CRISPR-associated (Cas) protein families and multiple CRISPR/Cas subtypes exist in prokaryotic genomes. *Plos Computational Biology 1*, 474-483.

(239) Makarova, K. S., Grishin, N. V., Shabalina , S. A., Wolf, Y. I., and Koonin, E. V. (2006) A putative RNA-interference-based immune system in prokaryotes: computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action. *Biology Direct 1*, 1-26.

(240) Nekrasov, S. V., Agafonova, O. V., Belogurova, N. G., Delver, E. P., and Belogurov, A. A. (2006) Plasmid-encoded Antirestriction Protein ArdA Can Discriminate between Type I Methyltransferase and Complete Restriction-Modification System *Journal of Molecular Biology in press*.

(241) Moult, J., and Melamud, E. (2000) From fold to function. *Current Opinion in Structural Biology 10*, 384-389.

(242) Rossman, M. G. (1981) Evolution of Glycolytic-Enzymes. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences 293*, 191-203.

(243) Buehner, M., Ford, C. C., Moras, D., Olsen, K. W., and Rossman, M. G. (1973) Structure of D-Glyceraldehyde 3-Phosphate Dehydrogenase at 3 a Resolution. *Abstracts of Papers of the American Chemical Society*, 186-186.

(244) Helmercitterich, M., Anceschi, M. M., Banner, D. W., and Cesareni, G. (1988) Control of Cole1 Replication - Low Affinity Specific Binding of Rop (Rom) to Rnai and Rnaii. *Embo Journal 7*, 557-566.

(245) Tomizawa, J. (1985) Control of Cole1 Plasmid Replication - Initial Interaction of Rna-I and the Primer Transcript Is Reversible. *Cell 40*, 527-535

.

(246) Tomizawa, J., and Som, T. (1984) Control of Cole1 Plasmid Replication - Enhancement of Binding of Rna-I to the Primer Transcript by the Rom Protein. *Cell 38*, 871-878.

(247) Aravind, L., Anantharaman, V., Balaji, S., Babu, M. M., and Iyer, L. M. (2005) The many faces of the helix-turn-helix domain: Transcription regulation and beyond. *Fems Microbiology Reviews 29*, 231-262.

(248) Song, J. J., Liu, J. D., Tolia, N. H., Schneiderman, J., Smith, S. K., Martienssen, R. A., Hannon, G. J., and Joshua-Tor, L. (2003) The crystal structure of the Argonaute2 PAZ domain reveals an RNA binding motif in RNAi effector complexes. *Nature Structural Biology 10*, 1026-1032.

(249) Song, J. J., Smith, S. K., Hannon, G. J., and Joshua-Tor, L. (2004) Crystal structure of argonaute and its implications for RISC slicer activity. *Science 305*, 1434-1437.

(250) Duggin, I. G., and Bell, S. D. (2006) The chromosome replication machinery of the archaeon Sulfolobus solfataricus. *Journal of Biological Chemistry 281*, 15029-15032.

(251) Robinson, N. P., Dionne, I., Lundgren, M., Marsh, V. L., Bernander, R., and Bell, S. D. (2004) Identification of two origins of replication in the single chromosome of the Archaeon Sulfolobus solfataricus. *Cell 116*, 25-38.

(252) Lundgren, M., Andersson, A., Chen, L. M., Nilsson, P., and Bernander, R. (2004) Three replication origins in Sulfolobus species: Synchronous initiation of chromosome replication and asynchronous termination. *Proceedings of the National Academy of Sciences of the United States of America 101*, 7046-7051.

(253) Lipps, G. (2004) The replication protein of the Sulfolobus islandicus plasmid pRN1. *Biochemical Society Transactions 32*, 240-244.

(254) Dismukes, G. C. (2001) Photosynthesis - Splitting water. *Science 292*, 447-448.

(255) Dismukes, G. C., Klimov, V. V., Baranov, S. V., Kozlov, Y. N., DasGupta, J., and Tyryshkin, A. (2001) The origin of atmospheric oxygen on Earth: The innovation of oxygenic photosynthesis. *Proceedings of the National Academy of Sciences of the United States of America 98*, 2170-2175.

(256) Yu, B. P. (1994) Cellular Defenses against Damage from Reactive Oxygen Species. *Physiological Reviews 74*, 139-162.

(257) Mittler, R., and Zilinskas, B. A. (2004) *Molecular Ecotoxicology of Plants* Springer, Berlin.

(258) Kehrer, J. P. (2000) The Haber-Weiss reaction and mechanisms of toxicity. *Toxicology 149*, 43-50.

(259) Sawyer, D. T., Sobkowiak, A., and Matsushita, T. (1996) Metal [ML(x); M=Fe, Cu, Co, Mn]/hydroperoxide-induced activation of dioxygen for the oxygenation of hydrocarbons: Oxygenated Fenton chemistry. *Accounts of Chemical Research 29*, 409-416.

(260) Carlioz, A., and Touati, D. (1986) Isolation of Superoxide-Dismutase Mutants in Escherichia-Coli - Is Superoxide-Dismutase Necessary for Aerobic Life. *Embo Journal 5*, 623-630.

(261) Imlay, J. A. (2003) Pathways of oxidative damage. *Annual Review of Microbiology 57*, 395-418.

(262) Pierre, J. L., and Fontecave, M. (1999) Iron and activated oxygen species in biology: The basic chemistry. *Biometals 12*, 195-199.

(263) Lewin, A., Moore, G. R., and Le Brun, N. E. (2005) Formation of protein-coated iron minerals. *Dalton Transactions*, 3597-3610.

(264) Hempstead, P. D., Yewdall, S. J., Fernie, A. R., Lawson, D. M., Artymiuk, P. J., Rice, D. W., Ford, G. C., and Harrison, P. M. (1997) Comparison of the three-dimensional structures of recombinant human H and horse L ferritins at high resolution. *Journal of Molecular Biology 268*, 424-448.

(265) Yang, X. K., Chen-Barrett, Y., Arosio, P., and Chasteen, N. D. (1998) Reaction paths of iron oxidation and hydrolysis in horse spleen and recombinant human ferritins. *Biochemistry 37*, 9743-9750.

(266) Sun, S., and Chasteen, N. D. (1992) Ferroxidase Kinetics of Horse Spleen Apoferritin. *Journal of Biological Chemistry 267*, 25160-25166.

(267) Sun, S. J., Arosio, P., Levi, S., and Chasteen, N. D. (1993) Ferroxidase Kinetics of Human Liver Apoferritin, Recombinant H-Chain Apoferritin, and Site-Directed Mutants. *Biochemistry 32*, 9362-9369.

(268) Lawson, D. M., Treffry, A., Artymiuk, P. J., Harrison, P. M., Yewdall, S. J., Luzzago, A., Cesareni, G., Levi, S., and Arosio, P. (1989) Identification of the Ferroxidase Center in Ferritin. *Febs Letters 254*, 207-210.

(269) Douglas, T., and Ripoll, D. R. (1998) Calculated electrostatic gradients in recombinant human H-chain ferritin. *Protein Science 7*, 1083-1091.

(270) Toussaint, L., Bertrand, L., Hue, L., Crichton, R., and Declercq, J. (2006) High-resolution X-ray Structures of Human Apoferritin H-chain Mutants Correlated with Their Activity and Metal-binding Sites. *Journal of Molecular Biology Epub ahead of print*.

(271) Treffry, A., Zhao, Z., Quail, M. A., Guest, J. R., and Harrison, P. M. (1995) Iron(Ii) Oxidation by H-Chain Ferritin - Evidence from Site-Directed Mutagenesis That a Transient Blue Species Is Formed at the Dinuclear Iron Center. *Biochemistry 34*, 15204-15213.

(272) Bauminger, E. R., Harrison, P. M., Hechel, D., Hodson, N. W., Nowik, I., Treffry, A., and Yewdall, S. J. (1993) Iron (Ii) Oxidation and Early Intermediates of Iron-Core Formation in Recombinant Human H-Chain Ferritin. *Biochemical Journal 296*, 709-719.

(273) Bauminger, E. R., Harrison, P. M., Hechel, D., Nowik, I., and Treffry, A. (1991) Mossbauer Spectroscopic Investigation of Structure-Function Relations in Ferritins. *Biochimica Et Biophysica Acta 1118*, 48-58.

(274) Johnson, E., Cascio, D., Sawaya, M. R., Gingery, M., and Schroder, I. (2005) Crystal structures of a tetrahedral open pore ferritin from the hyperthermophilic Archaeon Archaeoglobus fulgidus. *Structure 13*, 637-648.

(275) Cobessi, D., Huang, L. S., Ban, M., Pon, N. G., Daldal, F., and Berry, E. A. (2002) The 2.6 angstrom resolution structure of Rhodobacter capsulatus bacterioferritin with metal-free dinuclear site and heme iron in a crystallographic 'special position'. *Acta Crystallographica Section D-Biological Crystallography 58*, 29-38.

(276) Macedo, S., Romao, C. V., Mitchell, E., Matias, P. M., Liu, M. Y., Xavier, A. V., LeGall, J., Teixeira, M., Lindley, P., and Carrondo, M. A. (2003) The nature of the di-iron site in the bacterioferritin from Desulfovibrio desulfuricans. *Nature Structural Biology 10*, 285-290.

(277) Andrews, S. C., Lebrun, N. E., Barynin, V., Thomson, A. J., Moore, G. R., Guest, J. R., and Harrison, P. M. (1995) Site-Directed Replacement of the Coaxial Heme Ligands of Bacterioferritin Generates Heme-Free Variants. *Journal of Biological Chemistry 270*, 23268-23274.

(278) Keren, N., Aurora, R., and Pakrasi, H. B. (2004) Critical roles of bacterioferritins in iron storage and proliferation of cyanobacteria. *Plant Physiology 135*, 1666-1673.

(279) Miller, C. D., Kim, Y. C., Walsh, M. K., and Anderson, A. J. (2000) Characterization and expression of the Pseudomonas putida bacterioferritin alpha subunit gene. *Gene 247*, 199-207.

(280) Moore, G. R., Kadir, F. H. A., Almassad, F. K., Lebrun, N. E., Thomson, A. J., Greenwood, C., Keen, J. N., and Findlay, J. B. C. (1994) Structural Heterogeneity of Pseudomonas-Aeruginosa Bacterioferritin. *Biochemical Journal 304*, 493-497.

(281) Carrondo, M. A. (2003) Ferritins, iron uptake and storage from the bacterioferritin viewpoint. *Embo Journal 22*, 1959-1968.

(282) Baaghil, S., Lewin, A., Moore, G. R., and Le Brun, N. E. (2003) Core formation in Escherichia coli bacterioferritin requires a functional ferroxidase center. *Biochemistry 42*, 14047-14056.

(283) Bou-Abdallah, F., Lewin, A. C., Le Brun, N. E., Moore, G. R., and Chasteen, N. D. (2002) Iron detoxification properties of Escherichia coli bacterioferritin - Attenuation of oxyradical chemistry. *Journal of Biological Chemistry 277*, 37064-37069.

(284) Yang, X. O., Le Brun, N. E., Thomson, A. J., Moore, C. R., and Chasteen, N. D. (2000) The iron oxidation and hydrolysis chemistry of Escherichia coli bacterioferritin. *Biochemistry 39*, 4915-4923.

(285) Almiron, M., Link, A. J., Furlong, D., and Kolter, R. (1992) A Novel DNA-Binding Protein with Regulatory and Protective Roles in Starved Escherichia-Coli. *Genes & Development 6*, 2646-2654.

(286) 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. *Nature Structural Biology 5*, 294-303.

(287) Ilari, A., Stefanini, S., Chiancone, E., and Tsernoglou, D. (2000) The dodecameric ferritin from Listeria innocua contains a novel intersubunit iron-binding site. *Nature Structural Biology 7*, 38-43.

(288) Zhao, G. H., Ceci, P., Ilari, A., Giangiacomo, L., Laue, T. M., Chiancone, E., and Chasteen, N. D. (2002) Iron and hydrogen peroxide detoxification properties of DNA-binding protein from starved cells - A ferritin-like DNA-binding protein of Escherichia coli. *Journal of Biological Chemistry 277*, 27689-27696.

(289) Fetter, J., Cohen, J., Danger, D., SandersLoehr, J., and Theil, E. C. (1997) The influence of conserved tyrosine 30 and tissue-dependent differences in sequence on ferritin function: use of blue and purple Fe(III) species as reporters of ferroxidation. *Journal of Biological Inorganic Chemistry 2*, 652-661.

(290) Yamada, Y., Fujiwara, T., Sato, T., Igarashi, N., and Tanaka, N. (2002) The 2.0 angstrom crystal structure of catalase-peroxidase from Haloarcula marismortui. *Nature Structural Biology 9*, 691-695.

(291) Stillman, T. J., Upadhyay, M., Norte, V. A., Sedelnikova, S. E., Carradus, M., Tzokov, S., Bullough, P. A., Shearman, C. A., Gasson, M. J., Williams, C. H., Artymiuk, P. J., and Green, J. (2005) The crystal structures of Lactococcus lactis MG1363 Dps proteins reveal the presence of an N-terminal helix that is required for DNA binding. *Molecular Microbiology 57*, 1101-1112.

(292) Chen, L., and Helmann, J. D. (1995) Bacillus-Subtilis Mrga Is a Dps(Pexb) Homolog - Evidence for Metalloregulation of an Oxidative-Stress Gene. *Molecular Microbiology 18*, 295-300.

(293) Chen, L., James, L. P., and Helmann, J. D. (1993) Metalloregulation in Bacillus-Subtilis - Isolation and Characterization of 2 Genes Differentially Repressed by Metal-Ions. *Journal of Bacteriology 175*, 5428-5437.

(294) Touati, D., Jacques, M., Tardat, B., Bouchard, L., and Despied, S. (1995) Lethal Oxidative Damage and Mutagenesis Are Generated by Iron in Delta-Fur Mutants of Escherichia-Coli - Protective Role of Superoxide-Dismutase. *Journal of Bacteriology 177*, 2305-2314.

(295) Jenney, F. E., Verhagen, M. F. J. M., Cui, X. Y., and Adams, M. W. W. (1999) Anaerobic microbes: Oxygen detoxification without superoxide dismutase. *Science 286*, 306-309.

(296) Massey, V. (1994) Activation of Molecular-Oxygen by Flavins and Flavoproteins. *Journal of Biological Chemistry 269*, 22459-22462.

(297) Sawyer, D. T., and Valentine, J. S. (1981) How Super Is Superoxide. *Accounts of Chemical Research 14*, 393-400.

(298) Flint, D. H., Tuminello, J. F., and Emptage, M. H. (1993) The Inactivation of Fe-S Cluster Containing Hydro-Lyases by Superoxide. *Journal of Biological Chemistry 268*, 22369-22376.

(299) Liochev, S. I. (1996) The role of iron-sulfur clusters in in vivo hydroxyl radical production. *Free Radical Research 25*, 369-384.

(300) Liochev, S. I., and Fridovich, I. (1994) The Role of O-2-Center-Dot- in the Production of Ho-Center-Dot - in-Vitro and in-Vivo. *Free Radical Biology and Medicine 16*, 29-33.

(301) McCord, J. M., and Fridovic.I. (1969) Superoxide Dismutase. *Journal of Biological Chemistry 244*, 6049-6055.

(302) Adams, M. W. W., Jenney, F. E., Clay, M. D., and Johnson, M. K. (2002) Superoxide reductase: fact or fiction? *Journal of Biological Inorganic Chemistry 7*, 647-652.

(303) Su, M., Cavallo, S., Stefanini, S., Chiancone, E., and Chasteen, N. D. (2005) The So-Called Listeria innocua Ferritin Is a Dps Protein. Iron Incorporation, Detoxification, and DNA Protection Properties. *Biochemistry 44*, 5572-8.

(304) Kelley, L. A., MacCallum, R. M., and Sternberg, M. J. E. (2000) Enhanced genome annotation using structural profiles in the program 3D-PSSM. *Journal of Molecular Biology 299*, 499-520.

(305) Maniatis, T., Fritsch, E.F. & Shambrook, J. (1982) *Molecular Cloning: A Laboratory Manual*, Cold Springs Harbor Lab. Press, Plainview, NY.

(306) Allen, M., Willits, D., Young, M., and Douglas, T. (2003) Constrained synthesis of cobalt oxide nanomaterials in the 12-subunit protein cage from Listeria innocua. *Inorganic Chemistry 42*, 6300-6305.

(307) Baker, T. S., and Cheng, R. H. (1996) A model-based approach for determining orientations of biological macromolecules imaged by cryoelectron microscopy. *Journal of Structural Biology 116*, 120-130.

(308) Conway, J. F., Cheng, N., Zlotnick, A., Wingfield, P. T., Stahl, S. J., and Steven, A. C. (1997) Visualization of a 4-helix bundle in the hepatitis B virus capsid by cryo-electron microscopy. *Nature 386*, 91-94.

(309) Bozzi, M., Mignogna, G., Stefanini, S., Barra, D., Longhi, C., Valenti, P., and Chiancone, E. (1997) A novel non-heme iron-binding ferritin related to the DNA-binding proteins of the Dps family in Listeria innocua. *Journal of Biological Chemistry 272*, 3259-3265.

(310) Polidoro, M., De Biase, D., Montagnini, B., Guarrera, L., Cavallo, S., Valenti, P., Stefanini, S., and Chiancone, E. (2002) The expression of the dodecameric ferritin in Listeria spp. is induced by iron limitation and stationary growth phase. *Gene 296*, 121-128.

(311) Ren, B., Tibbelin, G., Kajino, T., Asami, O., and Ladenstein, R. (2003) The multi-layered structure of Dps with a novel di-nuclear ferroxidase center. *Journal of Molecular Biology 329*, 467-477.

(312) Roy, S., Gupta, S., Das, S., Sekar, K., Chatterji, D., and Vijayan, M. (2004) X-ray analysis of Mycobacterium smegmatis Dps and a comparative study involving other Dps and Dps-like molecules. *Journal of Molecular Biology 339*, 1103-1113.

(313) Ceci, P., Cellai, S., Falvo, E., Rivetti, C., Rossi, G. L., and Chiancone, E. (2004) DNA condensation and self-aggregation of Escherichia coli Dps are coupled phenomena related to the properties of the N-terminus. *Nucleic Acids Res 32*, 5935-44.

(314) Reindel, S., Anemuller, S., Sawaryn, A., and Matzanke, B. F. (2002) The DpsA-homologue of the archaeon Halobacterium salinarum is a ferritin. *Biochimica Et Biophysica Acta-Proteins and Proteomics 1598*, 140-146.

(315) Reindel, S., Schmidt, C. L., Anemuller, S., and Matzanke, B. F. (2002) Characterization of a non-haem ferritin of the Archaeon Halobacterium salinarum, homologous to Dps (starvation-induced DNA-binding protein). *Biochemical Society Transactions 30*, 713-715.

(316) Zeth, K., Offermann, S., Essen, L. O., and Oesterhelt, D. (2004) Iron-oxo clusters biomineralizing on protein surfaces: Structural analysis of Halobacterium salinarum DpsA in its low- and high-iron states. *Proceedings of the National Academy of Sciences of the United States of America 101*, 13780-13785.

(317) Ueshima, J., Shoji, M., Ratnayake, D. B., Abe, K., Yoshida, S., Yamamoto, K., and Nakayama, K. (2003) Purification, gene cloning, gene expression, and mutants of Dps from the obligate anaerobe Porphyromonas gingvalis. *Infection and Immunity 71*, 1170-1178.

(318) Nair, S., and Finkel, S. E. (2004) Dps protects cells against multiple stresses during stationary phase. *Journal of Bacteriology 186*, 4192-4198.

(319) Park, C. B., and Lee, S. B. (1999) Inhibitory effect of mineral ion accumulation on high density growth of the hyperthermophilic archaeon Sulfolobus solfataricus. *Journal of Bioscience and Bioengineering 87*, 315-319.

(320) Nakamura, Y., Kaneko, T., Sato, S., Mimuro, M., Miyashita, H., Tsuchiya, T., Sasamoto, S., Watanabe, A., Kawashima, K., Kishida, Y., Kiyokawa, C., Kohara, M., Matsumoto, M., Matsuno, A., Nakazaki, N., Shimpo, S., Takeuchi, C., Yamada, M., and Tabata, S. (2003) Complete genome structure of Gloeobacter violaceus PCC 7421, a cyanobacterium that lacks thylakoids. *DNA Research 10*, 137-145.

(321) Grossman, M. J., Hinton, S. M., Minakbernero, V., Slaughter, C., and Stiefel, E. I. (1992) Unification of the Ferritin Family of Proteins. *Proceedings of the National Academy of Sciences of the United States of America 89*, 2419-2423.

(322) Fridovich, I. (1986) Superoxide Dismutases. *Advances in Enzymology and Related Areas of Molecular Biology 58*, 61-97.

(323) Robb, F. T., Maeder, D. L., Brown, J. R., DiRuggiero, J., Stump, M. D., Yeh, R. K., Weiss, R. B., and Dunn, D. M. (2001) Genomic sequence of hyperthermophile, Pyrococcus furiosus: implications for physiology and enzymology. *Methods Enzymol 330*, 134-57.

(324) Bravo, J., Verdaguer, N., Tormo, J., Betzel, C., Switala, J., Loewen, P. C., and Fita, I. (1995) Crystal-Structure of Catalase Hpii from Escherichia-Coli. *Structure 3*, 491-502.

(325) Ko, T. P., Safo, M. K., Musayev, F. N., Di Salvo, M. L., Wang, C. Q., Wu, S. H., and Abraham, D. J. (2000) Structure of human erythrocyte catalase. *Acta Crystallographica Section D-Biological Crystallography 56*, 241-245.

(326) Shima, S., Netrusov, A., Sordel, M., Wicke, M., Hartmann, G. C., and Thauer, R. K. (1999) Purification, characterization, and primary structure of a monofunctional catalase from Methanosarcina barkeri. *Archives of Microbiology 171*, 317-323.

(327) Ward, D. E., Donnelly, C. J., Mullendore, M. E., van der Oost, J., de Vos, W. M., and Crane, E. J. (2001) The NADH oxidase from Pyrococcus furiosus - Implications for the protection of anaerobic hyperthermophiles against oxidative stress. *European Journal of Biochemistry 268*, 5816-5823.

(328) Touati, D., Jacques, M., Despied, S., Tardat, B., and Bouchard, L. (1995) Correlation between Iron-Metabolism, Superoxide-Dismutase Expression and Oxidative Stress in Escherichia-Coli. *Journal of Cellular Biochemistry*, 236-236.

(329) Blumentals, I. I., Itoh, M., Olson, G. J., and Kelly, R. M. (1990) Role of Polysulfides in Reduction of Elemental Sulfur by the Hyperthermophilic Archaebacterium Pyrococcus-Furiosus. *Applied and Environmental Microbiology 56*, 1255-1262.

(330) Madera, M., Vogel, C., Kummerfeld, S. K., Chothia, C., and Gough, J. (2004) The SUPERFAMILY database in 2004: additions and improvements. *Nucleic Acids Research 32*, D235-D239.

(331) Tatusov, R. L., Fedorova, N. D., Jackson, J. D., Jacobs, A. R., Kiryutin, B., Koonin, E. V., Krylov, D. M., Mazumder, R., Mekhedov, S. L., Nikolskaya, A. N., Rao, B. S., Smirnov, S., Sverdlov, A. V., Vasudevan, S., Wolf, Y. I., Yin, J. J., and Natale, D. A. (2003) The COG database: an updated version includes eukaryotes. *Bmc Bioinformatics 4*, -.

(332) Adams, M. W. W., Holden, J. F., Menon, A. L., Schut, G. J., Grunden, A. M., Hou, C., Hutchins, A. M., Jenney, F. E., Kim, C., Ma, K. S., Pan, G. L., Roy, R., Sapra, R., Story, S. V., and Verhagen, M. F. J. M. (2001) Key role for sulfur in peptide metabolism and in regulation of three hydrogenases in the hyperthermophilic archaeon Pyrococcus furiosus. *Journal of Bacteriology 183*, 716-724.

(333) Tusher, V. G., Tibshirani, R., and Chu, G. (2001) Significance analysis of microarrays applied to the ionizing radiation response. *Proceedings of the National Academy of Sciences of the United States of America 98*, 5116-5121.

(334) Bolstad, B. M., Irizarry, R. A., Astrand, M., and Speed, T. P. (2003) A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics 19*, 185-193.

(335) Zeller, T., Moskvin, O. V., Li, K. Y., Klug, G., and Gomelsky, M. (2005) Transcriptome and physiological responses to hydrogen peroxide of the facultatively phototrophic bacterium Rhodobacter sphaeroides. *Journal of Bacteriology 187*, 7232-7242.

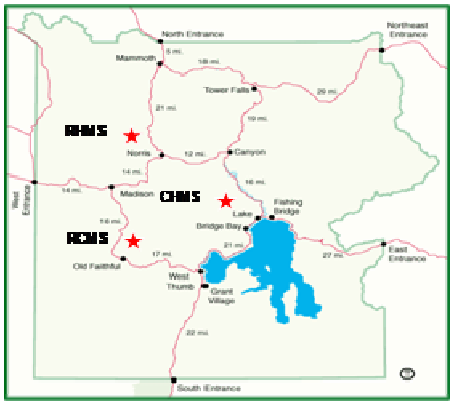
APPENDIX A:

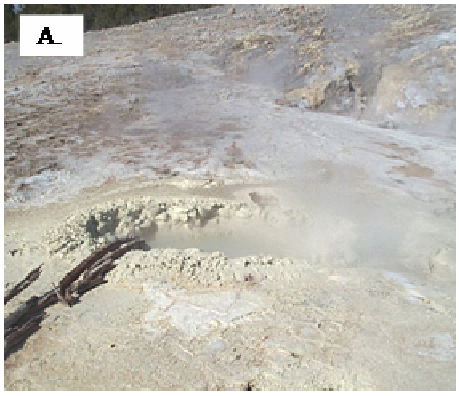
SUPPLEMENTARY INFORMATION

APPENDIX A

Chapter 4

Figure S4.1 Map of Yellowstone National Park and the three monitor sites. Red stars indicate site locations: A. Crater Hills Monitor Site (CHMS) (44°39.31´N, 110°28.95'W) B. Rabbit Creek Monitor Site (RCMS) (44°31.287´N, 110°48.647'W) C. Ragged Hills Monitor Site (RHMS) (44°43.653´N, 110°42.862'W).





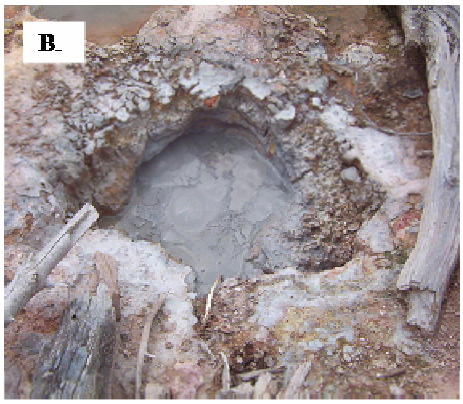
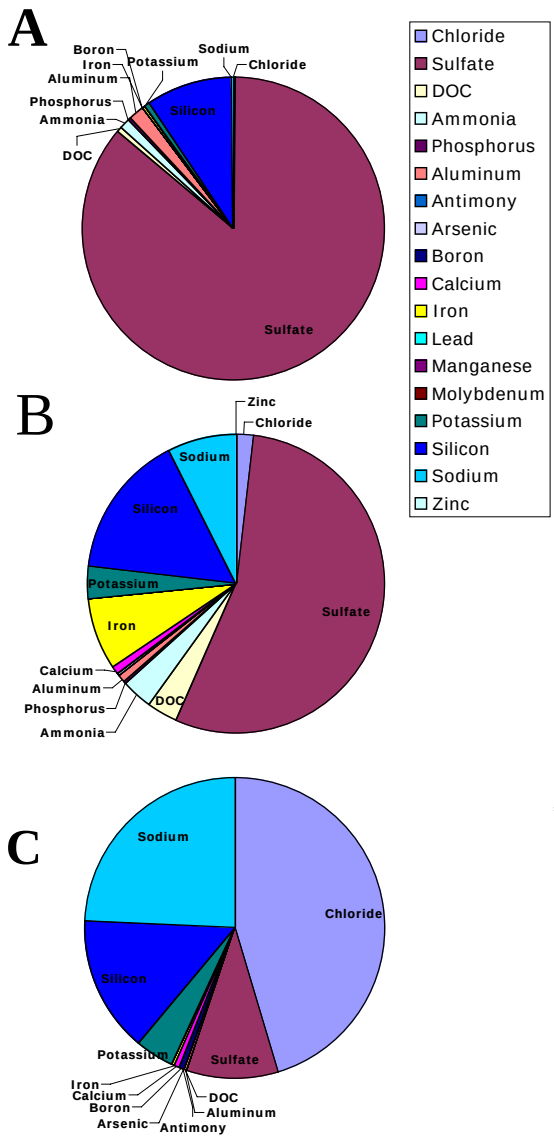


Figure S4.2 Each hot spring has a distinct geochemical signature. Although not entirely static, the overall geochemical signature of each spring appeared to remain relatively stable over the two year sampling course. A. Crater Hills Monitor Site B. Rabbit Creek Monitor Site C. Ragged Hills Monitor Site



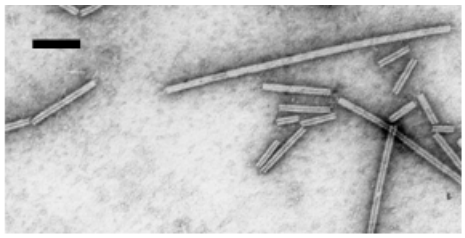
	CHMS	RCMS	RHMS
Chloride	1.8 *	3	611
Sulfate	1022	87	135
DOC	5.74	5.3	2.5
Ammonia	14.8	5.3	0.85
Nitrogen	0	0	0
Phosphorus	0.47	0.7	0.203
Aluminum	20.8	1.4	1.9
Antimony	0	0	0.125
Arsenic	0.01	0.04	1.01
Barium	0	0	0
Boron	1.4	0	7.9
Cadmium	0	0	0
Calcium	0	1.5	5.8
Copper	0	0	0
Iron	1.47	12.2	2.7
Lead	0.012	0	0
Magnesium	0	0	0
Manganese	0.02	0.08	0.03
Mercury	0	0	0
Molybdenum	0	0	0.18
Nickel	0	0	0
Potassium	9	5.8	59
Selenium	0	0	0
Silicon	108	24.9	198.5
Sodium	2.8	11.7	326
Zinc	0.07	0.13	0.04

* Units are mg/L

Figure S4.3 Transmission electron micrographs (TEMs) of SSV (A) and SIRV (B), and sequence alignments that were used to design these viral primer sets. SSV primers (3F and 4R) were designed using the complete genome sequences of four SSV isolates (Wiedenheft et al 2004). These primers amplify 250-260 base pairs (bp) in the largest and one of the most highly conserved open reading frames (ORF) of each sequenced SSV isolate. SIRV primers (DBPF and DBPR) were designed using the complete genome sequences of two Icelandic isolates (SIRV1 and SIRV2) and partial genomes sequences form two Yellowstone isolates (RC7#4 and 9Y-NL) (Peng X, Virology 2001). (Scale bars =100nm) (Red= nucleotides are conserved in 3 or more of the available complete genome sequences, Blue = conserved in 2 of the available sequences, Green=not conserved)

A.

B.



```
3F          CAATCGCCATAGGCTACGG
SSV1   CGGTGCAATCGCCATAGGCTACGGAGTGAACTTTCCCGC…
SSV2   CGGTTCAATCGCCATAGGCTACGGTGTAAATTTCCCTTC…
SSVRH  CGGTTCAATCGCCATAGGCTACGGTGTTAATTTTCCTTC…
SSVK1  CGGTTCAATCGCCATAGGCTACGGTGTAAATTTCCCTTC…

SSV1   …TTAGGGTTCTTTCACTTTACGTTTACTACTATAACGGTACTT
SSV2   …GGTACTTACCCCAACTTTACGTTTATTACTATAACGGTACTT
SSVRH  …AGGGTTCTGGTCAACTTTACGTTTATTACTATAACGGTACTC
SSVK1  …AGGGTTTTGGTCAACTTTACGTTTACTACTATAACGGTACTC
4R                          CGTTTAYTACTATAACGGTAC
```

```
DBPF              GATATTGACCAAAAATGGCAAAAGG
SIRV2  GAATAAAATTGGATATTGACCAAATATGGCAAAAGGTCACACATCAAGAAGT
SIRV1  GAATAAAATTGGATATTGACCAAAAATGGCAAAAGGTCACACAAAAAGATCA
RC7#4  ------------------------ATGGCAAAAGGTAGAACACCAAGATCA
9Y-NL  ------------------------ATGGCAAAAGGTAGAACACCAAGATCA

SIRV2  CATAACTGGAAGCCCAACTGGATATGTAAGTTAAAATATTAAGCTATTTTTT
SIRV1  AATAACTGGAAGCCCGACTGGATACGTAAGTTAAAATATTAAGCTATTTTTT
RC7#4  AATAACCGGTAGTCCAACTGGATACGTAAGTTAA----------------
9Y-NL  AATAACCGGTAGTCCAACTGGATACGTAAGTTAA----------------
DBPR                    GGATACGTAAGTTAAAATATTAAGC
```

*RC7#4 and 9Y-NL are two partially sequenced isolates from YNP.

Figure S4.5 Sequence accumulation curves for both (A) SSV and (B) SIRV are not asymptotic, indicating many more sequences need to be sampled before capturing the complete viral diversity. Maybe a sentence about the predicted size

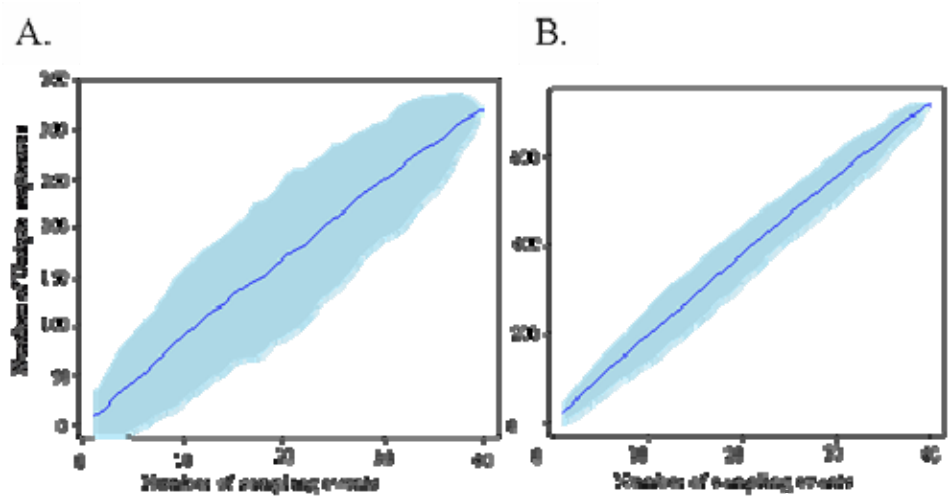
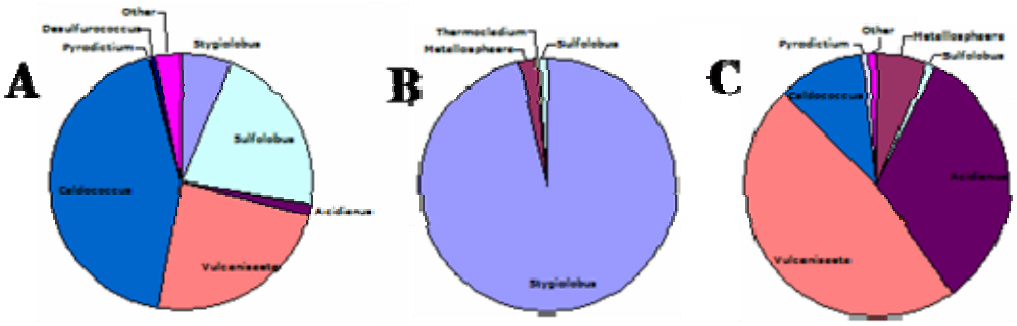


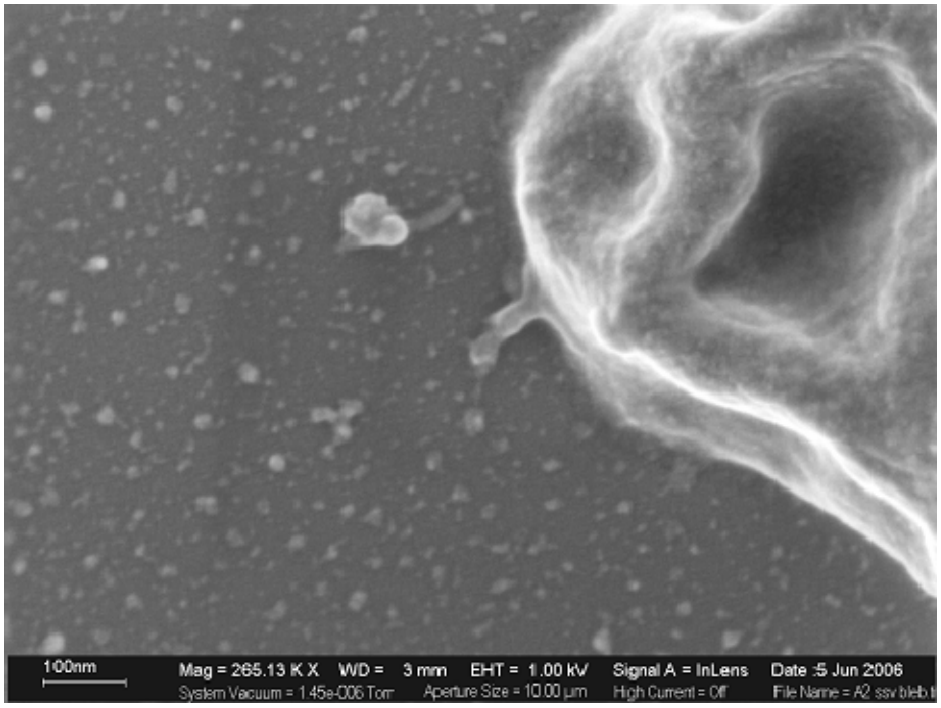
Figure S4.5. A 16S comparison of the Archaeal community detected at each of the three springs A. Crater Hills B. Rabbit Creek C. Ragged Hills. Community composition was relatively stable over the two year time course, with *Sulfolobus* being the only common, but never the dominate resident in each of the three springs.



Organism Name	Number of clones and percent match range		
	Crater Hills	Ragged Hills	Rabbit Creek
Caldococcus	126 (86-97%)	ND	10 (90-97%)
Vulcanisaeta	69 (87-98%)	ND	46 (94-98%)
Sulfolobus	63 (87-94%)	3 (98%)	1 (95%)
Acidianus	4 (93-99%)	ND	32 (96-99%)
Metallosphaera	ND	6 (94-97%)	6 (97%)
Pyrodictium	1 (90%)	ND	1 (91%)
Stygiolobus	15 (88-95%)	269 (95-99%)	ND
Thermocladium	ND	2 (96-97%)	ND
Desulfurococcus	1 (83%)	ND	ND

Chapter 5

Fig. S5.1 A bleb consistent with both SSV size and morphology forms on the surface of an infected *S. solfataricus* cell . Scanning electron micrograph taken with assistance from Dr. Peter Suci



Chapter 8

Figure S8.8 *Sulfolobus solfataricus* DNA-binding protein from starved cells (*SsDps*) expression in response to oxidative stress. (*A*) Northern blot analysis of RNA extracted from exponential growth phase (OD_{650} ≈0.5) *Sulfolobus solfataricus* cells. Lane 1, standard growth conditions (0 mM H_2O_2); lane 2, 5 mM H_2O_2; lane 3, 10 mM H_2O_2; lane 4, 20 mM H_2O_2; lane 5, 30 mM H_2O_2. RNA gels were stained with ethidium bromide (EtdBr) before membrane transfer (*Middle*). Blots were hybridized with a probe specific for *ssdps* and then stripped and rehybridized with a probe specific for the single-stranded binding (SSB) protein, as an additional load control (*Bottom*) (1, 2). Approximately 1.5 mg of total RNA was loaded in each lane. (*B*) Western blot analysis of total protein extract from logarithmic (OD_{650} ≈0.5) *S. solfataricus* cells; loads are normalized according to number of cells. Lane 1, 0 m M H_2O_2 (standard growth conditions); lane 2, 5 mM H_2O_2; lane 3, 10 mM H_2O_2; lane 4, 20 mM H_2O_2; lane 5, 30 mM H_2O_2; lane 6, 2.0 ng of purified recombinant *Ss*Dps; lane 7, 1.0 ng of purified recombinant *Ss*Dps; lane 8, 0.5 ng of purified recombinant *Ss*Dps. Densitometry was performed using AlfaImager (Alpha Innotech, San Leandro, CA) imaging system (IS-2200) according to average pixel density.

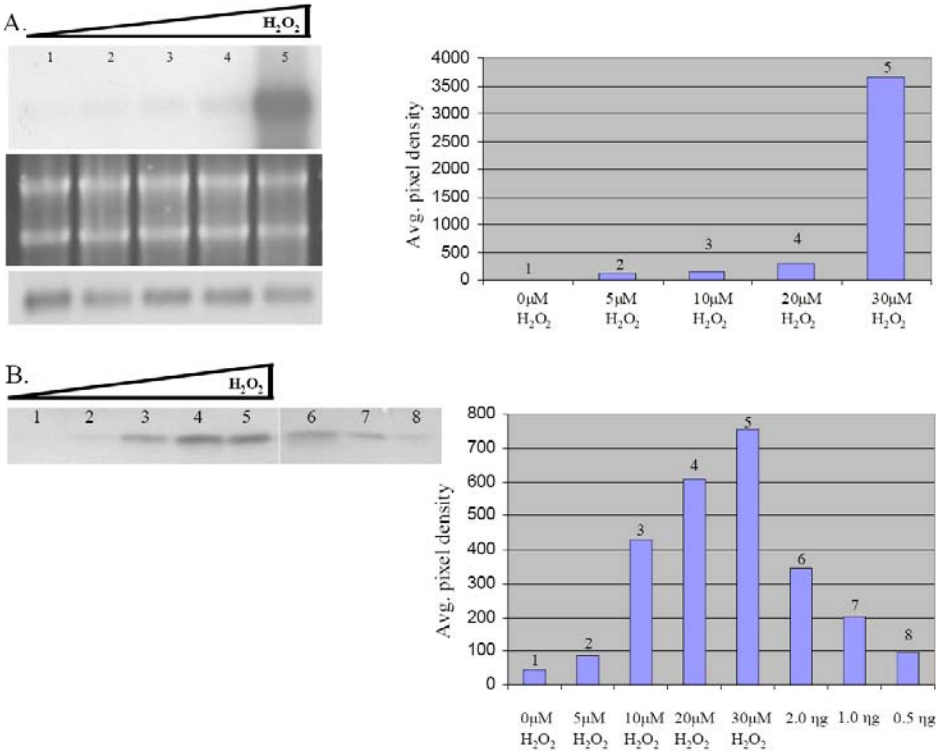


Figure S8.9. *SsDPSL*expression is specific to oxidative stress. Northern blot analysis of RNA extracted from *S. solfataricus* cultures grown under standard conditions (lane 1), 30 mM H_2O_2 (lane 2), late log phase cells (OD_{650} ≈1.2) (lane 3); 300 KJ UV, which accounts for RNA degradation as shown in the load control (lane 4); sucrose as the sole carbon source (lane 5); *Sulfolobus* spindle-shaped virus, Raggged Hills (SSV RH) virus-infected (3) (lane 6); 11 h heat shock at 90°C (lane 7); 11 h cold shock at 60°C. Approximately 0.75 mg of total RNA was loaded in each lane. The gel was stain with EtdBr before membrane transfer and serves as the loading control because concentrations of the SSB control were not consistent over this broad range of stress conditions.

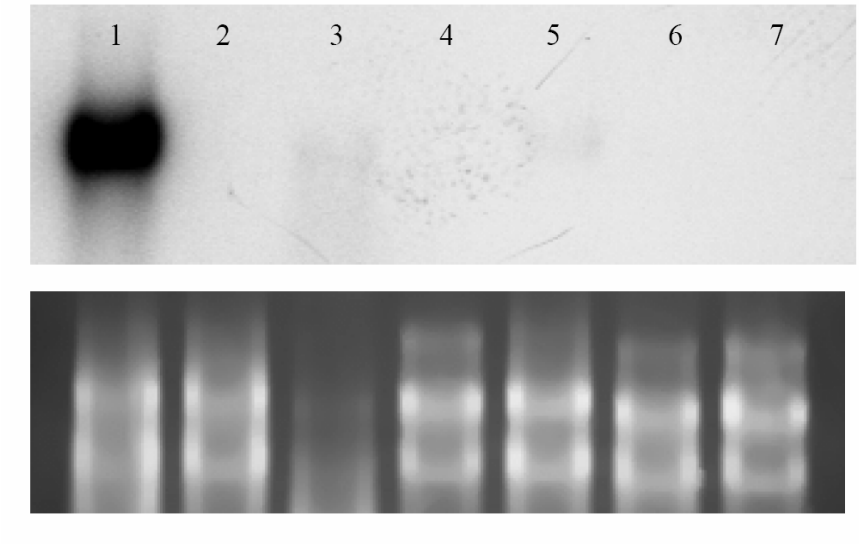


Figure S8.10 Heterologous expression and purification of *Ss*Dps. Coomassie-stained SDS/PAGE of BL21 *Escherichia coli* cells transformed with the *Ss*Dps gene. Lane 1, the molecular mass standard; lane 2, whole cell lysate; lane 3, after heat denaturation at 70°C for 10 min; lane 4, after size exclusion liguid chromatography.

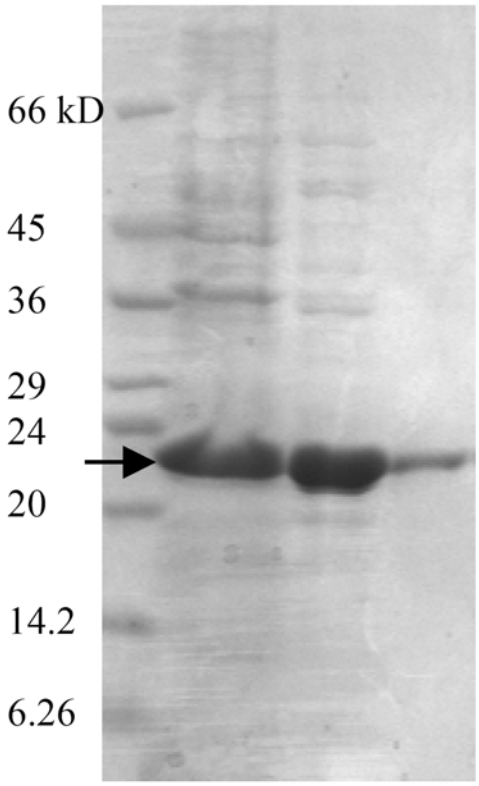
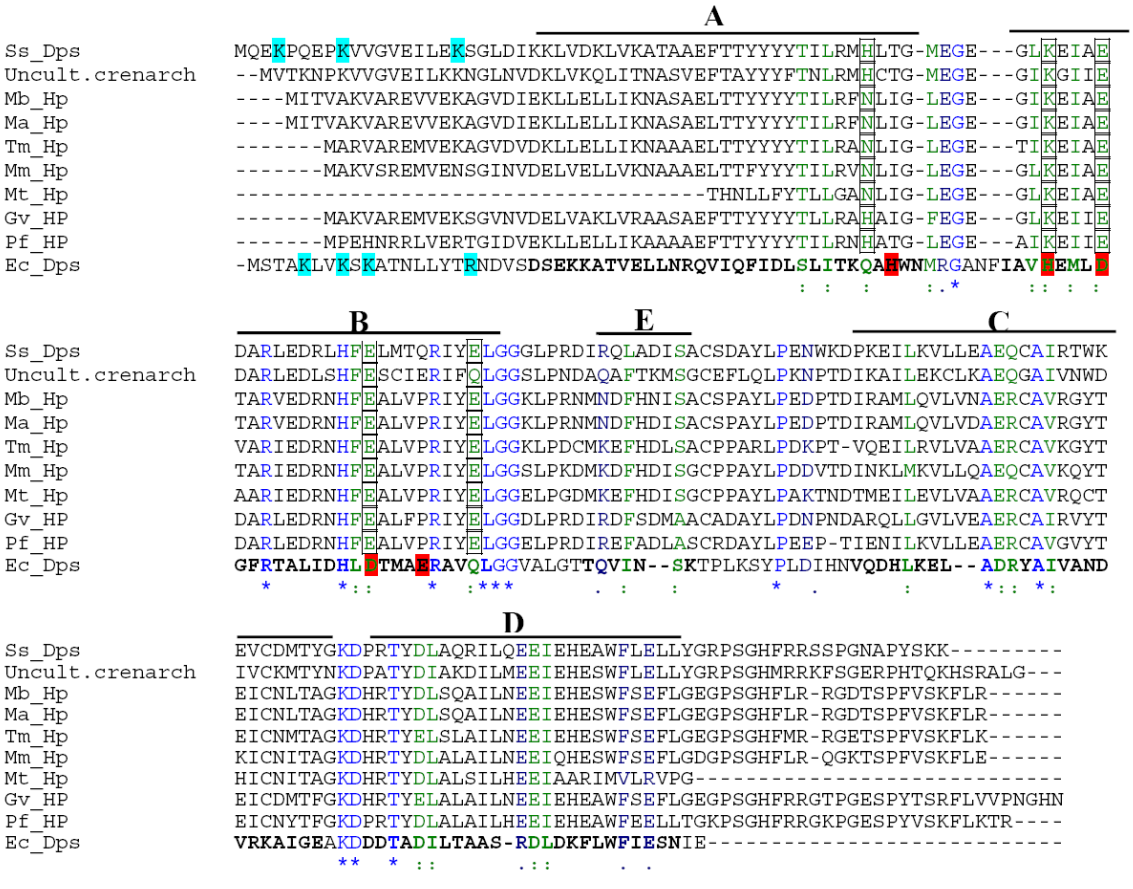


Figure S8.11 Proteins within the monophyletic cluster of Dps-like proteins all contain a putative Dps-like metal-binding domain. A putative diiron-binding motif, like that defined in all known Dps proteins, is identifiable in the primary structure of all of the hypothetical proteins in the Dps-like cluster (boxed amino acids in hypothetical proteins and in red in *E.coli* Dps). Although the amino acid sequences of this diiron-binding motif are not fully conserved, amino acid charge and motif spacing are conserved. Sequences were aligned with clustalw and NCBI gene identification numbers (gi), genus and species names of aligned sequences are provided in the phylogenetics legend to Fig. 7. Basic residues shaded in cyan are putatively involved in DNA binding. *, Fully conserved residue; colon, conservation of strong groups; period, conservation of weak groups; -, no consensus. Five separate alpha helixes (A-E) are evident in the *E. coli* Dps crystal structure (bold) and predicted in secondary structures of the new Dps-like sequences.



Chapter 10

Fig. S10.1 The *dpsl* disruption mutant is substantially more sensitive to H2O2 then the wild type. Mid-log phase cultures of the wild type and dpsl mutant strain were exposed to increasing amounts of H2O2. The dip in the growth curve at ~20hrs. clearly indicates the H2O2 delivery point. A. Growth curves of wild type *S. solfataricus* 98/2 and *dpsl* mutant strains are indistinguishable in when grown under standard aerobic conditions. B. At 20 µM H2O2 the dpsl mutant reaches a maximum optimal density (OD650) that is only half that of the wild type 98/2 strain. C. The trend is not consistent at 30 µM H2O2 where the growth curves are indistinguishable. D. At 30 µM H2O2 the mutant is not viable while the wild type strain rapidly recovers, reaching normal cell densities. E. At 40 µM H2O2 the dpsl mutant strain is not viable, while wild type only partially recovers. F. 60 µM H2O2 is lethal for both the wild type and the mutant.

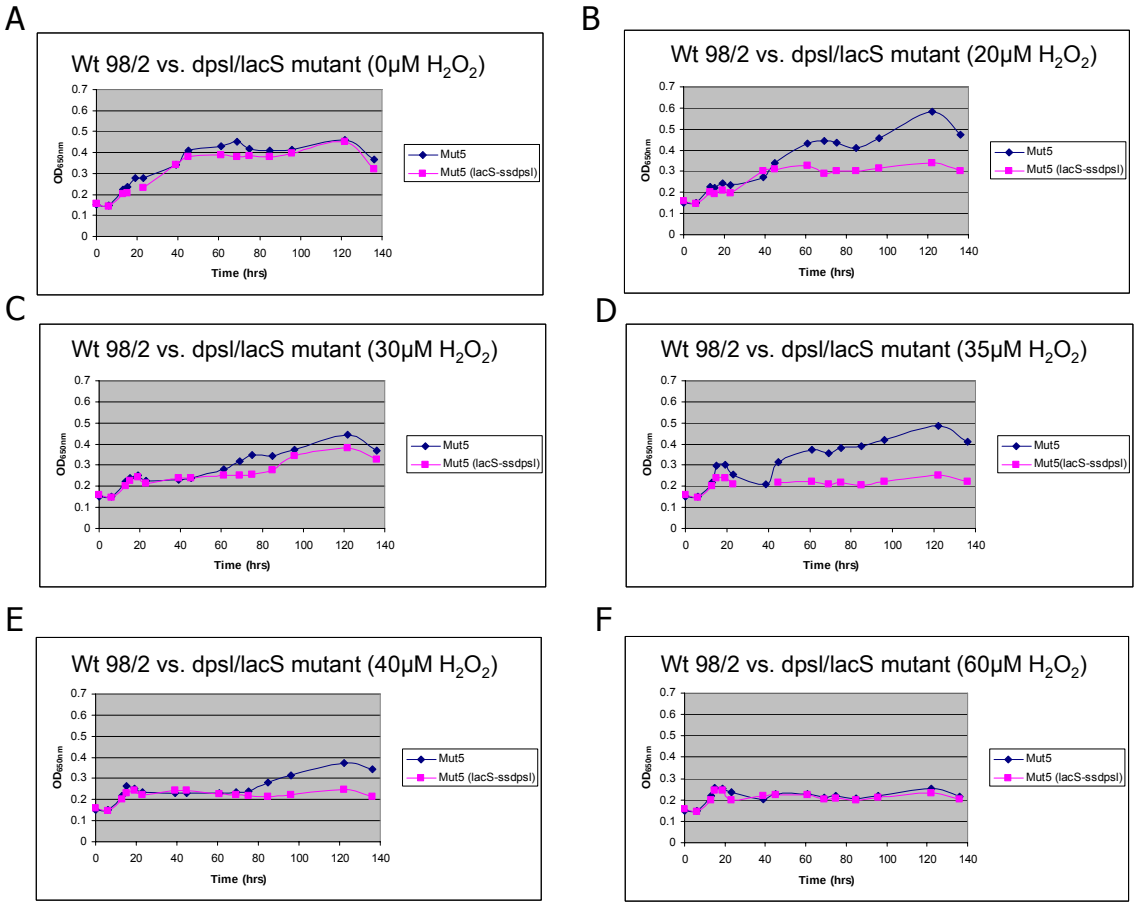
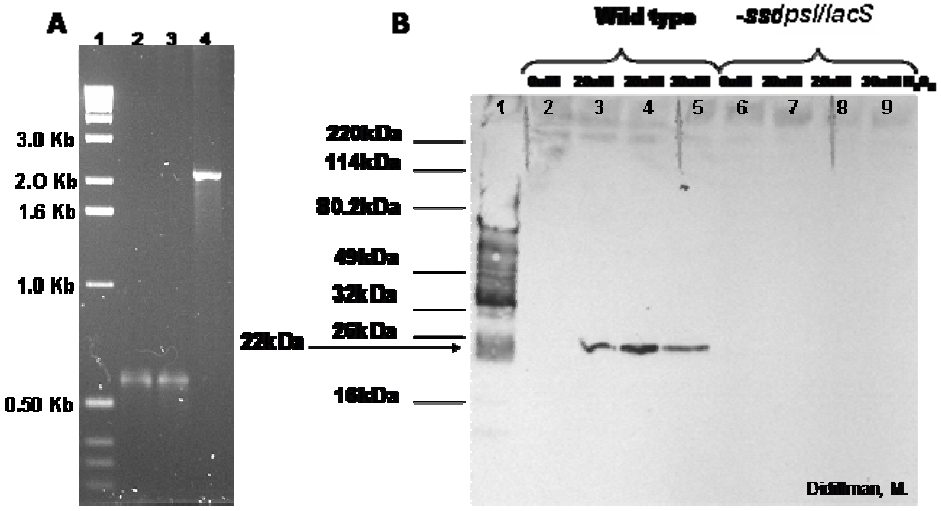


Fig. S10.2 LacS disruption mutant of the *ssdpsl* gene in the 98/2 strain of *S. solfataricus*. A. PCR amplification of the dpsl gene from genomic DNA isolated from, Lane 2) *S. solfataricus*, strain P2; Lane 3) *S. solfataricus* strain 98/2; Lane 4) *lacS* insertion into the *S. solfataricus* strain 98/2 *dpsl gene.* DNA sequencing identified a single nucleotide difference between the *S. solfataricus* P2 and 98/2 *dpsl* genes. B. Western Blot performed on wild type and *dpsl* mutant cells stressed with 0, 20 25 and 30μM H2O2. Approximately 8μg of protein was loaded in each lane and electrophoretically separated on a 15% SDS-polyacrylamide gel. Proteins were transferred to a nitrocellulose membrane and probed with polyclonal antibodies raised against purified recombinant *Ss*DPSL protein (*7*).



Microbial Community in Select Hot Springs of the Geyser Valley and Uzon Caldara (8/2006)

A total of 14 thermal features were sampled from two distinct locations on the Kamchatka peninsula; known as the Uzon Caldara (N 54º30.005; E 160º00.256) and Geyser Valley (N 54º26.264; E 160º08.416). Total DNA was extracted from 6 of these samples; all form the Uzon Caldara location. Total environmental DNA was used as a template for 16S amplification using primers 1391R and 515F. Sixteen clones from each site were sequenced and analyzed. The results are summarized below (A=Archaea, B=Bacteria E=Eukarya, X=failed reaction). Universally SSV primers (3F and 4R; in the C792 region) and SIRV primers (DBP F and DBP R; coat protein region) were used in an effort to detect known viral signatures. A visible PCR product was never obtain from any of the environmental DNA samples.

Four of these samples, collected from springs most closely repressing an environment suitable for Sulfolobus, were used to inoculate Sulfolobus media 182 at either pH 2, 3 or 4. All cultures were placed at ~78ºC and at 3 weeks post inoculation none of four samples at any of the tested pHs exhibited noticeable growth. This is consistent with the 16S analysis that did not detect a Sulfolobus-like signature for any of the environmental samples and the inability to PCR amplify viral signatures commonly associated with Sulfolobus spp.

*Geochemical analysis of indicted thermal features was conducted by Andrew Bychkov (Bychkov@geol.msu.Ru). Results were never obtained. $ Enrichment Culture (EC)

Region/GPS	Environment	Sample type	chemistry *	Total DNA	16S rDNA	SSV/ SIRV	EC $
U.Z. #1 (N 54º30.005; E 160º00.256)	pH5/86ºC	a. 900mls onto 0.45 µm filter b. 500mls to Diversa c. 1.5ml epi x 2		+	A=0 B=16 E=0 X=0	–/–	
U.Z. #2 (N 54º30.053; E 160º00.124)	pH5.5/83ºC	a. 900mls onto 0.45 µm filter b. 1.5ml epi x 2	Chem #5	+	A=0 B=15 E=0 X=1	–/–	_
U.Z. #3	pH5.5/95ºC	a. 900mls onto 0.45 µm filter b. 500mls to Diversa c. 1.5ml epi x 2	Chem #6	+	A=10 B=3 E=0 X=3	–/–	
U.Z. #4	pH7/90ºC	a. 900mls onto 0.45 µm filter b. 500mls to Diversa c. 1.5ml epi x 2	Chem #7	+	A=0 B=13 E=1 X=2	–/–	
U.Z. #5	pH2/78ºC	a. 900mls onto 0.45 µm filter b. 1.5ml epi x 2		+	A=10 B=3 E=0 X=3	–/–	_
U.Z. #6 (N 54º29.956; E 160º00.658)	pH3.5/80ºC			+	A=6 B=5 E=0 X=5	–/–	_
G.V. #1 (N 54º26.264; E 160º08.416)	pH5/85ºC	a. 50mls onto 0.45 µm filter b. 1.5ml epi x 4				–/–	
G.V. #2	pH5/82ºC	a. 50mls onto 0.45 µm filter b. 1.5ml epi x 4				–/–	
G.V. #3	pH7/85ºC	a. 1.5ml epi x 4				–/–	
G.V. #4	pH5/80ºC	a. 1.5ml epi x 4				–/–	
G.V. #6	pH3.5/80ºC	a. 1.5ml epi x 4				–/–	_
G.V. #7	pH4.5/80ºC	a. 50mls onto 0.45 µm filter b. 1.5ml epi x 4				–/–	
G.V. #8	pH5.5/80ºC	a. 500mls to Diversa c. 1.5ml epi x 4				–/–	
G.V. #9	pH5/83ºC	c. 1.5ml epi x 4				–/–	