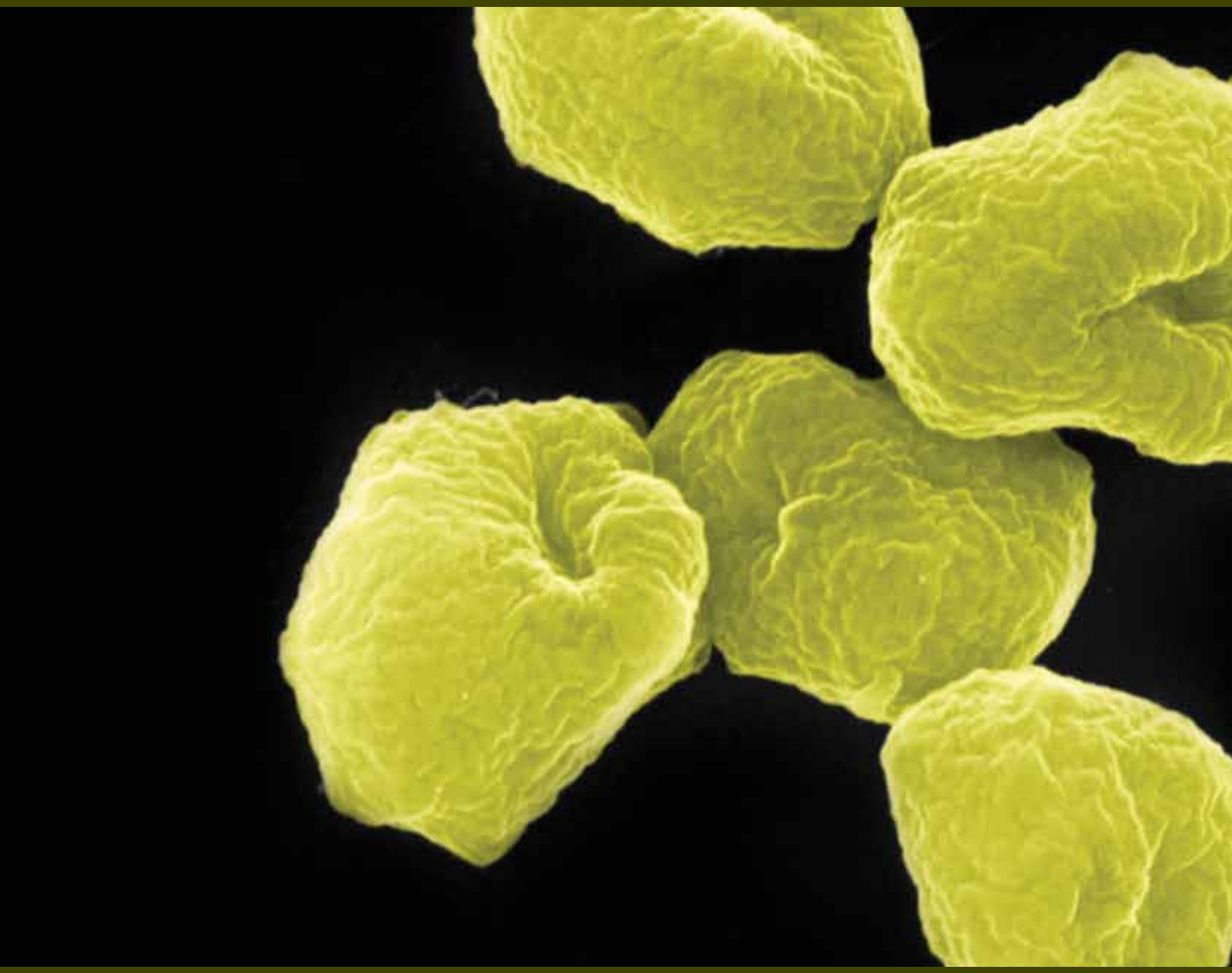
Archaea

Archaeal Protein Biogenesis: Posttranslational Modification and Degradation

Guest Editors: Jerry Eichler, Julie Maupin-Furlow, and Joerg Soppa



Archaeal Protein Biogenesis: Posttranslational Modification and Degradation

Archaea

Archaeal Protein Biogenesis: Posttranslational Modification and Degradation

Guest Editors: Jerry Eichler, Julie Maupin-Furlow, and Joerg Soppa

Editorial Board

Contents

Hindawi Publishing Corporation
Archaea
Volume 2010, Article ID 643046, 2 pages
doi:10.1155/2010/643046

Editorial

Archaeal Protein Biogenesis: Posttranslational Modification and Degradation

Jerry Eichler,[1] Julie Maupin-Furlow,[2] and Joerg Soppa[3]

[1] *Department of Life Sciences, Ben-Gurion University of the Negev, Beersheva, Israel*
[2] *Department of Microbiology and Cell Science, University of Florida, Gainesville, FL 32611, USA*
[3] *Institute for Molecular Biosciences, Frankfurt University, 60325 Frankfurt, Germany*

Correspondence should be addressed to Jerry Eichler, jeichler@bgu.ac.il

Received 1 August 2010; Accepted 1 August 2010

The study of Archaea at the DNA and RNA levels has provided considerable insight into replication, transcription, and other information-associated events which are either unique to this remarkable group of organisms or which were later found to also occur in Bacteria and/or Eukarya. In contrast, largely due to a lack of suitable model systems and a limited number of appropriate molecular tools, considerably less was known about archaeal proteins in terms of their biogenesis, modification, trafficking, or degradation. In recent years, however, we have witnessed major advances in proteomics, successful in vitro reconstitutions, and the development of reporter systems compatible with extreme conditions. Relying on such tools, insight into different stages in the life of archaeal proteins has begun to accumulate. In this special issue of Archaea, we present a series of articles addressing the current state of understanding of selected facets of archaeal protein biogenesis and processing.

An article by De Koning et al. discussing how fidelity in archaeal information processing at the DNA, RNA, and protein levels is achieved begins this special issue. Rother and Krzycki then address proteins containing the unusual animo acids, selenocysteine, and pyrrolysine, as related to the unique energy metabolism of methanogenic archaea. Soppa compares one specific posttranslational modification, acetylation, across evolutionary lines, while Botting et al. discuss the importance of lysine methylation in hyperthermophilic crenarchaeota. Questions related to protein assembly are considered when Iwasaki addresses the iron-sulfur world in hyperthermoacidophilic archaea.

Protein degradation is the focus of two articles in this special issue. Makarova and Koonin rely on comparative genomic analysis to reveal functional versatility of archaeal ubiquitin-like proteins, while Humbard et al. report on the phosphorylation and methylation of *Haloferax volcanii* proteasomal proteins.

Archaeal cell surface proteins undergo a variety of post-translational modifications. However, such proteins must first be targeted to and traverse the plasma membrane. Accordingly, Zwieb and Bhuiyan discuss the latest findings on the archaeal signal recognition particle targeting system. Storf et al. address questions related to the biogenesis of one class of membrane proteins, namely, lipoproteins. An article by Ellen et al. considers archaeal protein export and describes different cell surface structures, while a report by Jarrell et al. focuses on the S-layer glycoprotein and flagella as reporters of choice of different protein processing events. With this in mind, Peyfoon and colleagues describe the N-linked glycan decorating the S-layer glycoprotein of *Sulfolobus acidocaldarius*. Finally, Kaminski and Eichler consider the workings of AglD, one of the enzymes which is involved in *Haloferax volcanii* S-layer glycoprotein N-glycosylation.

In this, the first special issue of Archaea dedicated to archaeal protein biogenesis, we have tried to give the reader a sampling of the current research scene. It is our sincere hope that the community will find interest in the articles included here. More importantly, it is our intention that the work presented will stimulate other laboratories to begin studying questions related to archaeal protein biogenesis,

hopefully in time for the next installment of this special issue.

Jerry Eichler
Julie Maupin-Furlow
Joerg Soppa

Hindawi Publishing Corporation
Archaea
Volume 2010, Article ID 960298, 15 pages
doi:10.1155/2010/960298

Review Article

Fidelity in Archaeal Information Processing

Bart de Koning, Fabian Blombach, Stan J. J. Brouns, and John van der Oost

Laboratory of Microbiology, Wageningen University, Dreijenplein 10, 6703 HB Wageningen, The Netherlands

Correspondence should be addressed to Bart de Koning, bart.dekoning@gen.unimaas.nl

Received 15 May 2010; Accepted 12 July 2010

Academic Editor: J. Soppa

A key element during the flow of genetic information in living systems is fidelity. The accuracy of DNA replication influences the genome size as well as the rate of genome evolution. The large amount of energy invested in gene expression implies that fidelity plays a major role in fitness. On the other hand, an increase in fidelity generally coincides with a decrease in velocity. Hence, an important determinant of the evolution of life has been the establishment of a delicate balance between fidelity and variability. This paper reviews the current knowledge on quality control in archaeal information processing. While the majority of these processes are homologous in Archaea, Bacteria, and Eukaryotes, examples are provided of nonorthologous factors and processes operating in the archaeal domain. In some instances, evidence for the existence of certain fidelity mechanisms has been provided, but the factors involved still remain to be identified.

1. Introduction

Francis Crick first announced his central dogma of molecular biology in 1958: the flow of sequential information that occurs in living cells, including replication of stored information (DNA), as well as expression of this information via messengers (mRNA) to functional proteins [1]. This dogma turned out to be a solid basis for molecular biology, although additional roles of (small) regulatory and metabolic RNA have been recognized more recently [2]. A key element during this transfer of genetic information is fidelity: the final accuracy depends on the combined error rates of the processes that constitute the whole chain.

From the ancient RNA world on, replication fidelity has been a major limiting factor of the amount of information stored. It has been proposed that on average less than one error per replicated genome is tolerated, as higher error rates lead to a so-called "error catastrophe" with a fatal amount of progeny not being viable [3–5]. The same rule applies also for extant cellular life in which double-stranded DNA is used for storage of genetic information. The increase in genome size was allowed by the increased stability of DNA [6] and by considerably lower error rates in DNA replication [7]. One might expect a continuous selection towards the highest possible fidelity. However, a very high level of fidelity in replication will negatively affect both the genome's adaptation potential, and the replication velocity and costs, posing the risk of being out-competed by more efficient rival organisms [8, 9]. Overall, the delicate balance between fidelity and mutation rate is in itself a trait of organisms and can differ between individuals and species [10]. For some species it is even known to change upon environmental signals and may vary between different locations within the same genome [11]. Fidelity of information processing is thus a major factor driving the evolution of cellular life.

Transcription and translation show significantly higher error rates than replication. Although the risk on affecting progeny is lower, erroneous gene expression might influence the error rate of replication indirectly, for example, when the replication machinery is affected [12]. On the one hand, inaccurate gene expression may lead to the production of nonfunctional proteins, and as such to a decreased fitness, that is, generating selective pressure for increasing fidelity. On the other hand, increasing the fidelity of transcription and translation also correlates with decreasing velocity, what also has an impact on fitness. Hence, natural evolution leaves a narrow range for varying the level of fidelity [13, 14].

In this paper, we will, whenever possible, focus on the systems of Archaea that contribute to accurate replication and expression of their genetic information. While the majority of the archaeal processes are well conserved in Bacteria

and/or Eukaryotes, a number of examples will be described of factors and processes that appear to be restricted to the archaeal domain. Despite the fact that research on Archaea is generally lagging behind that of the other two domains, the successful development of several Archaea as model organisms has recently lead to some first insight in their mechanisms to control fidelity of information processing.

2. Replication

Fidelity in replication is the result of three separate processes: (i) base selection, (ii) proofreading, and (iii) postsynthetic correction [15, 16]. These three processes contribute to very accurate DNA replication: incorporating a mistake only once every 10^6–10^{10} nucleotides for DNA-based microorganisms. Interestingly, the genomic mutation rate (the number of mutations per replicated genome) is quite constant for all DNA-based microorganisms, including bacteriophages, bacteria, and fungi: roughly 0.003-0.004 (Drake's rule [7]), what is largely below the above mentioned predicted upper limit of 1 error per replicated genome [4]. Surprisingly, it has recently been found that a thermophilic bacterium (*Thermus thermophilus*) and a thermophilic archaeum (*Sulfolobus acidocaldarius*) have error rates that are 5-fold lower, supporting the concept that there is an evolved balance between the need for fidelity and the cost of reducing the mutation rate [17].

After a brief description of polymerases in living systems, the three separate processes will be discussed in more detail. The last paragraph will discuss systems that organisms have evolved to overcome misincorporations.

2.1. DNA Polymerases. DNA is polymerized by DNA-dependent DNA polymerases (DNAPs) that can be classified into various families based upon their sequence similarity. Most replication-related DNAPs and primases belong to DNAP family B. Like Bacteria and Eukaryotes, Archaea contain multiple DNAPs. *Sulfolobus solfataricus*, for example, contains three family B DNAPs (B1 to B3) and one family Y DNAP (Dpo4) [18]. Crenarchaeota are restricted to family B for their replicative polymerases, while Euryarchaeota, Korarchaeota, Nanoarchaeota, and Thaumarchaeota use both a family B and a family D DNAP [19]. There is biochemical evidence that in these species the family B DNAP replicates the leading strand, while the family D DNAP replicates the lagging strand [20]. Deviation between leading and lagging strand replication has been found in other domains of life as well [21, 22]. Lagging strand replication involves Okazaki fragments that are produced by a lagging strand replicative DNAP, initially extending an RNA primer generated by a primase, a family B RNA polymerase. Archaea possess homologs of eukaryotic primase proteins (PriS and PriL) that can synthesize both RNA as DNA oligonucleotides in vitro, but seem to prefer RNA polymerization in vivo [23, 24]. Interestingly the B family replicative DNAPs of Archaea contain an uracil-specific pocket that scans the template for the presence of uracil ahead of the polymerase. This feature is apparently lost in eukaryotic and bacterial DNAPs, although they still possess the reminiscent pocket structure. If uracil is encountered the archaeal polymerase stalls, presumably until the uracil is removed by Base Excision Repair (BER) or until a Translesion Synthesis (TLS) DNAP takes over [25, 26]. TLS is a process in which the regular replicative DNAP is substituted by a translesion DNAP. Translesion polymerases, often family Y DNAPs, allow replication to occur past otherwise impassable DNA lesions. This adaptation however has led to a considerably lower fidelity than in case of replicative DNAPs. Dpo4 from *Sulfolobus solfataricus* is a family Y TLS DNAP. Dpo4 has a spacious solvent-exposed active site in comparison to replicative DNAPs that permits accurate bypass of the 8-oxoguanine oxidation product of guanine. 8-oxoguanine preferentially base-pairs to adenine, however in the active site a stabilizing hydrogen bond network fixes 8-oxoguanine in such position that the correct preference for cytosine is restored [15, 27].

2.2. Base Selection. The highest contribution to fidelity during DNA replication is brought about by base selection. Soon after the initial suggestion by Watson and Crick that selection was the result of hydrogen bonding of complementary bases [28], it became clear that the free-energy differences between correct and incorrect base-pairs could only account for error rates of approximately 0.01 [16]. Although the removal of water from the active site of the DNAP leads to elevated ΔG values, improving the selectivity between correct and incorrect base pairings [29], studies with base analogs that lost the capacity to create hydrogen bonds revealed the importance of base pair geometry. In addition, structural studies showed that a Watson-Crick pair, of which all four are nearly identical in shape and size, fits nicely into the base pair binding pocket of DNAP, while non-Watson-Crick base pairs presumably cause steric clashes (reviewed in [15, 30]).

2.3. Proofreading. Like the replicative DNAPs of the Bacteria and Eukaryotes, both family B as family D DNAPs from Archaea possess intrinsic proofreading capabilities [26, 31, 32]. Because these enzymes are thermostable, and have intrinsic proofreading, they are of commercialy interest as exemplified by the high-fidelity Pfu DNAP from *Pyrococcus furiosus* in polymerase chain reactions. Comparisons between wild-type polymerases with intrinsic proofreading capabilities and exonuclease-deficient mutants show that on average proofreading improves fidelity between 3–100 fold. For *Sulfolobus solfataricus* DNAP B1, the commercially available DNAP (Vent pol) from *Thermococcus litoralis* and their respective exonuclease-deficient mutants, it was measured to improve approximately 3 fold, a similar increase as observed for *E. coli* DNA pol III [15, 32].

DNAPs have prolonged interaction with the newly generated duplex DNA. Mismatches are recognized because of abnormal base pair geometry, and generally result in considerably decreased elongation rate. In DNAPs that have intrinsic or associated $3' \rightarrow 5'$ exonuclease activity, elongation rate drops below the exonuclease rate upon mismatch recognition, leading to removal of mismatched nucleotides. Polymerases without intrinsic exonuclease activity can either recruit another protein that has exonuclease activity, or

can dissociate and allow another polymerase with intrinsic exonuclease activity to take over.

Other errors generated during elongation, at approximately the same rate as mismatches, are single-base deletions and slightly less frequently single-base insertions. These "indels" can occur by (i) DNA strand slippage, (ii) misinsertion that is followed by primer relocation, or (iii) misalignment at the polymerase active site, and can occur especially at repetitive sequences. Whereas proofreading corrects mismatches at a high rate, this mechanism is relatively inefficient in correcting indels, especially if the repetitive elements are longer. Strand slippage, for example, occurs often upstream of the polymerase, is therefore not sensed and does not decrease the elongation rate, preventing the exonuclease activity from taking over (reviewed in [15, 30]).

2.4. Postsynthetic Correction. Mismatches or indels that slipped through the proofreading process, or that are introduced by mutagenic factors, are to be repaired by postsynthetic correction. Organisms generally have a set of distinct systems, designed to repair a specific class of damage, each with a different fidelity rate. A repair system directly connected to replication is Mismatch Repair (MMR). This system removes base substitutions and indels on the newly synthesized strand directly after replication. MMR increases fidelity of replication almost 100-fold [15]. In Bacteria and Eukaryotes, essential proteins required for MMR belong to the MutS and MutL family. These two families are largely absent in the archaeal domain. Archaeal homologs have only been found in some euryarchaeal species, probably the result of a horizontal gene transfer from bacterial origin [33]. Deletion mutants of a variety of MutS and MutL homologs in *Halobacterium salinarum,* including a MutS double mutant, had only little effect on mutation rates, indicating that these genes are not essential for MMR in this species [34]. A MutS2 ortholog is also present in the euryarchaeote *Pyrococcus furiosus* and it was shown to have ATPase and DNA binding activity, but no specific MMR activity [35]. Despite the general absence of MutS and MutL in Archaea, it is found that spontaneous base pair substitution rates in *S. acidocaldarius* are an order of a magnitude lower than MMR-proficient *E. coli* suggesting the existence of a powerful, yet unknown MMR system in Archaea [17].

During MMR, a key step is to identify which of the two strands is the (correct) parental strand and which one the (mutated) daughter. In some bacterial systems, the methylated strand is considered to be the parental strand, a signal for MutH to cleave opposite of a methylated GATC sequence near the mismatch [36]. Other Bacteria, Eukaryotes, and Archaea use other mechanisms to distinguish between the strands that are not yet fully understood. It is believed that in Eukaryotes the newly synthesized daughter strand contains discontinuities, caused by the separate Okazaki fragments during lagging strand replication and by reinitiation or low-level incorporation of dUMP during leading strand replication. Archaea may also use the incorporation of uracils as a marker for the daughter strand as well, in line with the fact that DNA replication in Archaea cannot pass uracils on the template strand [37].

2.5. Excision Repair. Two additional repair systems that repair single strand damage by using the complementary strand as a template include (i) Base Excision Repair (BER) used to remove regularly occurring small, nonhelix-distorting base lesions (e.g., modification by depurinations and deaminations) and involves DNA glycosylases, and (ii) Nucleotide Excision Repair (NER) used to remove bulky distortions in the helix (e.g., thymine dimers formed by oxidative stress or UV). The BER system appears to be functional in Archaea, as archaeal BER-related thermostable N-glycosylases have been characterized [38–42]. In contrast, the archaeal NER system appears to lack important damage-recognition proteins, but has structure-specific nucleases, homologous to eukaryotic NER nucleases [37]. UV stress experiments with *Sulfolobus acidocaldarius* show evidence for the existence of an archaeal NER system, as its repair capacity is at least half the capacity of NER-proficient *E. coli* [43, 44]. Especially life at elevated temperatures asks for efficient repair systems, as spontaneous decomposition reactions are accelerated under these conditions [6]. The high temperatures characterising the habitat of *Sulfolobus* species causes high rates of depurinations and deaminations. Although most of these types of damage are removed by BER, the apparent absence of key factors for both NER and MMR has been referred to as "the great irony" [37].

3. Transcription

During transcription mRNAs are generated by a DNA-dependent RNA polymerase (RNAP). The polymerization reactions of RNA and DNA show several similarities, for example, the course of nucleic acids through the active centre, and the mechanism of substrate binding, as reflected by the similar location of the two metal binding sites in the active sites of both polymerases [45]. Despite these similarities there are also several differences: (i) RNA polymerization incorporates NTPs instead of dNTPs, (ii) most RNAPs, with the exception of bacteriophage and mitochondrial RNAPs, are complexes that consist of 5–15 polypeptide subunits, in contrast to most DNAPs and primases that contain a single or only a few subunits, and (iii) while in DNAPs the newly formed DNA duplex persists, the newly formed RNA is removed from the DNA-RNA hybrid in RNAPs after which the original DNA duplex is restored [45].

Two processes are relevant in terms of transcription fidelity: base selection, and proofreading; post-synthetic correction of RNA does not exist, although some systems exist to monitor the quality of the transcripts that are used as templates during translation. These surveillance systems occur mainly during translation and will be discussed in that section (later). Although the fidelity of the transcription process is considerably lower than that of the replication process, it has been reported to be less than one error every 10^5 nucleotides that are being transcribed in organisms ranging from *E. coli* to wheat [46–48].

3.1. RNA Polymerases. The RNAP of Bacteria is a relatively simple complex consisting of 5 subunits. In addition, a set of up to 20 sigma factors allows for promoter selection in response to changing conditions. Eukaryotes use up to five variant RNAP complexes (I–V) that are responsible for transcription of distinct genes: ribosomal RNAs (RNAP I), protein-coding messenger RNAs (RNAP II), transfer RNAs, and other small noncoding RNAs (RNAP III). RNAP IV and RNAP V are restricted to plants and transcribe small RNAs involved in silencing [49]. RNAP I and III are similar to RNAP II, but have some additional subunits that vary between the two. Archaea, in contrast, have only a single RNAP complex that contains 12 orthologous subunits of the eukaryotic RNAP II. There appear to be minor variations among the complexes of the archaeal phyla [50, 51]. For instance, the RNAP from *Sulfolobus shibatae* has an additional subunit in comparison to the eukaryotic RNAP II (Rpo13) that has been proposed to play a role in the formation of the transcription bubble [52]. The subunits of these RNA polymerases can be assigned to three different functional groups: (i) the "catalytic core" (the large subunits A′A″, and B′B″; in some Archaea these subunits are fused as in Bacteria and Eukaryotes) that harbours the active site, (ii) the "assembly platform" (D, N, L, and P), and (iii) the "auxiliary subunits" (H, K, F, E, and Rpo13). The latter auxiliary set is the part of the complex that differs between the archaeal and the different eukaryotic RNAPs. These subunits that are not required for in vitro transcription, but important to stabilize interactions with RNA (F/E stalk), DNA (H and Rpo13), and transcription factors (F/E stalk). Additionally, the F/E stalk is found to be important for processivity during elongation, and correct recognition of weak terminators during termination [51, 53]. Recently it was shown that subunit H is required during promoter opening and initial transcription, and that it, in contrast to its eukaryotic counterpart Rpb5, undergoes a structural rearrangement in the transition from initiation complex to elongation complex that might be specific for archaeal RNAPs [54]. It was also shown recently that in vitro reconstitution of the archaeal RNAP is similar in the presence or absence of subunit P. Apparently it does not play a key role in establishing the assembly platform in vitro. In addition, subunit P seems to be involved in open complex formation [55]. Interestingly, a putative ortholog of Rpc34, which is a part of the eukaryotic RNAP III, has recently been found to be present in all crenarchaeal and thaumarchaeal genomes, as well as in several euryarchaeal genomes. This finding suggests that in Archaea the single RNAP might use a variable set-up of auxiliary proteins to transcribe different sets of transcripts [56]. Archaeal RNAPs can be reconstituted from single heterologously expressed subunits in contrast to eukaryotic RNAPs [57, 58]. Recent success with a hybrid archaeal enzyme that contain subunits Rpb5 and Rpb12 from Eukaryotes confirms the high structural similarity of the archaeal and the eukaryotic RNAPs [55, 59].

3.2. NTP Selection and Induced Fit. RNAPs discriminate NTPs over dNTPs by recognizing the 2′-hydroxyl group of incoming NTPs. Selection of NTPs by RNAPs is performed by measuring the base pair geometry, in a similar manner as in DNAPs, in a two step process. In the preinsertion state of the open active center the NTP can come in. If the NTP is complementary to the template nucleotide, the catalytic subunit undergoes a conformational change to the closed state, after which NTP is delivered to the insertion site. This rearranges the active site in such a way that it promotes polymerization by induced fit. If a noncomplementary nucleotide is incorporated, the complex enters an off-line state, in which elongation is slowed down considerably [60].

3.3. Proofreading. Incorporation of a noncomplementary nucleotide induces an inactivated state, in which the nucleotide is frayed. The fraying sites of the RNAP overlap with the NTP-binding site, and as such the frayed nucleotide does not allow elongation to proceed. This paused RNAP complex favours backtracking, a process in which the RNAPs moves one nucleotide backwards. During this process the misincorporated nucleotide is moved from the fraying site to the proofreading site. Multisubunit RNA polymerases contain an intrinsic nucleolytic RNA cleavage activity that hydrolyses a phosphodiester bond to remove the last two nucleotides as a dinucleotide, resulting in a new RNA 3′-OH group and an empty NTP-binding site. This restores an active on-line state ready for elongation again [60]. This process of backtracking and subsequent cleavage is transcriptional proofreading, and was also described in Archaea. In contrast to Bacteria and Eukaryotes, it was found that elongation in Archaea could not continue after misincorporation, but stalled completely instead. TFS, like its eukaryotic homolog TFIIS and its bacterial non-orthologous counterparts GreA/GreB, is known to induce the cleavage activity by direct interaction with the active centre of the polymerase through the nucleotide entrance pore, and could therefore rescue stalled elongation complexes. Stalling of the elongation complex in Archaea appears to be an important trigger for TFS induced cleavage in vitro [61]. *Methanopyrus kandleri* has lost TFS during its evolution. Interestingly, this organism shows a higher mutation rate in comparison with closely related organisms, making it difficult to reconstruct its phylogeny. Especially genes encoding proteins related to transcription are affected, and could include compensatory mutations for the loss of TFS [12].

4. Protein Synthesis

The overall missense substitution rate of in vivo bacterial protein synthesis by ribosomes is in the range of 6×10^{-4} to 5×10^{-3} per amino acid [62, 63]. In line with those findings are measurements of the rate of misreading in *Sulfolobus* in vitro translation systems: 3×10^{-3} incorrect leucine incorporations per amino acid on a poly(U) template [64]. Rates of misincorporations during replication, transcription, and aminoacyl-tRNA synthesis are all lower, showing that the final step, the translation process itself, is decisive with respect to fidelity of protein synthesis. The importance of fidelity during protein synthesis is reflected

in the organization and evolution of the genetic code. The presumable primordial genetic code that codes for an original set of 10 amino acids [65], as well as the 20 amino acid genetic code, operating in extant cellular life forms, are relatively robust, as most misincorporations will result in substitutions by physicochemically related amino acids that only in rare occasions will lead to a nonfunctional protein [65]. Fidelity was thus a key determinant in the evolution of the genetic code. Two separate processes are distinguished during protein synthesis: the coupling of amino acids to their respective tRNAs by a set of specific aminoacyl-tRNA synthetases, and the actual translation itself by ribosomes. In the next paragraphs both processes will be discussed, after which it will be concluded with an overview of the mRNA surveillance systems that are used to avoid the reuse of erroneous templates.

4.1. tRNA Modification. tRNA molecules are among the most strongly modified RNAs. This mainly concerns nucleotides that are located within the 3D-core and in the anticodon arm, especially at the wobble position N34 and N37 (conventional numbering). At present, over 120 different posttranscriptional modifications of nucleotides have been described, ranging from quite simple methylations to very complex multistep transformations [66]. These nucleotide modifications are important for cellular functionality of tRNAs: they lower conformational flexibility, improve (thermal) stability, and improve aminoacylation rate and specificity. Interestingly, it is known that lack of modification in in vitro translation systems can be compensated by excess of magnesium ions, indicating the importance to lower flexibility of tRNAs for translation (reviewed in [67]). Modifications of the wobble position N34 are common in all three domains of life and contribute to accuracy and efficacy of decoding during translation. These modifications are specific and vary between tRNAs. In contrast to unsplit codon boxes in the genetic code, tRNAs coding the split codon boxes are always modified at N34, suggesting that modifications play an important role in increasing the discriminative characteristics between near-identical codons. Remarkably, many modifications of N34 are restricted to specific phylogenetic Domains, or even to lower taxonomic groups, and come with an enormous diversity. This suggests that the corresponding modification enzymes evolved after the divergence of the three domains, and that the extension of the primordial code, and the accompanying increasing need for higher discrimination capacity, has led to a multitude of solutions (reviewed in [68]).

One of these wobble modifications in Bacteria and Eukaryotes is the conversion from G34 to queuosine. The replacement of guanine in this process is catalysed by the enzyme tRNA-guanine Transglycosylase. In Archaea, a related enzyme catalyses also the replacement step in the conversion from a guanine to a the positively charged archaeosine at position 15 [69]. G15 is part of the Levitt base-pair, which is the base-pair between N15 of the D-loop and N48 in the variable loop at the start of the T-loop, [70] and also interacts with N59 in the T-loop (Figure 2) [71]. These interactions between the D- and the T-loops establish the L-shape of the tRNA, indicating that formation of archaeosine is involved in stabilization of the RNA molecule. In Eukaryotes and Bacteria, where position 15 is not restricted to a G, and other variants of the Levitt base-pair exists, stabilization of the Levitt base-pair is brought about by Mg^{2+} binding. Interestingly, binding of a metal, which is less stable at high temperatures than chemical modification, is not compatible with archaeosine formation, suggesting distinct evolutionary mechanisms to stabilize the L-shaped structure of tRNAs between the Domains [72]. For modification of the deeply buried position 15, but probably also for other modifications, the tRNA has to adopt a different configuration, the λ-form. The energetics involved in such rearrangement suggest that modification enzymes might act together in a tRNA maturation complex [71]. Modifications in tRNAs are important for fidelity, processivity, and velocity of translation as they can directly affect decoding for example by modifications in the anticodon loop or in sites that are recognized by aminoacyl-tRNA synthetases, or indirectly by decreasing the flexibility and increasing the stability of the molecule.

4.2. Aminoacylation. The specific coupling of amino acids to their tRNAs yields aminoacyl-tRNAs (aa-tRNAs) and is catalyzed by specific aminoacyl-tRNA synthetases (aaRSs). Two classes (I and II) of aaRSs are distinguished on the basis of their structural topology of the active site [73]. Class I aaRSs, are generally monomeric, attach to the minor groove of the tRNA acceptor stem, and aminoacylate the terminal adenosine of the tRNA at the 2′-OH position, while Class II are generally multimeric, attach to the major groove, and aminoacylate the 3′-OH position [74]. Aminoacylation is a two-step process. First the amino acid is activated using ATP, forming the intermediate aminoacyl-adenylate. Once activated, the amino acid is transferred to the 3′ adenosine of the corresponding tRNA [74]. In Archaea and Eukaryotes, aaRSs are often organized in higher-order complexes that contain multiple aaRSs and other cellular factors, for example, the large multiaminoacyl-tRNA synthetase complex in *Haloarcula marismortui* that might harbour all aaRSs [75], or the LysRS-LeuRS-ProRS complex in *Methanothermobacter thermoautotrophicus* that increases the kinetics of LysRS and ProRS [76]. In Eukaryotes complex formation is sometimes also associated with other noncanonical functions like translational silencing, transcriptional control, or antiapoptosis (reviewed in [74]).

A cell might contain over 25 different types of aa-tRNAs [77]. For translation purposes there are 20 canonical elongator tRNAs, usually acylated by the corresponding synthetases, and an initiator aa-tRNA, acylated by methionyl tRNA synthetase. In Bacteria and eukaryotic organelles, the initiator Met-$tRNA^{Met}$ is subsequently formylated by a specific formyltransferase, in contrast to the situation in Eukaryotes and Archaea. In addition, a small number of noncanonical elongator tRNAs have been discovered (selenocysteinyl-tRNA, and pyrrolysyl-tRNA; see below). After coupling, aa-tRNAs are screened for their correctness by the translation elongation factor EF-Tu (eEF1A/aEF1α in Eukaryotes and Archaea) and delivered to the ribosome,

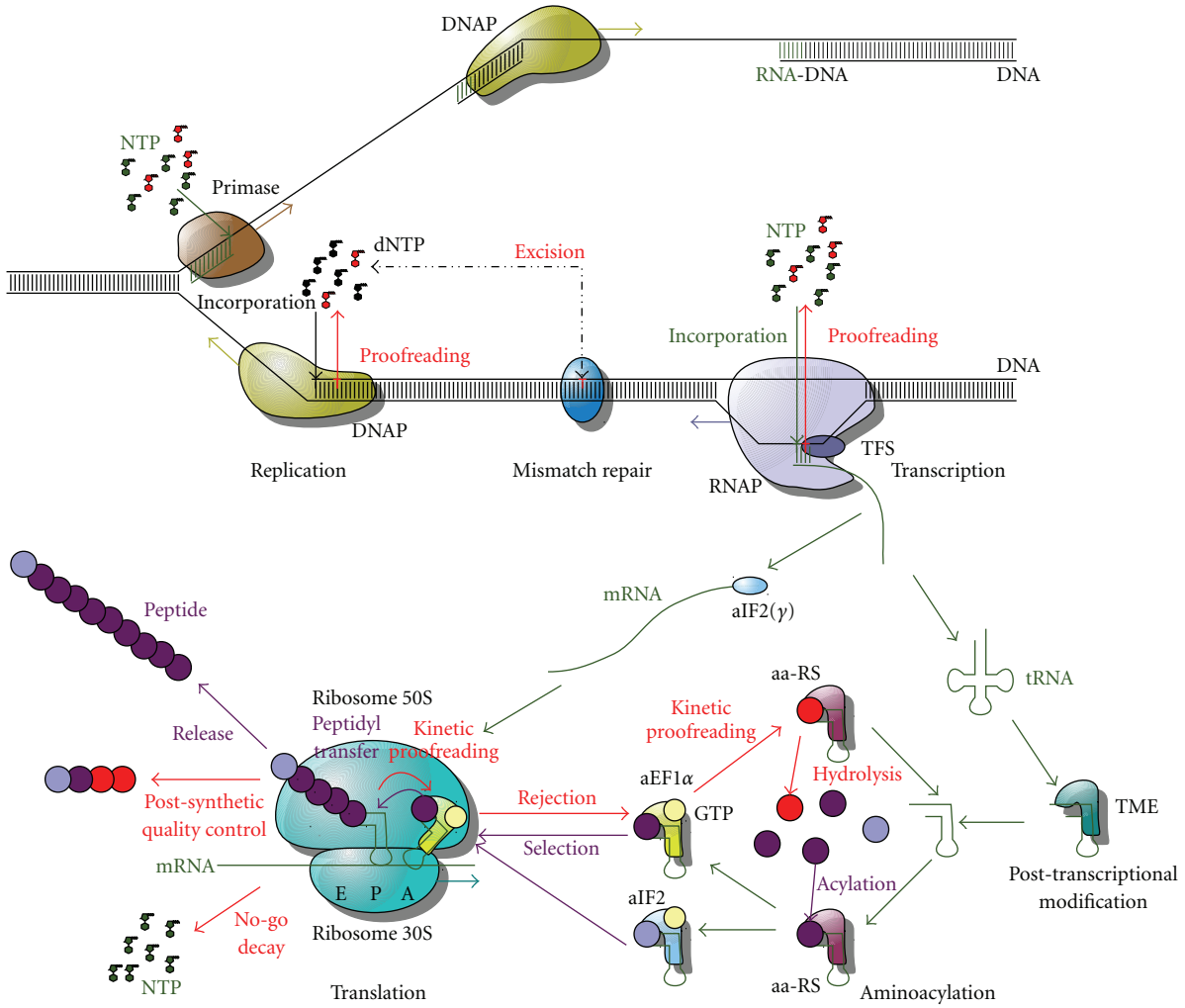


Figure 1: Overview of the processes involved in genetic Information Processing in Archaea (TFS: Transcription Factor S; TME: tRNA modifying enzymes; aa-RS: aminoacyl-tRNA synthetase; aIF2(γ): archaeal Initiation Factor 2(γ); aEF1α: archaeal Elongation Factor 1α).

with the exceptions of the initiator aa-tRNA that is verified and delivered by translation initiation factors, and selenocysteinyl-tRNA that is verified and delivered by SelB.

The second major group of aa-tRNAs is composed by misacylated translation substrates. A part is due to mistakes by the synthetases. Because elongation factor EF-Tu verifies aa-tRNAs before delivery to the ribosomes, and due to rapid editing by synthetases these errors are low: in most cases once in 10^6 events or less [78, 79]. aaRSs have special editing domains, which are located at a distant position from the synthetic domain, to decouple amino acids from misacylated tRNAs. It has been suggested that amino acid selection of the aaRSs depends on a double sieve mechanism, in which the substrate selection at the editing site is the inverse of the substrate selection at the synthetic site. For example, during coupling at the synthetic site, the amino acids larger than the cognate will be rejected. Then subsequent translocation to the editing site takes place where amino acids smaller than the cognate will be removed [80]. Unfortunately, this model is not complete as the editing site of some aaRSs can still edit on the basis of substrate selection present at the synthetic site. In a more recent model for class I aaRSs, it is proposed that the resting state of an aaRSs has the CCA of a bound tRNA at the editing site. When the intermediate aminoacyl-adenylate is formed in the synthetic site, the CCA of the tRNA is translocated to the synthetic site, allowing aminoacyl transfer from the adenylate to the CCA. After that the aminoacylated-CCA is translocated back to the editing site, allowing inspection, and subsequent hydrolysis or release of the aa-tRNA. This model uses two translocation actions providing the opportunity for kinetic proofreading (discussed later) [78]. Besides the editing domains available in aaRSs themselves, free-standing editing proteins, homologs to aaRSs that lack the acylation domain, also exists in all three domains [81–83].

In addition to accidentally misacylated tRNAs, there is also a group of aa-tRNAs that is deliberately misacylated by aminoacyl-tRNA synthetases, and are subjected to pretranslational amino acid modification. In a large number of Archaea, for example, *Methanothermobacter thermautotrophicus* [84], and Bacteria glutamate and aspartate are coupled to $tRNA^{Gln}$ and $tRNA^{Asn}$, respectively, by a

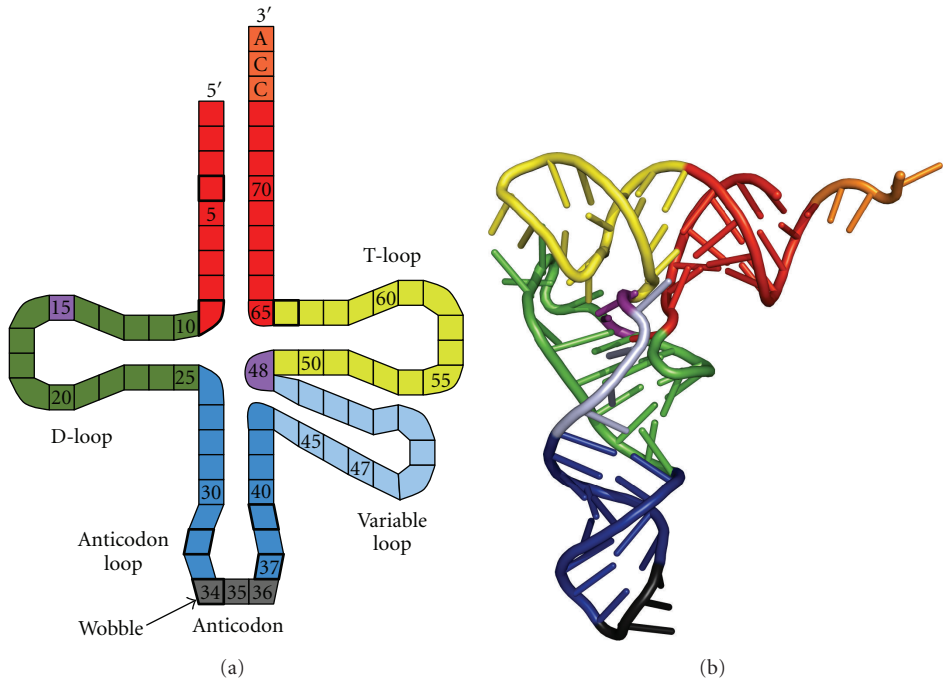


Figure 2: tRNAs. (a) Schematic representation, showing the D-loop (green), anticodon loop (blue) that harbours the anticodon (grey), variable loop that is variable in length (light blue), the T-loop (yellow), the acceptor stem (red), and the CCA aminoacyl binding site (orange). The Levitt base pair is coloured purple. Nucleotides with thick boxes are often modified with variable modifications. (b) Tertiary structure of a yeast $tRNA^{Phe}$, coloured similar to (a). Figure is rendered with PyMOL from data deposited in the Protein Data Bank (1 ehz).

nondiscriminating aaRS, and then converted by a tRNA amidotransferase into Gln-$tRNA^{Gln}$ and Asn-$tRNA^{Asn}$. Other deliberate mis-acylation pathways include cysteinyl-$tRNA^{Cys}$ (via *O*-phosphoseryl-$tRNA^{Cys}$) in methanogenic Archaea [85], and selenocysteinyl-$tRNA^{Sec}$ (via seryl-$tRNA^{Sec}$) [86] (reviewed in [77]).

4.3. Translation. Polymerization of amino acids is catalyzed by the ribosome, a large ribonucleoprotein complex that consists of 3-4 ribosomal RNAs and a large number of ribosomal proteins [87]. Archaeal translation is initiated by recognition of the small ribosomal subunit (30S) of an initiation codon, and the formation of the initiation complex, which includes the initiation factors, the initiator tRNA (Met-$tRNA^{Met}$) and mRNA. When the initiation complex is formed, the large subunit (50S) joins and the monomeric 70S ribosome is formed. Several mechanisms are known for initiation site recognition. Best known for prokaryotes is the mechanism that is associated with a Shine-Dalgarno (SD) motif that is recognized by the anti-SD motif on the 16S rRNA of the 30S. Although it is best known, it is not primarily used by all Bacteria or Archaea. *Sulfolobus* and *Pyrobaculum,* for example, use the SD mechanism only on distal cistrons of polycistronic transcripts, and not for the first cistron [88, 89]; in addition, *Haloarchaea,* hardly make use of this mechanism at all [90]. In Eukaryotes, that are devoid of the SD mechanism, the 40S cannot interact directly with mRNA, but needs mediation by the 5′-cap binding complex eIF4F. After binding of mRNA, it scans the RNA for an initiation codon by moving in the 3′ direction. Once located, the 60S joins the complex, the initiation factors leave, and elongation can start [91]. Less frequently, Eukaryotes use an IRES-dependent recognition mechanism, in which the complex IRES structures, that are located in the 5′-UTR, are recognized by IRES-binding transacting factors that are involved in recruitment of the small subunit [92]. All three domains of life also contain leaderless mRNAs, transcripts that start with 5′-terminal initiation codons, and that can be efficiently translated by all ribosomes regardless of the source [93, 94]. While leaderless transcripts are rare in Bacteria and Eukaryotes, they are abundant in many Archaeal species, being the primary mechanisms for monocistronic mRNAs and opening cistrons [88–90, 95]. It is thought that these leaderless transcripts are relics of primitive translation systems [93]. Recently, a novel mechanism has been identified in *Haloarchaea,* and although the exact molecular details are unknown, it has been demonstrated to act on transcripts that do not contain SD nor IRES motifs, however the efficiency of their translation depends on the 5′-UTR sequence involved [94].

On the basis of structural and chemical similarities between the homologous systems, translation elongation in Archaea is most likely very similar to that in Bacteria and Eukaryotes. Bacterial translation elongation occurs as follows. First, a ternary complex, which consist of an aminoacyl-tRNA, elongation factor EF-Tu (eEF1A/aEF1α in Eukaryotes/Archaea), and GTP is delivered to the Aminoacyl (A)-site. This complex reacts with the peptidyl-tRNA harboring the Peptidyl (P)-site. During this reaction, that is

discussed below in more detail, the peptidyl is transferred to the aminoacyl-tRNA, elongating the nascent chain by one amino acid. Third the peptidyl-tRNA in the A-site and the deacylated tRNA in the P-site move one position to the P and the Exit (E)-site respectively, leaving the A-site empty and ready for a new round. Energy for this translocation, in which also the accompanying mRNA moves accordingly, is delivered by GTP hydrolysis by EF-G (eEF2/aEF2 in Eukaryotes and Archaea). Accuracy of the ribosome depends on (i) kinetic proofreading, (ii) induced fit, and (iii) postpeptidyl transfer quality control that will be discussed in more detail in the next paragraphs (reviewed in [14]).

4.4. tRNA Selection by Kinetic Proofreading and Induced Fit. Kinetic proofreading is a mechanism that allows discrimination between small energetic differences with low error rates by repeated usage of those differences in distinct separate steps and by coupling them to high-energy intermediates. The error rate drops exponentially proportional to the number of repetitions [96, 97]. During translation elongation, the energetic difference between the codon and anticodon is measured first during the encounter between ribosome and the ternary complex (initial selection), and then again after hydrolysis of GTP, which is irreversible, when the ribosome associates with either the ternary complex (with GDP instead of GTP), or the free aminoacyl-tRNA when EF-Tu is dissociated (proofreading) [14].

Recent models based on bacteria show that decoding is composed out of seven steps [98]. (1) *Initial binding*: the exceptionally fast codon-independent interaction of the ternary complex to the ribosome is determined by EF-Tu and the ribosome, probably with a key role for the L7/L12 stalk. (2) *Codon recognition*: the formation of a complementary codon-anticodon at the decoding centre, what is reflected by a correct (presumably Watson-Crick) geometry, induces conformational changes in the 16S rRNA, while near-cognate geometry induces a different structural change that leads to an almost 1000-fold higher dissociation rate, although recognition rates remain almost similar. (3) *GTPase activation*: the GTP hydrolysis rate is increased by binding of cognate tRNAs compared to near-cognate binding. The local 16S rRNA conformational changes upon cognate binding (step (2)) lead to a closed conformation of the 30S ribosomal subunit. This conformational signal is communicated to the 50S ribosomal subunit and affects EF-Tu GTP hydrolysis. Near-cognate binding induces a different structural change in the decoding centre what most probably does not lead to the closed conformation of the 30S subunit, and thereby does not affect EF-Tu GTP hydrolysis. Slowing down hydrolysis increases discrimination capacity, however at the cost of velocity. (4) *GTP hydrolysis*: the rate of GTP hydrolysis by EF-Tu depends on the activation state of EF-Tu. (5) *Conformational change of EF-Tu*: EF-Tu changes from the GTP-form to the GDP form. This conformational change is limited by the rate of inorganic phosphate release. EF-Tu releases the aa-tRNA probably during the transition. (6) *Accommodation or rejection*: after release, the 3′ end of the aa-tRNA has to move almost 70 Å from its binding site on EF-Tu to the Peptidyl Transferase Centre (PTC), while the codon-anticodon interaction should remain intact. Accommodation in the PTC of cognate aa-tRNA is rapid and efficient, in contrast to near-cognate aa-tRNA that is mostly rejected because of the low stability of binding and lower rate for accommodation. (7) *Peptidyl-transfer*: the Peptidyl chain is transferred from the aa-tRNA on the P-site to the aa-tRNA on the A-site, elongating the nascent peptide with one amino acid. Initial selection occurs in steps (1) to (3), while proofreading occurs in step (6).

The ribosome not only uses kinetic proofreading to improve its selectivity it also uses an additional principle to further improve it: induced-fit. Induced-fit is a principle in which the correct substrate induces a conformational change leading to an acceleration of the desired process, while an incorrect substrate has the opposite effect. During decoding a correct codon-anticodon interaction accelerates both GTPase activation and accommodation steps, while a noncorrect near-cognate interaction inhibits both steps, leading to rejection [98]. It was reported that near-cognate tRNAs showed an increase in GTP consumption relative to the amount of amino acids incorporated, while other noncognate tRNAs did not. This suggests that noncognate tRNAs are already rejected during the second step whereas near-cognate are expelled during the fifth step after GTP hydrolysis, showing the importance of kinetic proofreading and induced fit for reliable discrimination between cognate and near-cognate tRNAs [99].

4.5. Postpeptidyl Transfer Quality Control. The molecular characteristics of Postpeptidyl transfer quality control (PPQC) have only recently been discovered in detail using bacterial in vitro systems [100], but it is likely to occur also in Eukaryotes and Archaea. Like proofreading in nucleotide polymerization, the ribosome senses mismatching after the polymerization reaction (peptidyl transfer in this case). However, where in replication and transcription exonucleases could erase the mistake, the ribosome should take more drastic actions to undo translation errors: abortive termination of the nascent peptide chain. Trigger for PPQC is mismatching between tRNA and template at the P-site of the ribosome. A mismatch at the P-site increases selection of noncognate tRNA at the A-site dramatically. After peptidyl transfer and subsequent translocation, the nascent chain contains two wrong subsequent amino acids, and both E-site as P-site harbour a mismatching tRNA. Mismatching at both E- and P-site leads to strongly stimulated release of the nascent peptide chain, increasing the rate constants for release in a range comparable to tRNA selection due to increased binding of Release Factors [100].

4.6. Termination. Translation terminates when a stop-codon reaches the A-site. Unlike other codons a stop-codon is recognized by proteins that mimic tRNAs: class-1 release factors (RF1s). These factors induce hydrolysis of peptidyl-tRNA, disconnecting the nascent chain from the tRNA. While Bacteria use two release factors (RF1 and RF2) that recognize different stop-codon pairs (UAA/UAG and UAA/UGA,

resp.), most Archaea and Eukaryotes have a single one (aRF1 and eRF1, resp.) that recognizes all three stop-codons. Results from experiments with genuine archaeal release factors and archaeal/eukaryotic chimeras in eukaryotic in vitro translation systems suggest similar mechanisms for both [101]. An interesting variant on this theme is found in pyrrolysine-utilizing Archaea. Pyrrolysine (Pyl), the 22nd amino acid, is only found in some archaeal species belonging to the Methanosarcinaceae and two Bacteria (*Desulfitobacterium hafniense* and an uncultured *δ-proteobacteria)* [102]. It is encoded by the amber stop codon (UAG). Pyl-tRNA^{Pyl} is normally recognized by EF-Tu, implicating normal incorporation during elongation [103]. *Methanosarcina barkeri* contains two RFs of which only one appears to be active in termination: it was found that aRF1-1 (at least when combined as a archaeal/eukaryotic chimera) had a lower release efficiency for the UAG codon than for UAA or UGA. Comparative genomics also showed that pyrrolysine-utilizing Archaea avoid UAG as a stop codon. This suggests that in these Archaea the genetic code is changed to incorporate Pyrrolysine instead of termination [101]. Reassigning stop codons is not restricted to Archaea: the Eukaryotic ciliates *Tetrahymena thermophila* and *Euplotes aediculatus* reassigned stop codons, UAG and UAA to glutamine, and UGA to cysteine, respectively, and changed specificity of their eRF1s accordingly [104, 105]. More prominent and present in all domains of life is selenocysteine (Sec), the 21st amino acid. In Archaea selenoproteins are found in *Methanococcus, Methanocaldococcus*, and *Methanopyrus* species [106]. Sec is encoded by the opal stop codon (UGA). However in contrast to Pyr incorporation, Sec incorporation needs a special elongation factor (SelB) that via an extended domain recognizes an mRNA hairpin loop downstream of a UGA codon (the selenocysteine insertion element or SECIS) [107]. Binding of Sec-tRNA^{Sec}-SelB-GTP to this structure leads to insertion of Sec at in-frame UGA codons. In contrast to Bacteria, the SECIS element is located outside of the coding region in Archaea and Eukaryotes, while the archaeal and eukaryotic SelB contain considerably shorter extensions. To overcome the distance, Eukaryotes evolved an additional adapter protein (SBP2). Additionally, it was found that the ribosomal protein L30 binds SECIS elements and influences Sec insertion. Although a similar mechanism is proposed for Archaea the adapter protein is not found [86].

4.7. mRNA Surveillance. In Eukaryotes mRNA quality control processes exist that act during translation to ensure the quality of the transcripts. These processes, called mRNA surveillance, are dependent on the eukaryotic release factors eRF1 and eRF3 and their paralogs Dom34 (synonym Pelota), and Hbs1 and Ski7, respectively. Three mRNA surveillance pathways are known in Eukaryotes. (i) Nonsense Mediated Decay (NMD): when premature stop codons are encountered, (ii) No-go Decay (NGD): to release stalled ribosomes, and (iii) Non-stop Decay (NSD): to rescue ribosomes that have read through a stop codon. NMD and NSD are restricted to Eukaryotes: as eRF3 and other components of the NMD system are missing in Archaea. Ski7, necessary for NSD, is even only present in the *Saccharomycetales,* although recent findings suggest that Hbs1 could take over what could mean that NSD is more widespread in Eukaryotes. In contrast, Dom34, necessary for No-go Decay is also found in Archaea. This suggests that NGD might be functionally present, although Hbs1p and eRF3 are missing in Archaea [108]. In Eukaryotes a ternary complex Dom34-Hbs1-GTP is formed, similar to the formation of the eRF1-eRF3-GTP complex used in eukaryotic translation termination. This Dom34-Hbs1-GTP ternary complex is able to recognize stalled ribosomes, leading to endonucleolytic cleavage of mRNA by Dom34 [109, 110]. In Archaea, aRF1 is able to terminate translation without the help of a RF3 ortholog, what could imply that the paralogous Dom34 might be able to perform NGD without the help of an Hbs1p ortholog [108].

To rescue stalled ribosomes, Bacteria have a system that uses an intermediate between tRNA and mRNA: tmRNA, an RNA molecule with a tertiary structure similar to tRNAs, but with an extended anticodon loop that contains an mRNA-like ORF. If the ribosome stalls, because a transcript is finished without a proper termination, tmRNA in concert with SmpB and EF-Tu binds to the empty A-site of the stalled ribosome. After translocation to the P-site, the mRNA-like ORF located in the anticodon loop of the tmRNA takes over the role of messenger, and encodes for a degradation tag and ends with a proper stop-codon. After release the nascent peptide is thus tagged for degradation, and the ribosomal subunits are released again. This system seems to be restricted to Bacteria as tmRNA genes have not been identified in Eukaryotes, with the small exception of a few eubacterial-like organelles, or Archaea [111]. Interestingly, in investigations of archaeal protein degradation in *Methanococcus jannaschii,* green fluorescent proteins tagged with a ssrA-extension were used. The ssrA extension is the 11 amino acid degradation tag encoded on the tmRNA, which gene was designated ssrA. Tagged proteins showed a rapid unfolding and degradation while untagged proteins did not [112].

5. Turnover of RNA and Proteins

5.1. RNA Decay. Beside above mentioned mRNA surveillance during translation, more general systems are involved in RNA turnover. Main component in these mechanisms in Archaea is the exosome, a protein complex that includes Rrp41 and Rrp42, a homolog of RNasePH, a bacterial phosphorolytic nuclease, and Rrp4 and Csl4, containing KH/S1 RNA-binding domains. The archaeal exosome is responsible for $3' \rightarrow 5'$ degradation of RNA, as well as for $3'$ polyadenylation. This complex is similar to the bacterial PNPase and the eukaryotic exosome. All three have a double-doughnut-like structure with a central hole with a core ring of six RNasePH-type subunits. The narrow neck of the archaeal structure only allows single-stranded RNA devoid of secondary structures, suggesting a regulatory role for cofactors, as observed in Eukaryotes [113–115].

Polyadenylation occurs mainly on fragmented molecules as part of an RNA decay pathway in Bacteria, Archaea, and eukaryotic cell organelles, and recently has been described for nuclear genes from Eukaryotes as well. Although, in contrast, in Eukaryotes poly(A) tails are also added to mature 3′ ends of most nuclear encoded, full-length, mRNAs for proper translation initiation, and mRNA stability. The general scheme of RNA 3′ → 5′ degradation in prokaryotes is as follows: (1) removal of the 5′ pyrophosphate, (2) endonucleolytic cleavage of the transcript, (3) poly-adenylation of cleavage products, and (4) rapid exonucleolytic degradation of polyadenylated products. In *Sulfolobus*, the exosome is able to generate a heteromeric poly(A)-rich tail and use NDPs as a substrate. It has been suggested that polyadenylation is used to overcome secondary RNA structures that otherwise cannot pass the exosome neck. Interestingly, halophilic Archaea, together with several methanogenic Archaea, like *Haloferax*, and *Methanococcus*, are the only known organisms that lack polyadenylation, and do not contain an exosome or PNPase. In these organisms poly(A)-independent RNA degradation is performed by RNase R [116–119].

Eukaryotes also use another pathway for mRNA degradation that involves 5′ → 3′ exonucleases, like Xrn1p. Eukaryotic transcripts are protected against this rapid form of decay by a 5′ cap. To prevent transcripts from being decapped unintentionally, they are protected by the eukaryotic translation initiation factor eIF4E [120]. In Archaea, mRNAs are similarly protected from 5′ → 3′ decay by binding of the γ-subunit of the archaeal translation initiation factor aIF2 to the 5′ end. The similarities between both systems suggests that 5′ → 3′ decay is common to all domains of life [121]. Additionally, this protection offers a mechanism to discriminate between new versus already translated transcripts. After translation aIF2 is removed from the mRNA, making the mRNA vulnerable to 5′ → 3′ decay as soon as translation is terminated. Interestingly, a tight coupling beyond the use of an initiation factor to protect mRNA, exists between transcription and translation in Archaea, as it has been found that multiple rounds of translation already start before transcription is finished [122]. This tight interplay might have to be extended to mRNA degradation as well, what would provide Archaea with a very efficient and short information processing pipeline.

5.2. The Protein Waste Bin. The 20S proteasome, present in Eukaryotes, Archaea, and actinobacteria, is a barrel-shaped complex that consists of four heptameric rings of α- and β-type subunits in an $\alpha 7\beta 7\beta 7\alpha 7$ configuration. Other Bacteria use the simpler HslV protease, that is structurally related to the β-type subunits of the 20S proteasomes. The function of the proteasome is to breakdown proteins into short peptides that in turn can be further degraded to amino acids by peptidases to be recycled in protein synthesis or in metabolism. The proteasome is therefore an essential component for protein turnover and to maintain protein quality control by degrading misfolded and denatured proteins [123, 124].

The protease domains of β-type subunits are located on the inside of the barrel. This creates a tightly regulated environment, to circumvent uncontrolled protein breakdown. In Eukaryotes, the 20S proteasome can be capped by 19S regulatory particles (a combination of a Rpt and Rpn proteins forming a base and a lit), on one side (26S proteasome), or on both sides (30S proteasome). These caps play a role in recognition and degradation of polyubiquitin tagged substrates. Archaea encode orthologs of Rpt called Proteasome-activating Nucleotidase (PAN). PAN is able to unfold proteins in a ATP-dependent manner, can open the axial gate of the 20S proteasome, and subsequently translocates the substrate into the 20S core. Interestingly, as mentioned earlier, the archaeal PAN is able to distinguish between a ssrA-tagged or untagged green fluorescent protein, which suggests a role for ssrA-tagging in peptide degradation in Archaea, although a tmRNA system responsible for ssrA-tagging has not been identified [112, 123, 124].

In Eukaryotes, ubiquitin and ubiquitin-like proteins, small stable proteins that contain a β-grasp fold and that can be attached to a wide variety of other proteins, play an important role in targeted degradation by the proteasome. Ubiquitylation is also used in a number of other nonproteolytic mechanisms like endocytosis, intracellular trafficking, chromatin-mediated regulation of transcription, and DNA repair. Discrimination between those target processes is thought to be dependent on the differences in ubiquitin chains [125]. Ubiquitin-targeted degradation is used for quality control in Eukaryotes. Misfolded proteins in the cytosol are recognized by chaperones, because of their toxic hydrophobic surfaces. These chaperones recruit ubiquitylation enzymes (e.g., CHIP), that attach a polyubiquitin chain to the misfolded protein after which it is degraded or refolded [126]. In the lumen of the endoplasmatic reticulum, N-linked oligosaccharides indicate the folding stage, but also appear to keep track of the time a polypeptide resides within the lumen. If misfolding occurs and the polypeptide is trapped within the lumen, they are directed to a ubiquitin ligase and targeted for destruction [127].

Although ubiquitin-like tagging was long thought to be restricted to Eukaryotes, an ubiquitin-like tagging system was recently revealed in *Haloferax volcanii* that is able to tag proteins with small archaeal modifier proteins (SAMPs). SAMPs are small proteins that contain a β-grasp fold and a C-terminal diglycine motif similar to ubiquitin, and are widespread among the Archaea. It was shown that SAMPs are coupled to a wide range of proteins. SAMP1 appears to target proteins for destruction by the proteasome [128]. Alternative signalling objectives might also be present, as SAMP2 was also found to be coupled to a wide range of proteins like SAMP1, but showed decreasing levels in proteasomal mutant strains [128]. It seems to be likely that systems similar to eukaryotic ubiquitin-targeted systems, like targeted destruction, are also present in the archaeal domain. Opening up a potential role for the proteasome in regulation of protein levels, quality control against misfolded proteins, and recycling of nonfunctional polypeptides in archaeal cells.

6. Concluding Remarks

It is obvious that a certain level of fidelity of genetic information processing is of major importance to the cell, in order to maintain the delicate balance between accuracy on the one hand and velocity on the other. At the moment, a rather complete picture is emerging for the three main polymerization reactions related to genetic information processing in living cells in general. More and more is known about the mechanisms and the role of factors that contribute to fidelity in these systems. To some extent these crucial cellular processes have successfully been studied in selected Archaea. Despite this progress, however, it is obvious that insight in fidelity-related mechanisms in Archaea is still relatively scarce. For that reason, extrapolations on the basis of analogous systems of Bacteria and Eukaryotes have been used in this overview to bridge the gaps in our understanding of the archaeal counterparts. Although we think that most of the described processes work similarly in Archaea, we cannot rule out that such generalisations may in some instances turn out to be an oversimplification of the actual situation. As many Archaea thrive in extreme environments, it will be very interesting to learn how fidelity mechanisms of these extremophilic organisms are adapted to overcome these harsh conditions. It is therefore anticipated that Archaea will continue to play an important role in future research to elucidate details on the intriguing systems that control the fidelity of information processing.

References

[1] F. Crick, "Central dogma of molecular biology," *Nature*, vol. 227, no. 5258, pp. 561–563, 1970.

[2] J. S. Mattick, "Challenging the dogma: the hidden layer of non-protein-coding RNAs in complex organisms," *BioEssays*, vol. 25, no. 10, pp. 930–939, 2003.

[3] M. Eigen, "Selforganization of matter and the evolution of biological macromolecules," *Die Naturwissenschaften*, vol. 58, no. 10, pp. 465–523, 1971.

[4] M. Eigen and P. Schuster, "The hypercycle. A principle of natural self organization. Part A: emergence of the hypercycle," *Naturwissenschaften*, vol. 64, no. 11, pp. 541–565, 1977.

[5] Y. I. Wolf and E. V. Koonin, "On the origin of the translation system and the genetic code in the RNA world by means of natural selection, exaptation, and subfunctionalization," *Biology Direct*, vol. 2, article 14, 2007.

[6] T. Lindahl, "Instability and decay of the primary structure of DNA," *Nature*, vol. 362, no. 6422, pp. 709–715, 1993.

[7] J. W. Drake, B. Charlesworth, D. Charlesworth, and J. F. Crow, "Rates of spontaneous mutation," *Genetics*, vol. 148, no. 4, pp. 1667–1686, 1998.

[8] J. A. G. M. De Visser and D. E. Rozen, "Limits to adaptation in asexual populations," *Journal of Evolutionary Biology*, vol. 18, no. 4, pp. 779–788, 2005.

[9] V. Furió, A. Moya, and R. Sanjuán, "The cost of replication fidelity in human immunodeficiency virus type 1," *Proceedings of the Royal Society B*, vol. 274, no. 1607, pp. 225–230, 2007.

[10] E. Loh, J. J. Salk, and L. A. Loeb, "Optimization of DNA polymerase mutation rates during bacterial evolution," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 107, no. 3, pp. 1154–1159, 2010.

[11] D. Metzgar and C. Wills, "Evolutionary changes in mutation rates and spectra and their influence on the adaptation of pathogens," *Microbes and Infection*, vol. 2, no. 12, pp. 1513–1522, 2000.

[12] C. Brochier, P. Forterre, and S. Gribaldo, "Archaeal phylogeny based on proteins of the transcription and translation machineries: tackling the Methanopyrus kandleri paradox," *Genome Biology*, vol. 5, no. 3, p. R17, 2004.

[13] R. C. Thompson and A. M. Karim, "The accuracy of protein biosynthesis is limited by its speed: high fidelity selection by ribosomes of aminoacyl-tRNA ternary complexes containing GTP[γS]," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 79, no. 16, pp. 4922–4926, 1982.

[14] H. S. Zaher and R. Green, "Fidelity at the molecular level: lessons from protein synthesis," *Cell*, vol. 136, no. 4, pp. 746–762, 2009.

[15] T. A. Kunkel, "Evolving views of DNA replication (In)fidelity," *Cold Spring Harbor Symposia on Quantitative Biology*, vol. 74, pp. 91–101, 2009.

[16] L. A. Loeb and T. A. Kunkel, "Fidelity of DNA synthesis," *Annual Review of Biochemistry*, vol. 51, pp. 429–457, 1982.

[17] J. W. Drake, "Avoiding dangerous missense: thermophiles display especially low mutation rates," *PLoS Genetics*, vol. 5, no. 6, Article ID e1000520, 2009.

[18] Q. She, R. K. Singh, F. Confalonieri et al., "The complete genome of the crenarchaeon *Sulfolobus solfataricus* P2," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 98, no. 14, pp. 7835–7840, 2001.

[19] T. H. Tahirov, K. S. Makarova, I. B. Rogozin, Y. I. Pavlov, and E. V. Koonin, "Evolution of DNA polymerases: an inactivated polymerase-exonuclease module in Pol ε and a chimeric origin of eukaryotic polymerases from two classes of archaeal ancestors," *Biology Direct*, vol. 4, article 11, 2009.

[20] G. Henneke, D. Flament, U. Hübscher, J. Querellou, and J.-P. Raffin, "The hyperthermophilic euryarchaeota *Pyrococcus abyssi* likely requires the two DNA polymerases D and B for DNA replication," *Journal of Molecular Biology*, vol. 350, no. 1, pp. 53–64, 2005.

[21] T. Fukui, K. Yamauchi, T. Muroya et al., "Distinct roles of DNA polymerases delta and epsilon at the replication fork in *Xenopus* egg extracts," *Genes to Cells*, vol. 9, no. 3, pp. 179–191, 2004.

[22] G. M. Sanders, H. G. Dallmann, and C. S. McHenry, "Reconstitution of the *B. subtilis* replisome with 13 proteins including two distinct replicases," *Molecular Cell*, vol. 37, no. 2, pp. 273–281, 2010.

[23] S.-H. Lao-Sirieix and S. D. Bell, "The heterodimeric primase of the hyperthermophilic archaeon *Sulfolobus solfataricus* possesses DNA and RNA primase, polymerase and 3′-terminal Nucleotidyl transferase activities," *Journal of Molecular Biology*, vol. 344, no. 5, pp. 1251–1263, 2004.

[24] L. Liu, K. Komori, S. Ishino et al., "The archaeal DNA primase: biochemical characterization of the p41-p46 complex from *Pyrococcus furiosus*," *The Journal of Biological Chemistry*, vol. 276, no. 48, pp. 45484–45490, 2001.

[25] E. R. Barry and S. D. Bell, "DNA replication in the archaea," *Microbiology and Molecular Biology Reviews*, vol. 70, no. 4, pp. 876–887, 2006.

[26] H. J. Russell, T. T. Richardson, K. Emptage, and B. A. Connolly, "The 3′-5′ proofreading exonuclease of archaeal family-B DNA polymerase hinders the copying of template

strand deaminated bases," *Nucleic Acids Research*, vol. 37, no. 22, pp. 7603–7611, 2009.

[27] S. Broyde, L. Wang, O. Rechkoblit, N. E. Geacintov, and D. J. Patel, "Lesion processing: high-fidelity versus lesion-bypass DNA polymerases," *Trends in Biochemical Sciences*, vol. 33, no. 5, pp. 209–219, 2008.

[28] J. D. Watson and F. H. C. Crick, "Genetical implications of the structure of deoxyribonucleic acid," *Nature*, vol. 171, no. 4361, pp. 964–967, 1953.

[29] J. Petruska, L. C. Sowers, and M. F. Goodman, "Comparison of nucleotide interactions in water, proteins, and vacuum: model for DNA polymerase fidelity," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 83, no. 6, pp. 1559–1562, 1986.

[30] T. A. Kunkel and K. Bebenek, "DNA replication fidelity," *Annual Review of Biochemistry*, vol. 69, pp. 497–529, 2000.

[31] M. Jokela, A. Eskelinen, H. Pospiech, J. Rouvinen, and J. E. Syväoja, "Characterization of the 3′ exonuclease subunit DP1 of *Methanococcus jannaschii* replicative DNA polymerase D," *Nucleic Acids Research*, vol. 32, no. 8, pp. 2430–2440, 2004.

[32] L. Zhang, H. Lou, L. Guo et al., "Accurate DNA synthesis by *Sulfolobus solfataricus* DNA polymerase B1 at high temperature," *Extremophiles*, vol. 14, no. 1, pp. 107–117, 2009.

[33] Z. Lin, M. Nei, and H. Ma, "The origins and early evolution of DNA mismatch repair genes—multiple horizontal gene transfers and co-evolution," *Nucleic Acids Research*, vol. 35, no. 22, pp. 7591–7603, 2007.

[34] C. R. Busch and J. DiRuggiero, "MutS and MutL are dispensable for maintenance of the genomic mutation rate in the halophilic archaeon *Halobacterium salinarum* NRC-1," *PLoS ONE*, vol. 5, no. 2, Article ID e9045, 2010.

[35] R. Vijayvargia and I. Biswas, "MutS2 family protein from *Pyrococcus furiosus*," *Current Microbiology*, vol. 44, no. 3, pp. 224–228, 2002.

[36] K. G. Au, K. Welsh, and P. Modrich, "Initiation of methyl-directed mismatch repair," *The Journal of Biological Chemistry*, vol. 267, no. 17, pp. 12142–12148, 1992.

[37] D. W. Grogan, "Stability and repair of DNA in hyperthermophilic archaea," *Current Issues in Molecular Biology*, vol. 6, no. 2, pp. 137–144, 2004.

[38] N.-K. Birkeland, H. Ånensen, I. Knævelsrud et al., "Methylpurine DNA glycosylase of the hyperthermophilic archaeon *Archaeoglobus fulgidus*," *Biochemistry*, vol. 41, no. 42, pp. 12697–12705, 2002.

[39] S. Kiyonari, S. Tahara, M. Uchimura, T. Shirai, S. Ishino, and Y. Ishino, "Studies on the base excision repair (BER) complex in *Pyrococcus furiosus*," *Biochemical Society Transactions*, vol. 37, no. 1, pp. 79–82, 2009.

[40] S. Kiyonari, S. Tahara, T. Shirai, S. Iwai, S. Ishino, and Y. Ishino, "Biochemical properties and base excision repair complex formation of apurinic/apyrimidinic endonuclease from *Pyrococcus furiosus*," *Nucleic Acids Research*, vol. 37, no. 19, pp. 6439–6453, 2009.

[41] I. Knævelsrud, P. Ruoff, H. Ånensen, A. Klungland, S. Bjelland, and N.-K. Birkeland, "Excision of uracil from DNA by the hyperthermophilic Afung protein is dependent on the opposite base and stimulated by heat-induced transition to a more open structure," *Mutation Research*, vol. 487, no. 3-4, pp. 173–190, 2001.

[42] A. A. Sartori, P. Schär, S. Fitz-Gibbon, J. H. Miller, and J. Jiricny, "Biochemical characterization of uracil processing activities in the hyperthermophilic archaeon *Pyrobaculum aerophilum*," *The Journal of Biological Chemistry*, vol. 276, no. 32, pp. 29979–29986, 2001.

[43] K. J. Schmidt, K. E. Beck, and D. W. Grogan, "UV stimulation of chromosomal marker exchange in *Sulfolobus acidocaldarius*: implications for DNA repair, conjugation and homologous recombination at extremely high temperatures," *Genetics*, vol. 152, no. 4, pp. 1407–1415, 1999.

[44] E. R. Wood, F. Ghané, and D. W. Grogan, "Genetic responses of the thermophilic archaeon *Sulfolobus acidocaldarius* to short-wavelength UV light," *Journal of Bacteriology*, vol. 179, no. 18, pp. 5693–5698, 1997.

[45] P. Cramer, "Common structural features of nucleic acid polymerases," *BioEssays*, vol. 24, no. 8, pp. 724–729, 2002.

[46] A. Blank, J. A. Gallant, R. R. Burgess, and L. A. Loeb, "An RNA polymerase mutant with reduced accuracy of chain elongation," *Biochemistry*, vol. 25, no. 20, pp. 5920–5928, 1986.

[47] R. F. Rosenberger and G. Foskett, "An estimate of the frequency of in vivo transcriptional errors at a nonsense codon in *Escherichia coli*," *Molecular and General Genetics*, vol. 183, no. 3, pp. 561–563, 1981.

[48] L. de Mercoyrol, Y. Corda, C. Job, and D. Job, "Accuracy of wheat-germ RNA polymerase II. General enzymatic properties and effect of template conformational transition from right-handed B-DNA to left-handed Z-DNA," *European Journal of Biochemistry*, vol. 206, no. 1, pp. 49–58, 1992.

[49] R. Landick, "Functional divergence in the growing family of RNA polymerases," *Structure*, vol. 17, no. 3, pp. 323–325, 2009.

[50] E. V. Koonin, K. S. Makarova, and J. G. Elkins, "Orthologs of the small RPB8 subunit of the eukaryotic RNA polymerases are conserved in hyperthermophilic Crenarchaeota and "Korarchaeota"," *Biology Direct*, vol. 2, article 38, 2007.

[51] F. Werner, "Structure and function of archaeal RNA polymerases," *Molecular Microbiology*, vol. 65, no. 6, pp. 1395–1404, 2007.

[52] Y. Korkhin, U. M. Unligil, O. Littlefield et al., "Evolution of complex RNA polymerases: the complete archaeal RNA polymerase structure," *PLoS Biology*, vol. 7, no. 5, Article ID e1000102, 2009.

[53] A. Hirtreiter, D. Grohmann, and F. Werner, "Molecular mechanisms of RNA polymerase—the F/E (RPB4/7) complex is required for high processivity in vitro," *Nucleic Acids Research*, vol. 38, no. 2, pp. 585–596, 2009.

[54] S. Grünberg, C. Reich, M. E. Zeller, M. S. Bartlett, and M. Thomm, "Rearrangement of the RNA polymerase subunit H and the lower jaw in archaeal elongation complexes," *Nucleic Acids Research*, vol. 38, no. 6, pp. 1950–1963, 2009.

[55] C. Reich, M. Zeller, P. Milkereit et al., "The archaeal RNA polymerase subunit P and the eukaryotic polymerase subunit Rpb12 are interchangeable in vivo and in vitro," *Molecular Microbiology*, vol. 71, no. 4, pp. 989–1002, 2009.

[56] F. Blombach, K. S. Makarova, J. Marrero, B. Siebers, E. V. Koonin, and J. V. D. Oost, "Identification of an ortholog of the eukaryotic RNA polymerase III subunit RPC34 in Crenarchaeota and Thaumarchaeota suggests specialization of RNA polymerases for coding and non-coding RNAs in Archaea," *Biology Direct*, vol. 4, article 1745, p. 39, 2009.

[57] S. Naji, S. Grünberg, and M. Thomm, "The RPB7 orthologue E′ is required for transcriptional activity of a reconstituted archaeal core enzyme at low temperatures and stimulates open complex formation," *The Journal of Biological Chemistry*, vol. 282, no. 15, pp. 11047–11057, 2007.

[58] F. Werner and R. O. J. Weinzierl, "A recombinant RNA polymerase II-like enzyme capable of promoter-specific

transcription," *Molecular Cell*, vol. 10, no. 3, pp. 635–646, 2002.

[59] M. Thomm, C. Reich, S. Grünberg, and S. Naji, "Mutational studies of archaeal RNA polymerase and analysis of hybrid RNA polymerases," *Biochemical Society Transactions*, vol. 37, no. 1, pp. 18–22, 2009.

[60] J. F. Sydow and P. Cramer, "RNA polymerase fidelity and transcriptional proofreading," *Current Opinion in Structural Biology*, vol. 19, no. 6, pp. 732–739, 2009.

[61] U. Lange and W. Hausner, "Transcriptional fidelity and proofreading in Archaea and implications for the mechanism of TFS-induced RNA cleavage," *Molecular Microbiology*, vol. 52, no. 4, pp. 1133–1143, 2004.

[62] F. Bouadloun, D. Donner, and C. G. Kurland, "Codon-specific missense errors in vivo," *EMBO Journal*, vol. 2, no. 8, pp. 1351–1356, 1983.

[63] P. Edelmann and J. Gallant, "Mistranslation in *E. coli*," *Cell*, vol. 10, no. 1, pp. 131–137, 1977.

[64] P. Londei, S. Altamura, J. L. Sanz, and R. Amils, "Aminoglycoside-induced mistranslation in thermophilic archaebacteria," *MGG Molecular & General Genetics*, vol. 214, no. 1, pp. 48–54, 1988.

[65] A. S. Novozhilov and E. V. Koonin, "Exceptional error minimization in putative primordial genetic codes," *Biology Direct*, vol. 4, article 44, 2009.

[66] A. Czerwoniec, S. Dunin-Horkawicz, E. Purta et al., "MODOMICS: a database of RNA modification pathways. 2008 update," *Nucleic Acids Research*, vol. 37, no. 1, pp. D118–D121, 2009.

[67] M. Helm, "Post-transcriptional nucleotide modification and alternative folding of RNA," *Nucleic Acids Research*, vol. 34, no. 2, pp. 721–733, 2006.

[68] H. Grosjean, V. de Crécy-Lagard, and C. Marck, "Deciphering synonymous codons in the three domains of life: co-evolution with specific tRNA modification enzymes," *FEBS Letters*, vol. 584, no. 2, pp. 252–264, 2010.

[69] J. Sabina and D. Söll, "The RNA-binding PUA domain of archaeal tRNA-guanine transglycosylase is not required for archaeosine formation," *The Journal of Biological Chemistry*, vol. 281, no. 11, pp. 6993–7001, 2006.

[70] M. Levitt, "Detailed molecular model for transfer ribonucleic acid," *Nature*, vol. 224, no. 5221, pp. 759–763, 1969.

[71] R. Ishitani, O. Nureki, N. Nameki, N. Okada, S. Nishimura, and S. Yokoyama, "Alternative tertiary structure of tRNA for recognition by a posttranscriptional modification enzyme," *Cell*, vol. 113, no. 3, pp. 383–394, 2003.

[72] R. Oliva, A. Tramontano, and L. Cavallo, "Mg^{2+} binding and archaeosine modification stabilize the G15-C48 Levitt base pair in tRNAs," *RNA*, vol. 13, no. 9, pp. 1427–1436, 2007.

[73] G. Eriani, M. Delarue, O. Poch, J. Gangloff, and D. Moras, "Partition of tRNA synthetases into two classes based on mutually exclusive sets of sequence motifs," *Nature*, vol. 347, no. 6289, pp. 203–206, 1990.

[74] C. D. Hausmann and M. Ibba, "Aminoacyl-tRNA synthetase complexes: molecular multitasking revealed," *FEMS Microbiology Reviews*, vol. 32, no. 4, pp. 705–721, 2008.

[75] Y. Goldgur and M. Safro, "Aminoacyl-tRNA synthetases from Haloarcula marismortui: an evidence for a multienzyme complex in a procaryotic system," *Biochemistry and Molecular Biology International*, vol. 32, no. 6, pp. 1075–1083, 1994.

[76] M. Prætorius-Ibba, C. D. Hausmann, M. Paras, T. E. Rogers, and M. Ibba, "Functional association between three archaeal aminoacyl-tRNA synthetases," *The Journal of Biological Chemistry*, vol. 282, no. 6, pp. 3680–3687, 2007.

[77] M. Ibba and D. Söll, "Aminoacyl-tRNAs: setting the limits of the genetic code," *Genes and Development*, vol. 18, no. 7, pp. 731–738, 2004.

[78] C. S. Francklyn, "DNA polymerases and aminoacyl-tRNA synthetases: shared mechanisms for ensuring the fidelity of gene expression," *Biochemistry*, vol. 47, no. 45, pp. 11695–11703, 2008.

[79] H. Jakubowski and E. Goldman, "Editing of errors in selection of amino acids for protein synthesis," *Microbiological Reviews*, vol. 56, no. 3, pp. 412–429, 1992.

[80] A. R. Fersht, "Enzymic editing mechanisms and the genetic code," *Proceedings of the Royal Society of London*, vol. 212, no. 1189, pp. 351–379, 1981.

[81] D. Korencic, I. Ahel, J. Schelert et al., "A freestanding proofreading domain is required for protein synthesis quality control in Archaea," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 101, no. 28, pp. 10260–10265, 2004.

[82] F.-C. Wong, P. J. Beuning, C. Silvers, and K. Musier-Forsyth, "An isolated class II aminoacyl-tRNA synthetase insertion domain is functional in amino acid editing," *The Journal of Biological Chemistry*, vol. 278, no. 52, pp. 52857–52864, 2003.

[83] I. Ahel, D. Korencic, M. Ibba, and D. Söll, "Trans-editing of mischarged tRNAs," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 100, no. 26, pp. 15422–15427, 2003.

[84] H. Oshikane, K. Sheppard, S. Fukai et al., "Structural basis of RNA-dependent recruitment of glutamine to the genetic code," *Science*, vol. 312, no. 5782, pp. 1950–1954, 2006.

[85] A. Sauerwald, W. Zhu, T. A. Major et al., "RNA-dependent cysteine biosynthesis in archaea," *Science*, vol. 307, no. 5717, pp. 1969–1972, 2005.

[86] T. Stock and M. Rother, "Selenoproteins in Archaea and Gram-positive bacteria," *Biochimica et Biophysica Acta*, vol. 1790, no. 11, pp. 1520–1532, 2009.

[87] O. Lecompte, R. Ripp, J.-C. Thierry, D. Moras, and O. Poch, "Comparative analysis of ribosomal proteins in complete genomes: an example of reductive evolution at the domain scale," *Nucleic Acids Research*, vol. 30, no. 24, pp. 5382–5390, 2002.

[88] D. Benelli, E. Maone, and P. Londei, "Two different mechanisms for ribosome/mRNA interaction in archaeal translation initiation," *Molecular Microbiology*, vol. 50, no. 2, pp. 635–643, 2003.

[89] M. M. Slupska, A. G. King, S. Fitz-Gibbon, J. Besemer, M. Borodovsky, and J. H. Miller, "Leaderless transcripts of the crenarchaeal hyperthermophile *Pyrobaculum aerophilum*," *Journal of Molecular Biology*, vol. 309, no. 2, pp. 347–360, 2001.

[90] M. Brenneis, O. Hering, C. Lange, and J. Soppa, "Experimental characterization of Cis-acting elements important for translation and transcription in halophilic archaea.," *PLoS Genetics*, vol. 3, no. 12, p. e229, 2007.

[91] M. Kozak, "Pushing the limits of the scanning mechanism for initiation of translation," *Gene*, vol. 299, no. 1-2, pp. 1–34, 2002.

[92] M. López-Lastra, A. Rivas, and M. I. Barría, "Protein synthesis in eukaryotes: the growing biological relevance of cap-independent translation initiation," *Biological Research*, vol. 38, no. 2-3, pp. 121–146, 2005.

[93] D. Benelli and P. Londei, "Begin at the beginning: evolution of translational initiation," *Research in Microbiology*, vol. 160, no. 7, pp. 493–501, 2009.

[94] O. Hering, M. Brenneis, J. Beer, B. Suess, and J. Soppa, "A novel mechanism for translation initiation operates in haloarchaea," *Molecular Microbiology*, vol. 71, no. 6, pp. 1451–1463, 2009.

[95] O. Wurtzel, R. Sapra, F. Chen, Y. Zhu, B. A. Simmons, and R. Sorek, "A single-base resolution map of an archaeal transcriptome," *Genome Research*, vol. 20, no. 1, pp. 133–141, 2010.

[96] J. J. Hopefield, "Kinetic proofreading: a new mechanism for reducing errors in biosynthetic processes requiring high specificity," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 71, no. 10, pp. 4135–4139, 1974.

[97] J. Ninio, "Kinetic amplification of enzyme discrimination," *Biochimie*, vol. 57, no. 5, pp. 587–595, 1975.

[98] M. V. Rodnina, K. B. Gromadski, U. Kothe, and H.-J. Wieden, "Recognition and selection of tRNA in translation," *FEBS Letters*, vol. 579, no. 4, pp. 938–942, 2005.

[99] R. C. Thompson and P. J. Stone, "Proofreading of the codon anticodon interaction on ribosomes," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 74, no. 1, pp. 198–202, 1977.

[100] H. S. Zaher and R. Green, "Quality control by the ribosome following peptide bond formation," *Nature*, vol. 457, no. 7226, pp. 161–166, 2009.

[101] E. Alkalaeva, B. Eliseev, A. Ambrogelly et al., "Translation termination in pyrrolysine-utilizing archaea," *FEBS Letters*, vol. 583, no. 21, pp. 3455–3460, 2009.

[102] M. M. Lee, R. Jiang, R. Jain, R. C. Larue, J. Krzycki, and M. K. Chan, "Structure of Desulfitobacterium hafniense PylSc, a pyrrolysyl-tRNA synthetase," *Biochemical and Biophysical Research Communications*, vol. 374, no. 3, pp. 470–474, 2008.

[103] A. Théobald-Dietrich, M. Frugier, R. Giegé, and J. Rudinger-Thirion, "Atypical archaeal tRNA pyrrolysine transcript behaves towards EF-Tu as a typical elongator tRNA," *Nucleic Acids Research*, vol. 32, no. 3, pp. 1091–1096, 2004.

[104] K. Ito, L. Frolova, A. Seit-Nebi, A. Karamyshev, L. Kisselev, and Y. Nakamura, "Omnipotent decoding potential resides in eukaryotic translation termination factor eRF1 of variant-code organisms and is modulated by the interactions of amino acid sequences within domain 1," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 99, no. 13, pp. 8494–8499, 2002.

[105] S. Kervestin, L. Frolova, L. Kisselev, and O. Jean-Jean, "Stop codon recognition in ciliates: euplotes release factor does not respond to reassigned UGA codon," *EMBO Reports*, vol. 2, no. 8, pp. 680–684, 2001.

[106] T. Stock, M. Selzer, and M. Rother, "In vivo requirement of selenophosphate for selenoprotein synthesis in archaea," *Molecular Microbiology*, vol. 75, no. 1, pp. 149–160, 2010.

[107] M. Leibundgut, C. Frick, M. Thanbichler, A. Böck, and N. Ban, "Selenocysteine tRNA-specific elongation factor SelB is a structural chimaera of elongation and initiation factors," *EMBO Journal*, vol. 24, no. 1, pp. 11–22, 2005.

[108] G. C. Atkinson, S. L. Baldauf, and V. Hauryliuk, "Evolution of nonstop, no-go and nonsense-mediated mRNA decay and their termination factor-derived components," *BMC Evolutionary Biology*, vol. 8, no. 1, article 290, 2008.

[109] M. K. Doma and R. Parker, "Endonucleolytic cleavage of eukaryotic mRNAs with stalls in translation elongation," *Nature*, vol. 440, no. 7083, pp. 561–564, 2006.

[110] M. Graille, M. Chaillet, and H. Van Tilbeurgh, "Structure of Yeast Dom34: a protein related to translation termination factor Erf1 and involved in No-Go decay," *The Journal of Biological Chemistry*, vol. 283, no. 11, pp. 7145–7154, 2008.

[111] S. D. Moore and R. T. Sauer, "The tmRNA system for translational surveillance and ribosome rescue," *Annual Review of Biochemistry*, vol. 76, pp. 101–124, 2007.

[112] N. Benaroudj, P. Zwickl, E. Seemüller, W. Baumeister, and A. L. Goldberg, "ATP hydrolysis by the proteasome regulatory complex PAN serves multiple functions in protein degradation," *Molecular Cell*, vol. 11, no. 1, pp. 69–78, 2003.

[113] S. Hartung and K.-P. Hopfner, "Lessons from structural and biochemical studies on the archaeal exosome," *Biochemical Society Transactions*, vol. 37, no. 1, pp. 83–87, 2009.

[114] E. Lorentzen, J. Basquin, and E. Conti, "Structural organization of the RNA-degrading exosome," *Current Opinion in Structural Biology*, vol. 18, no. 6, pp. 709–713, 2008.

[115] S. Slomovic, V. Portnoy, S. Yehudai-Resheff, E. Bronshtein, and G. Schuster, "Polynucleotide phosphorylase and the archaeal exosome as poly(A)-polymerases," *Biochimica et Biophysica Acta*, vol. 1779, no. 4, pp. 247–255, 2008.

[116] H. Celesnik, A. Deana, and J. G. Belasco, "Initiation of RNA decay in *Escherichia coli* by 5′ pyrophosphate removal," *Molecular Cell*, vol. 27, no. 1, pp. 79–90, 2007.

[117] V. Portnoy, E. Evguenieva-Hackenberg, F. Klein et al., "RNA polyadenylation in Archaea: not observed in Haloferax while the exosome polynucleotidylates RNA in Sulfolobus," *EMBO Reports*, vol. 6, no. 12, pp. 1188–1193, 2005.

[118] V. Portnoy and G. Schuster, "RNA polyadenylation and degradation in different Archaea; roles of the exosome and RNase R," *Nucleic Acids Research*, vol. 34, no. 20, pp. 5923–5931, 2006.

[119] S. Slomovic, D. Laufer, D. Geiger, and G. Schuster, "Polyadenylation of ribosomal RNA in human cells," *Nucleic Acids Research*, vol. 34, no. 10, pp. 2966–2975, 2006.

[120] S. F. Newbury, "Control of mRNA stability in eukaryotes," *Biochemical Society Transactions*, vol. 34, no. 1, pp. 30–34, 2006.

[121] D. Hasenöhrl, T. Lombo, V. Kaberdin, P. Londei, and U. Bläsi, "Translation initiation factor a/eIF2(-γ) counteracts 5′ to 3′ mRNA decay in the archaeon *Sulfolobus solfataricus*," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 105, no. 6, pp. 2146–2150, 2008.

[122] S. L. French, T. J. Santangelo, A. L. Beyer, and J. N. Reeve, "Transcription and translation are coupled in Archaea," *Molecular Biology and Evolution*, vol. 24, no. 4, pp. 893–895, 2007.

[123] M. A. Humbard, G. Zhou, and J. A. Maupin-Furlow, "The N-terminal penultimate residue of 20S proteasome α1 influences its Nα acetylation and protein levels as well as growth rate and stress responses of *Haloferax volcanii*," *Journal of Bacteriology*, vol. 191, no. 12, pp. 3794–3803, 2009.

[124] J. A. Maupin-Furlow, M. A. Humbard, P. A. Kirkland et al., "Proteasomes from structure to function: perspectives from Archaea," *Current Topics in Developmental Biology*, vol. 75, pp. 125–169, 2006.

[125] M. Hochstrasser, "Origin and function of ubiquitin-like proteins," *Nature*, vol. 458, no. 7237, pp. 422–429, 2009.

[126] H. Kubota, "Quality control against misfolded proteins in the cytosol: a network for cell survival," *Journal of Biochemistry*, vol. 146, no. 5, pp. 609–616, 2009.

[127] C. Hirsch, R. Gauss, S. C. Horn, O. Neuber, and T. Sommer, "The ubiquitylation machinery of the endoplasmic reticulum," *Nature*, vol. 458, no. 7237, pp. 453–460, 2009.

[128] M. A. Humbard, H. V. Miranda, J.-M. Lim et al., "Ubiquitin-like small archaeal modifier proteins (SAMPs) in *Haloferax volcanii*," *Nature*, vol. 463, no. 7277, pp. 54–60, 2010.

Hindawi Publishing Corporation
Archaea
Volume 2010, Article ID 453642, 14 pages
doi:10.1155/2010/453642

Review Article

Selenocysteine, Pyrrolysine, and the Unique Energy Metabolism of Methanogenic Archaea

Michael Rother[1] and Joseph A. Krzycki[2]

[1] *Institut für Molekulare Biowissenschaften, Molekulare Mikrobiologie & Bioenergetik, Johann Wolfgang Goethe-Universität, Max-von-Laue-Str. 9, 60438 Frankfurt am Main, Germany*
[2] *Department of Microbiology, The Ohio State University, 376 Biological Sciences Building 484 West 12th Avenue Columbus, OH 43210-1292, USA*

Correspondence should be addressed to Michael Rother, m.rother@bio.uni-frankfurt.de and Joseph A. Krzycki, krzycki1@osu.edu

Received 15 June 2010; Accepted 13 July 2010

Academic Editor: Jerry Eichler

Methanogenic archaea are a group of strictly anaerobic microorganisms characterized by their strict dependence on the process of methanogenesis for energy conservation. Among the archaea, they are also the only known group synthesizing proteins containing selenocysteine or pyrrolysine. All but one of the known archaeal pyrrolysine-containing and all but two of the confirmed archaeal selenocysteine-containing protein are involved in methanogenesis. Synthesis of these proteins proceeds through suppression of translational stop codons but otherwise the two systems are fundamentally different. This paper highlights these differences and summarizes the recent developments in selenocysteine- and pyrrolysine-related research on archaea and aims to put this knowledge into the context of their unique energy metabolism.

1. Introduction

Expansion of the amino acid repertoire of proteins beyond the 20 "canonical" amino acids is a phenomenon observed almost 50 years ago [1]. Numerous modifications of the carboxyl- or amino-terminals or the individual side chains of amino acids after ribosomal synthesis of the respective polypeptide had finished were identified and the biosynthetic path elucidated (reviewed in [2]). It is thus not surprising that a similar process was assumed when selenocysteine, 2-selenoalanine, was discovered as constituent of eukaryal and bacterial proteins [3]. What made selenocysteine special is that subsequent efforts established the cotranslational nature of its insertion into proteins at the position of a UGA stop codon on the respective mRNA [4, 5]. Thus, selenocysteine was designated the 21st proteinogenic amino acid [6]. Discovery of pyrrolysine, lysine with $^{\varepsilon}$N in amide linkage to (4*R*,5*R*)-4-methyl-pyrroline-5-carboxylate, occurred in a different order, a single in-frame amber codon within the gene encoding the monomethylamine (MMA) methyltransferase in *Methanosarcina barkeri* [7, 8] was later found to correspond to pyrrolysine in the crystal structure [9, 10] and have its own tRNA [11]. As pyrrolysine was also shown to be inserted cotranslationally, it was designated the 22nd proteinogenic amino acid [10]. Beside the fact that translation of selenocysteine and pyrrolysine both involves suppression of stop codons the two systems have little in common (also reviewed in [12, 13]). To emphasize the differences between the mechanisms underlying selenocysteine and pyrrolysine translation, to summarize recent insights from efforts to better understand the biology of these two unusual amino acids, and to put this knowledge into the physiological context of the unique energy metabolism of methanogenesis are the aims of this paper.

2. Selenocysteine and Methanogenesis

The early observation that selenium supply influences growth performance of some methanogens [14–16] is due to the fact that most of their selenocysteine-(Sec-)containing enzymes are involved in the organism's primary metabolism,

methanogenesis. This process is of profound global importance as ca. 2% of the net CO_2 fixed into biomass is recycled through methane [17, 18]. All known methanogens are members of the domain archaea. The range of substrates methanogenic archaea use for methanogenesis is rather limited reflecting the narrow ecological niche methanogens occupy; mostly simple C1- and C2-compounds such as CO_2 (with hydrogen as reductant), carbon monoxide, methanol, methylamines (mono-, di-, and trimethylamine, as well as tetramethylammonium ion), methylsulfides, and acetate are converted to methane. The different substrate classes are metabolized via distinct, but overlapping, pathways of methanogenesis (for reviews, see [19–21]). Of the five established orders of methanogenic archaea [22], the *Methanococcales*, *Methanobacterales*, *Methanomicrobiales*, *Methanopyrales*, and *Methanosarcinales*, all (with very few exceptions) but the latter are strictly hydrogenotrophic, that is, only $H_2 + CO_2$ and/or formate serve as energy substrates.

It is intriguing that before Böck's and Stadtman's laboratories became famous for their selenium-related research they both studied (among other things) different aspects of methanogens [32–34]. It, thus, seems a striking coincidence that within the Archaea proteins containing selenocysteine [35], are until now restricted to methanogens (*Methanococcales* and the related *Methanopyrales*) [36, 37]. The known archaeal Sec-containing proteins are listed in Table 1. It should be emphasized that most methanogens do not employ selenocysteine, which poses the question as to why they get along just fine without it while others employ, and sometimes even depend on, the residue. There is no straightforward answer to this question but some considerations will be given below.

If the methanogenic growth substrate is formate, it is first oxidized to CO_2 via (sometimes Sec-containing) formate dehydrogenase (FDH, Figure 1) [24]. In the hydrogenotrophic pathway of methanogenesis, CO_2 is sequentially reduced to methane in seven steps via coenzyme-bound intermediates using H_2 as the electron donor (Figure 1). The eight-electron reduction proceeds through the redox levels of formate, formaldehyde, and methanol. Formyl-MF dehydrogenase (FMD), of which a subunit (FwuB) can contain Sec [25] catalyzes CO_2 reduction to the formyl-level and attachment to methanofuran (a 2-aminomethylfuran derivative). Of the three hydrogenases responsible for hydrogen activation in *Methanococcus,* subunits of two can contain Sec. These are the large subunit of the F_{420}-dependent hydrogenase Fru (FruA, F_{420} is a hydride carrier functionally analogous to dinucleotide cofactors) [38] and two subunits of the F_{420}-independent Vhu hydrogenase (VhuU and VhuD) [39]. Vhu and the heterodisulfide reductase (HDR), of which the large subunit (HdrA) can contain Sec [40], form a tight complex [41]. The heterodisulfide of coenzyme M and coenzyme B formed in the last step of methanogenesis serves as the terminal electron acceptor, which is reduced by HDR [18].

3. Pyrrolysine and Methanogenesis

The *Methanosarcinales* have the most diverse catabolism of all methanogens, largely due to the ability to use methylated substrates. Methylotrophic methanogenesis requires simultaneous oxidation and reduction of the methyl group (Figure 2) on the same C1 carriers employed during hydrogenotrophic methanogenesis. The reducing equivalents from the oxidation of one methyl group, at least some of which are produced as hydrogen [42], are used to reduce three methyl groups to methane. Several steps of both the reductive and oxidative branches of methylotrophic methanogenesis may rely on distinct isozymes from those essential for acetotrophic or hydrogenotrophic methanogenesis [43]. Both oxidative and reductive branches originate with methyl-Coenzyme M (methyl-CoM); thus, key to the diverse substrate range of *Methanosarcinales* are a number of methyltransferases with specificity for a particular methylated substrate such as methanol [44, 45], MMA [46], dimethylamine (DMA) [47], or trimethylamine (TMA) [48]. Each methyltransferase methylates a dedicated cognate corrinoid protein, subsequently demethylated by one of several corrinoid protein:CoM methyltransferases, such as MtbA, or MtaA [49]. An iron-sulfur protein, RamA, reductively activates the methylamine corrinoid proteins prior to initial methyl transfer [50]. The corrinoid proteins are homologous to the B_{12}-binding domain of methionine synthase [7, 8, 51, 52] and share the double Rossman fold which binds the cofactor, while the corrinoid:CoM methyltransferases are closely related to uroporphorinogen decarboxylases [51, 53]. Both corrinoid proteins and CoM methylases are often erroneously classified as their more commonly known homologs in genome annotations. In contrast, the methanol, MMA, DMA, and TMA methyltransferases share no sequence similarity with each other [7, 8]. However, both the methanol [54] and MMA methyltransferases [9] have TIM barrel structures in common with methyltransferases interacting with homologous corrinoid proteins [55].

Multiple isoforms of each type of methyltransferase and cognate corrinoid protein are often encoded in different *Methanosarcinales* genomes [56–59]. Mutagenesis of all three methanol methyltransferase genes is required to eliminate methanol dependent growth in *Methanosarcina acetivorans* [60]. Although two MMA methyltransferase genes (*mtmB1* and *mtmB2*) are found in *M. barkeri* MS, only MtmB1 is isolated as an abundant protein during growth on MMA [46, 61]. The *mtmB2* gene of *Methanosarcina mazei* is most highly expressed during growth on methanol, possibly representing a nitrogen scavenging strategy [62]. MttB1, the predominant TMA methyltransferase isoform [48, 61], may preferentially interact with a TMA-specific permease in *Methanococcoides burtonii* [63].

The abundance of the isolated methylamine methyltransferases added to the initial surprise of finding a single in-frame amber codon in *mtmB1* in two different strains of *M. barkeri* [7]. An amber codon was also found in the *M. barkeri mtbB1*, *mtbB2*, and *mtbB3* genes encoding isozymes of DMA methyltransferase, and the *M. barkeri* and *Methanosarcina thermophila* TMA methyltransferase gene *mttB* [8]. The genomes of *M. barkeri*, *M. acetivorans*, *M. mazei*, and *M. burtonii* have since shown the amber codon is a conserved trait in all methylamine methyltransferase genes [56–59]. In each class of methyltransferase the amber codon placement

Table 1: Sec-containing proteins of archaea.

Selenoprotein-containing enzyme	Subunit	Function	Characteristic organism	Reference
Formate dehydrogenase	FdhA	Methanogenesis	*Methanococcus vannielii*	[23][24]
Formyl-methanofuran dehydrogenase	FwuB	Methanogenesis	*Methanopyrus kandleri*	[25]
F_{420}-reducing hydrogenase	FruA	Methanogenesis	*Methanococcus voltae*	[26]
F_{420}-nonreducing hydrogenase	VhuD	Methanogenesis	*Methanococcus voltae*	[26][27]
	VhuU	Methanogenesis		
Heterodisulfide reductase	HdrA	Methanogenesis	*Methanocaldococcus*[a] *jannaschii*	[28]
Selenophosphate synthetase	homomeric	Sec synthesis	*Methanococcus maripaludis*	[28, 29]
HesB-like protein	unknown	Unknown (iron/sulfur cluster assembly?)	*Methanococcus maripaludis*	[29, 30]

[a]*Methanococcus jannaschii* was placed in a separate genus, *Methanocaldococcus* [31].

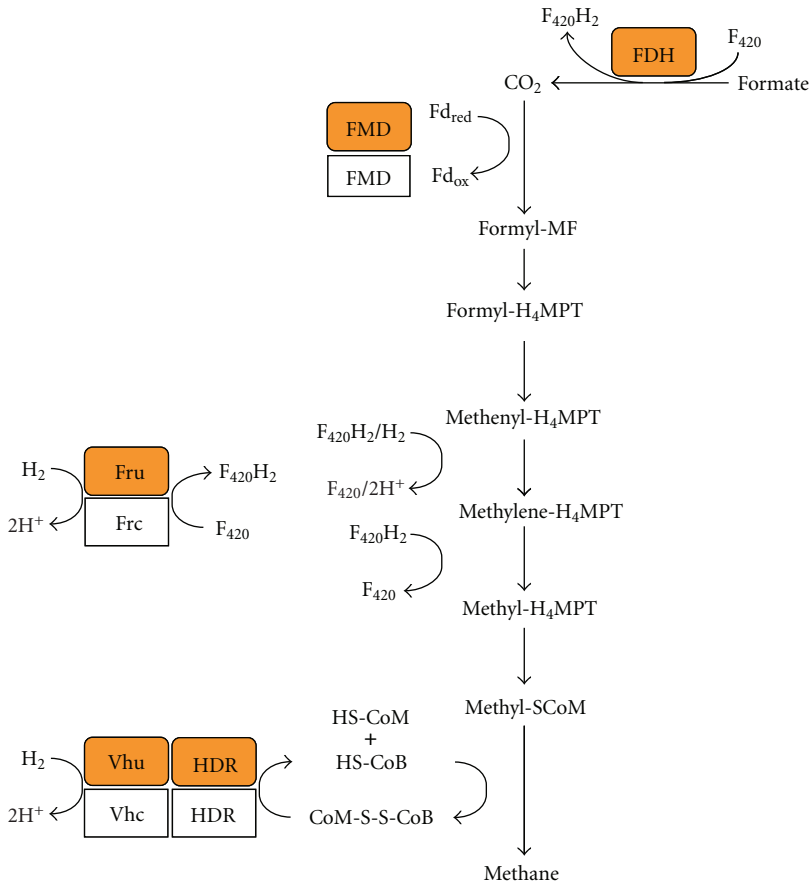


Figure 1: Scheme of hydrogenotrophic methanogenesis involving Sec-containing proteins (orange); the Cys-containing isoforms are in white. CoM-S-S-CoB, heterodisulfide of coenzyme M and coenzyme B; Fd, ferredoxin; Fd_{ox}, oxidized Fd; Fd_{red}, reduced Fd; FDH, formate dehydrogenase; FMD, formyl-methanofuran dehydrogenase; Fru, F_{420}-reducing hydrogenase; F_{420}, (oxidized) 8-hydroxy-5-deazariboflavin derivative; $F_{420}H_2$, reduced coenzyme F_{420}; H_4MPT, tetrahydromethanopterin, HDR, heterodisulfide reductase; HS-CoB, coenzyme B (*N*-7-mercaptoheptanoyl-*O*-phospho-*L*-threonine); HS-CoM, coenzyme M (2-mercaptoethanesulfonic acid); MF, methanofuran; Vhu, F_{420}-nonreducing hydrogenase.

is different, but conserved in genes encoding isozymes of a particular methyltransferase. A single exception is the *mttB3* gene of *M. burtonii* that lacks an in-frame amber codon [63]. This gene is not expressed during growth on TMA [63], but may instead be specific for a known or unknown methylotrophic substrate. For example, the encoding gene for tetramethylammonium chloride methyltransferase is not yet identified [64]. The UAG codon remains in methylamine methyltransferase transcripts [8], yet little detectable UAG termination product of *mtmB* is detectable in cell extracts [65, 66]. Peptide sequencing of HPLC isolated peptides revealed the reading frame conserved before and after the UAG-encoded position [8, 65], with a lysine observed at the UAG-encoded position. Lysine codon usage is normal in other genes in *M. barkeri*, and the possibility that a labile lysine residue could be present at the UAG position that was

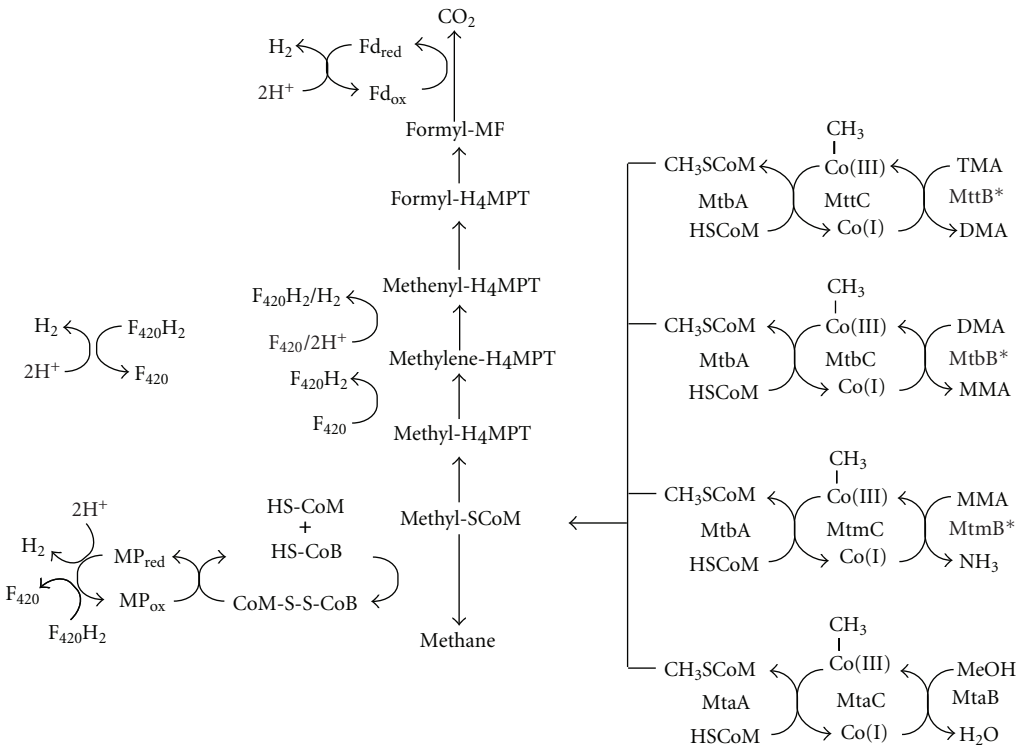


Figure 2: Scheme of methylotrophic methanogenesis. Methyl groups from methanol, TMA, DMA, and MMA are mobilized into metabolism by the action of a substrate-specific methyltransferase interacting with its cognate corrinoid protein in which the cofactor's cobalt ion cycles between methyl-Co(III) and Co(I) states. Methyltransferases that are pyrrolysyl-proteins are marked with an asterisk. The corrinoid cofactor is then demethylated by the action of a methylcobamide:CoM methyltransferase such as MtbA (for methylamines) or MtaA (for methanol). Adventitious oxidation can inactive the corrinoid proteins to the Co(II) state, which can be reductively reactivated by RamA (for methylamines) and possibly by RamA homologs for other pathways. Reducing equivalents in the form of hydrogen, $F_{420}H_2$, or Fd_{red} from the oxidation of methyl-CoM are used to reduce methanophenazine (MP) and subsequently CoM-S-S-CoB, thereby generating ATP via electron transport phosphorylation and the free HS-CoM and HS-CoB cofactors; CoM-S-S-CoB is then recycled by the reduction of methyl-CoM to methane. See Figure 1 for cofactor abbreviations.

destroyed during peptide isolation [65] was addressed by the structure of MtmB and the subsequent visualization of pyrrolysine [9, 10]. Mass spectroscopy of MtmB, MtbB, and MttB confirmed the mass of pyrrolysine corresponding to the proposed structure at the UAG encoded position of all three proteins [61].

4. Peculiarities of Archaeal Selenocysteine Synthesis and Incorporation

The pathway of Sec biosynthesis and incorporation is well understood in *E. coli* [67]. First, Sec-specific tRNA ($tRNA^{sec}$) is charged with serine by seryl-tRNA synthetase, and the seryl moiety is subsequently converted to a selenocysteyl-moiety by Sec synthase (SS). The selenium donor is selenomonophosphate generated by selenophosphate synthetase (SPS). The specialized translation elongation factor SelB (homologous to EF-Tu) delivers in its GTP-bound form the selenocysteylated tRNA to the ribosome via binding of a secondary structure on the selenoprotein mRNA, the SECIS element, located immediately adjacent to the UGA codon [68–71]. This binding triggers a conformational change in the quaternary Sec-$tRNA^{sec}$-SelB-GTP-SECIS complex, which allows for insertion of the charged tRNA into the ribosomal A site [72, 73].

Both Sec synthesis and Sec insertion differ in Archaea from the bacterial path (Figure 3). In fact, identical strategies appear to be employed by archaea and eukarya—to the exclusion of bacteria. Conversion of the seryl-moiety to Sec proceeds via two steps: seryl-$tRNA^{sec}$ is phosphorylated to *O*-phosphoseryl-$tRNA^{sec}$ in an ATP-dependent reaction by phosphoseryl-$tRNA^{sec}$ kinase (PSTK) [74, 75]; subsequently, the *O*-phosphoseryl-moiety is converted to the Sec-moiety by *O*-phosphoseryl-$tRNA^{sec}$:Sec synthase (SepSecS) [76]. SepSecS (also named SecS in the eukaryal system [77]) is (like SS) a pyridoxal phosphate-dependent enzyme, and its proposed reaction mechanism is analogous to that proposed for bacterial SS [76, 78]. Although it has been shown for *Trypanosoma brucei* that only the PSTK- and SepSecS-dependent pathway is present [79], and although high-resolution structures of both enzymes have recently become available [80–84], the physiological function of *O*-phosphoseryl-$tRNA^{sec}$ and, thus, the selective advantage for investing an additional ATP in Sec synthesis (as compared to the bacterial system) is not evident. The possibility that this

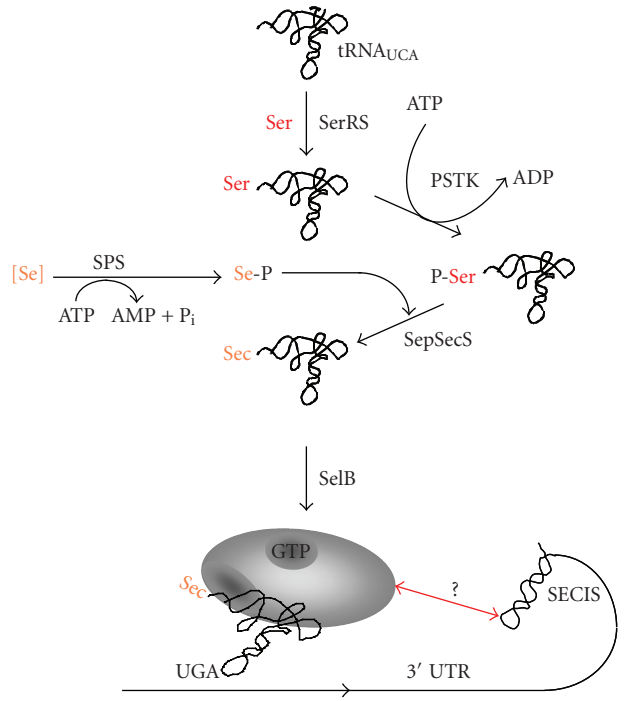


Figure 3: Schematic representation of selenocysteine biosynthesis and incorporation in Archaea. 3′UTR, 3′-untranslated region; PSTK, seryl-tRNA^{sec} kinase, [Se], reduced Se-species; SelB, Sec-specific elongation factor; SepSecS, *O*-phosphoseryl-tRNA^{sec}: selenocysteine synthase; Ser, serine; Se-P, seleno(mono)phosphate; SerRS, seryl-tRNA synthetase; SPS, selenophosphate synthetase; see text for details.

pathway renders activation of selenium in the SPS reaction unnecessary [85, 86], turned out not to apply [29]. Severely selenium deprived rats were shown to incorporate cysteine at the selenocysteine-position of selenium-dependent thioredoxin reductase, probably in order to salvage at least some enzymatic activity [87]. Furthermore, it was suggested that *O*-phosphoseryl-tRNA^{sec} might be converted to cysteinyl-tRNA^{sec} in *Methanococcus maripaludis* [88]. However, if no highly stringent mechanisms to ensure the fidelity of codon/amino acid correlation during translation under "normal" physiological conditions were operative, such a possibility would render any UGA codon within a selenoprotein mRNA ambiguous for the amino acid to be inserted, which in turn would have detrimental effects on the "mistranslated" protein's activity. Fortunately, *M. maripaludis*, for which this scenario was proposed, is the ideal model to study it and exciting new insights regarding the physiological meaningfulness of the additional phosphorylation step await us. The same is true for the *in vivo* role of selenium-binding protein (SeBP), a 81 amino acid polypeptide which binds one reduced selenium per tetrameric protein *in vitro* [89, 90] and could therefore be involved in transport and intracellular trafficking of selenium.

In eukarya and archaea, the SECIS element is located in the nontranslated regions of selenoprotein mRNAs; noteworthy, there is no relation between the SECIS elements in terms of structure and/or sequence, which indicates that they have distinct evolutionary origins [67] and that the modes of SECIS-function, that is, SECIS recognition, might differ. While the bacterial SECIS is specifically bound by the Sec-specific translation elongation factor SelB, the archaeal and eukaryal counterparts are not; there, the respective SelBs do not contain a C-terminal extension shown to be responsible for SECIS-binding in *E. coli* [91–93]. Instead eukaryal SECIS elements are bound by "SECIS-binding protein 2" (SBP2) and the ribosomal protein L30 [94, 95]. Other factors have also been shown to be involved in SECIS-dependent UGA recoding, but their exact function is not clear yet (current knowledge is summarized in [96, 97]). So far, no SECIS-binding factor has been identified in Archaea. Homologs of SBP2 are not encoded in any available archaeal genome and the two L30 homologs of *Methanococcus maripaludis* appear not to be involved in selenocysteine insertion because their homologous overproduction had no effect on selenoprotein formation of the organism (Sattler and Rother, unpublished data). Further, *Methanocaldococcus jannaschii* encodes three L30 homologes [40] and the one most similar to SBP2 was tested whether it could function in eukaryal selenoprotein synthesis but did not [94]. This may not be too surprising as both the structure/sequence and the distance of the archaeal SECIS elements to the respective UGA codons they recode is different to the eukaryal SECS elements [67], which may be as far as 5.4 kb downstream of their cognate UGA [98] with an average distance of 500–2,500 nt [30, 99]; the archaeal SECIS structures are usually located in closer proximity (80–1,300 nt [40, 100, 101]). It is possible that a distance constraint is the reason for a unique situation regarding the selenoprotein FdhA (encoding a subunit of formate dehydrogenase; see Table 1); the putative SECIS element is located in the 5′-nontranslated region of the respective deduced mRNA, maybe because the distance of the Sec codon and the 3′-nontranslated (>1,600 nt) region would be too great. However, this scenario needs to be verified by experimentation and fortunately, the tools required for such an *in vivo* analysis are now available [102, 103].

5. Pyrrolysine Synthesis and Incorporation Is Different from Sec

An amber decoding tRNA^{Pyl} (then called tRNA$_{CUA}$) was identified as the product of the *pylT* gene [11] simultaneously with the discovery of pyrrolysine. This gene is the first of the *pylTSBCD* gene cluster found in representative *Methanosarcinales* [11]. The five *pyl* genes have proven sufficient for the biosynthesis and genetic encoding of pyrrolysine [66].

Initial scenarios suggested that tRNA^{Pyl} might be charged with lysine, either by a complex of class I and class II lysyl-tRNA synthetases [104], or by the *pylS* gene product, identified as a homolog of class II aminoacyl-tRNA synthetases [11]. Both ideas proved incorrect. Loss of the genes encoding one or the other LysRS from *Methanosarcina acetivorans* does not affect UAG translation as pyrrolysine [105], and the substrate of PylS is not lysine, but pyrrolysine itself [106].

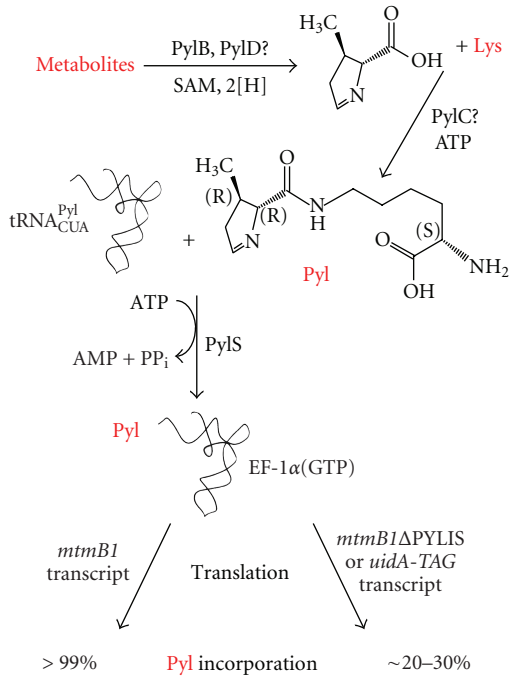


Figure 4: Schematic representation of pyrrolysine incorporation into protein. The *pylB*, *pylC*, and *pylD* genes have been shown to enable pyrrolysine biosynthesis in *E. coli*, but their exact roles are unknown. A conceptual and speculative scheme is shown which is keeping with the activities of proteins in their respective proteins families. Lysine is likely to form the acyl of pyrrolysine, but may also be coupled to an early precursor which subsequently cyclizes after amide bond formation. Pyrrolysine is given entrance to the genetic code by PylS in Archaea, the equivalent in Bacteria are the products of the split gene *pylSc* and *pylSn*. The direct formation of pyl-tRNA^{Pyl} is likely followed by binding to the elongation factor used by the canonical amino acids, that is, EF-1α in Archaea. The common bacterial elongation factor EF-Tu can bind charged tRNA^{Pyl} both in vivo and in vitro. The recognition of pyl-tRNA^{Pyl} by non-specialized elongation factors underlies the relatively high level of UAG translation in a reporter gene such as *uidA* with an introduced amber codon in organisms having tRNA^{Pyl}, PylS, and pyrrolysine or an analog to charge tRNA^{Pyl}. The PYLIS is not essential for this level of translation, as shown by replacement of PYLIS in *mtmB1*, but may enhance UAG translation to some extent. This effect is unlikely to require the structure formed by PYLIS. See text for further details.

Quite unexpected by analogy to selenocysteine, pyrrolysine is not made on tRNA^{Pyl}, but as a free amino acid which is directly ligated to the tRNA (Figure 4).

PylS is a pyrrolysyl-tRNA synthetase, as shown by *in vitro* activity with chemically synthesized pyrrolysine [10, 107, 108]. The 50 μM K_m of PylS for stereochemically pure pyrrolysine remains the lowest observed for any substrate or pyrrolysine analog tested [107]. The function of the *pylS* and *pylT* gene products was confirmed by *in vivo* synthesis of pyrrolysine-containing proteins in *E. coli* transformed with *pylT* and *pylS* and supplemented with exogenous pyrrolysine [107]. As an orthologous pair, PylS and tRNA^{Pyl} have been exploited in recent years to incorporate Pyl analogs with modifiable tags for production of recombinant proteins with derivatizable residues [109–111]. Several structures of the catalytic core of *M. mazei* PylRS enzymes are now available, although these lack the first 180 residues of the protein [112–115]. The activated pyrrolysyl-adenylate made prior to ligation to tRNA^{Pyl} lies in a deep groove with the pyrroline ring buried in a hydrophobic pocket [113, 114]. A mobile loop can bring a tyrosine into H-bonding distance of the imine nitrogen of pyrrolysine [113], but is not essential for activity [114]. Analogs of pyrrolysine lacking an electronegative group at this position are recognized with a lowered specificity constant [116]. Mutation of the hydrophobic pocket has led to enhanced activity with a derivatizable pyrrolysine analog [117], or $^{\varepsilon}$N-acetyl-lysine [118].

In contrast to the single archaeal *pylS* gene, bacteria such as *Desulfitobacterium hafniense* encode *pylS* in two separate genes, with the catalytic domain encoded by *pylSc* and the N-terminal region encoded by *pylSn* [11, 13, 106]. PylSc is competent *in vitro* as a pyrrolysyl-tRNA synthetase, but has minimal detectable activity *in vivo* [115, 119]. This may be due to high affinity binding of tRNA^{Pyl} by PylSn which presumably interacts with PylSc (Jiang and Krzycki, manuscript in preparation). The structure of PylSc and tRNA^{Pyl} complex [115] revealed the compact core of tRNA^{Pyl} enhancing interaction with the catalytic domain. Interestingly, unique elements of tRNA^{Pyl} such as the elongated anticodon stem, small variable loop, and T-loop without the classical TψC sequence [11, 120], were not directly contacted by PylSc.

Recombinant expression of *pylTSBCD* in *E. coli* results in translation of UAG as internally biosynthesized pyrrolysine in reporter proteins [66]. Transformation of only *pylBCD* leads to pyrrolysine production, as determined by PylS-based charging and amino acid activation assays [66]. The recombinantly produced amino acid comigrates with synthetic pyrrolysine in TLC (Gaston and Krzycki, unpublished data). The enzymatic activities of the pyrrolysine biosynthetic genes, that is, PylB, PylC, and PylD are yet unknown but they share homologies that lead to logical possibilities [11, 66, 121]. The *pylB* gene product is highly similar to biotin synthase, while lacking key residues for binding dethiobiotin but it has all other hallmarks of the radical SAM family whose members catalyze various intramolecular rearrangements, reductions, and methylation reactions [122]. PylB may catalyze formation of the methylated ring or ring precursor. The *pylC* gene product is related to the carbamoyl-phosphate synthetase and D-alanine-D-alanine ligase superfamily and might play a role in amide bond of pyrrolysine between lysine and the ring precursor. PylD has an NAD-binding signature, and could be involved in formation of the imine bond.

The ability of a single gene cluster, *pylTSBCD,* to transform a naïve organism to incorporate genetically encoded biosynthesized pyrrolysine could underlie the far-flung distribution of pyrrolysine genes in microbes [66]. Although only *Methanosarcinales* are known to have *pyl* genes in the Archaea, all five *pyl* genes have been noted in Gram-positive bacteria such as *Desulfitobacterium hafniense* [11], *Desulfotomaculum acetoxidans*, and *Acetohalobium arabaticum* [13].

A recently sequenced genome also reveals a complete *pyl* operon in *Thermincola* sp. JR as well. Examples of *pyl* genes are also found in Gram-negative bacteria such as a δ-proteobacterial worm intestinal symbiont [123]. In the Gram-positive bacteria, *pylScBCDSn* typically form a single gene cluster (with the exception of *A. arabaticum* where *pylSn* precedes *pylSc*), whereas in the δ-proteobacterium separate *pylBCD* and *pylTScSn* gene clusters are found on different contigs of the unclosed genome sequence.

UAG appears to serve globally as both sense and stop codon in archaea having the *pyl* genes. An *E. coli uidA* gene with an introduced amber codon transformed into *M. acetivorans* resulted in mostly amber-terminated gene product, but also produced full-length active β-glucuronidase, complete with pyrrolysine, at an apparent efficiency of 20%–30% [124]. When this data is considered in light of the translation of *mtmB1* or reporter genes with introduced amber codons in *E. coli* dependent on *pylT* and *pylS* [66, 107, 124], it appears very likely that amber suppression underlies this relatively high level of pyrrolysine incorporation.

Sec incorporation into protein requires the presence of the SECIS element in the transcript as discussed above. An analogous pyrrolysine insertion sequence (PYLIS) element was proposed [125] whose basic structure exists in solution [126]. In the initial sequencing of the first *mtmB* and *mttB* genes, this same element was observed, but nothing similar could be seen in *mtbB* [127], a result that was further emphasized by a later more exhaustive bioinformatics study [12]. Direct replacement of the PYLIS confirmed it was not essential for incorporation of pyrrolysine into MtmB, albeit with a decrease in full-length product in the absence of the element [124]. Concomitant increase in amber-truncated *mtmB1* product indicated that some portion of the PYLIS might enhance UAG translation or diminish UAG-directed termination. The effect observed is unlikely to require the entire structure formed by the PYLIS as comparison of the PYLIS from ten methanogen *mtmB* genes shows only limited evidence of covariance [128].

Again, unlike Sec, factors that might participate in highly efficient translation of UAG as pyrrolysine during *mtmB* expression have not been identified. Two release factors are encoded in the genomes of *M. mazei* and *M. acetivorans*, which were proposed to possibly participate in UAG translation as pyrrolysine by differential recognition of stop codons [12], but only one of these homologs is found in the genomes of *M. barkeri*, *M. burtonii*, or *Methanohalophilus mahii* suggesting this could not be a general method of enhancing UAG translation. The two release factors from *M. acetivorans* were tested and one was capable of recognizing all three stop codons, whereas the other had no activity, leaving its function an open question [129].

Unlike Sec-tRNA^{Sec}, Pyl-tRNA^{pyl} can be recognized by bacterial EF-Tu *in vitro* [120] and *in vivo* [107], and by eukaryotic EF-1α as well [110, 111]. This would seem to obviate the need for another elongation factor, although it would be possible that an elongation factor with a higher affinity for Pyl-tRNA^{Pyl} could function in pyrrolysine-utilizing archaea or bacteria. The *pyl* genes-containing *Methanosarcinales* encode an archaeal EF1α, as well as a SelB homolog that could enhance recognition of Pyl-tRNA^{Pyl} [130]. This idea has not yet been tested, but it should be noted that close homologs of this same SelB-like protein could be found in other methanogens that lack *pyl* genes, indicating the role of this factor may not be connected to pyrrolysine metabolism.

6. Why Use Sec?

Most of the organisms for which genome information is available obviously do not employ Sec at all. Although not abided by all bioinformaticians, more than a TGA-interrupted gene and an orphan translation factor are needed to conclude that an organism synthesizes selenoproteins [131]. The simplest way would be to conduct an experiment but at least tRNA^{sec} and the Sec-biosynthesis factors (SS or PSTK/SepSecS) have to be encoded as well.

Why organisms use Sec is still not understood, mostly because for most characterized selenoproteins the specific functions of Sec are still unknown. Furthermore, for all but one of the selenoproteins of prokaryotes (clostridial glycine reductase), homologous proteins with cysteine (Cys) at the respective position exist [132]. This is true even within one organism in Sec-utilizing methanogens. In *M. maripaludis* strain JJ, for example, all of its selenoproteins are dispensable during growth with H_2 + CO_2 because they can be substituted by a set of Cys-isoforms. On the other hand, a very close relative, *M. maripaludis* strain S2, cannot do without its selenoproteins, probably because for at least one of them no complementing Cys-isoform exists or is sufficiently active [29]. It was shown for strain JJ that the Cys-isoforms are present at a much higher level than the selenoproteins they replace, which was interpreted as a means to counteract decreased kinetic efficiency of the Cys-containing proteins in order to retain competitive substrate flux through the methanogenic pathway [133]. Thus, using selenoproteins could be a strategy to avoid unnecessary protein synthesis. However, the notion that selenoenzymes are "super-cysteine-enzymes", an argument often used and derived from the observed drastically lower enzymatic efficiencies of Sec to Cys mutant variants of selenoproteins [134, 135], is proven to not always be correct [136]. It can, thus, not be the sole explanation for the use of Sec. The same is true for the argument that Sec is more reactive than Cys due to the fact that the selenol group is mostly deprotonated at physiological pH while the thiol group is mostly protonated due to the different pK_a values (5.2 for Sec, 8.3 for Cys) [137, 138]. However, measuring the respective values in different Cys- and Sec-containing peptide- and protein-contexts showed that pK_a cannot serve as the sole explanation for the use of Sec [139, 140]. Selenoproteins have a lower redox potential than their Cys-homologs in the cases where this was determined [141, 142], a feature that is also often used to explain the use of Sec. It was recently pointed out that Sec has a higher nucleophilic character than Cys and as a consequence might better facilitate initial high rates in redox catalysis [143]. On the other hand, it has

been argued that higher electophilicity of Sec than Cys and ultimately protection of a Sec-containing enzyme from overoxidation could be the "biological rationale" to employ Sec [144]. However, all these arguments may be true for some, but most probably not for all Sec-containing proteins. Furthermore, a much more facile accessibility of the radical oxidation state of Sec as compared to Cys has been observed and although no known selenoproteins has been shown to use a radical mechanism, the electrochemical difference is remarkable. Therefore, it may have biological implications, such as in mediation of one-electron- and two-electron-transfer processes [145]. All of the mentioned potential advantages of Sec over Cys might apply for methanogenic archaea employing Sec, as most of the selenoproteins act in the central metabolic pathway, methanogenesis, and the organisms thrive at lowered redox potential conditions [146]. As such, methanogenic archaea would be ideal to study the differences between selenoproteins and their cysteine-containing isoforms in a naturally occurring system.

Today, there is broad consensus that Sec constitutes a very ancient trait, present already in the last universal common ancestor [76, 147, 148]. A simple explanation of why some methanogens use Sec and others do not is unequal loss of the trait due to different selective pressures during evolution. Closely related *Methanococcus* species, or even strains of the same species probably represent "moments" in this process [29, 149]. The Sec-utilization trait was and is being lost in archaea—and those still synthesizing selenoproteins might just thrive under conditions, like permanent absence of oxygen and low reductant concentration, which are not selecting against this trait. On the other hand, very low selenium availability should rapidly select against this trait and with typical environmental concentrations of selenate ranging from 20 nM to less than 100 pM [150, 151] such a scenario is plausible. Furthermore, microbial and chemical reduction of selenate [152, 153] and selenite [154] to insoluble elemental selenium and gaseous hydrogen selenide can deplete bioavailable selenium even further.

7. Why Use Pyl?

The major physiological reason apparent for pyrrolysine remains methylamine methyltransfer. The *pyl* operon lies adjacent to a separately transcribed MMA methyltransferase operon in all *Methanosarcina* species examined to date [11]. The only essential activity lost by deletion of *pylT* and the *pyl* promoter from *M. acetivorans* is the ability to use methylamines, resulting in cells unable to grow on MMA, DMA, or TMA, yet with no apparent defect in growth or methanogenesis from methanol or acetate, save that methylamines no longer are a nitrogen source [155]. This result suggests that the only viable routes to metabolism of methylamines in this organism are the corrinoid-dependent methyltransferases made via UAG translation as pyrrolysine.

To date, every bacterial species found to have a *pyl* operon has also had homologs of methylamine methyltransferase genes with conserved amber codons. Nearby genes often encode corrinoid proteins, and occasionally bacterial RamA homologs, and corrinoid:pterin methyltransferases. These genes suggest some of these bacteria have pathways to mobilize methylamine into metabolism as methylated pterins to serve as electron donors for anaerobic respirations and as a source of cellular carbon and nitrogen. Few examples exist of bacteria that use CoM (e.g., see [156]), however, in *D. acetoxidans* a cluster of genes encoding homologs of MtmB, its cognate corrinoid protein MttC, CoM methylase MtbA, and bacterial RamA are found. It is tempting to suggest this organism could employ CoM as a methyl-donor. At times, the link between methylamine metabolism and the *pyl* operon is even more intimate than seen with the methylamine utilizing methanogens. For example, the *pyl* gene cluster of *A. arabaticum* is interspersed with a gene encoding a trimethylamine methyltransferase homolog [13], and in *D. acetoxidans* an iron-sulfur protein encoded between *pylT* and *pylS* is a member of the bacterial family of proteins that are close homologs of RamA, demonstrated to activate the methylamine:corrinoid methyltransferase reaction in methanogenic archaea [50].

Homologs of the methanogen methylamine methyltransferases whose genes lack an amber codon are found in the genomes of many bacteria and a few nonmethanogenic archaea [11, 12, 123]. Such genomes always lack a complete set of *pyl* genes, unless they also possess methylamine methyltransferase genes with amber codons. Genes presumably encoding pyrrolysine-free homologs of the TMA methyltransferase are relatively prevalent, and BLAST [157] searches will readily retrieve such homologs predominantly from various α-proteobacteria, as well as in *Bacteriodes spp*, and the crenarchaeote *Thermofilum pendulans*. Rarer examples of DMA methyltransferase and MMA methyltransferase gene homologs lacking the amber codon can also be found. These proteins are diverged from the methylamine methyltransferases with pyrrolysine, and it remains to be seen if these genes are actual methylamine methyltransferases. The methylamine methyltransferases without amber codons have at the site corresponding to pyrrolysine a small or bulky hydrophobic residue, suggesting that these proteins will not have similar chemistry.

This begs the question as to what function pyrrolysine could serve in the demonstrated methylamine methyltransferases. Pyrrolysine brings a unique electrophilic nature to the repertoire of genetically encoded amino acids, one that can otherwise only be introduced into proteins by cofactors or residue modification [158]. Pyrrolysine reactivity with nucleophiles [9, 10] suggests the ability to participate in corrinoid dependent methylamine methyltransferase reactions by interacting with either the methylamine substrate or product. In the case of the MMA methyltransferase, pyrrolysine lies in the bottom of an anionic cleft that corresponds to the active site in corrinoid-dependent methyltransferases that are structurally analogous to MtmB [9, 10]. The pyrroline ring rotates by 90° upon forming an adduct with a nucleophile such as ammonia or hydroxylamine, and it is postulated that such a methylamine-pyrrolysine adduct could facilitate methyltransfer to the Co(I) form of the cognate corrinoid protein [9]. The ring rotation would bring the methyl-group of the MMA-pyrrolysine

adduct into roughly the same position occupied prior to corrinoid transfer of the methyl-group during function of the methyl-tetrahydrofolate:corrinoid methyltransferase domain of methionine synthase [55]. Recently, site directed mutagenesis and inhibitor studies showed that pyrrolysine is crucial for rapid transfer of the methyl-group to corrinoid cofactor or protein (Longstaff, Soares, and Krzycki, manuscript in preparation). However, while these studies will demonstrate the importance of pyrrolysine in methyltransfer, many aspects of the proposed model for pyrrolysine function remain completely untested and are a priority for the field.

Pyrrolysine has been physically observed in one protein beyond methylamine methyltransferases, Thg1 from *M. acetivorans* [159, 160]. Thg1 homologs are present in a diverse number of methanogens, and in *M. acetivorans* only, the gene has acquired an in-frame amber codon that can be translated as pyrrolysine and is not essential for activity. This result is not surprising, considering *Methanosarcina acetivorans* will incorporate pyrrolysine into a recombinant bacterial reporter protein whose gene has an introduced amber codon [124]. Other examples of the ambiguity of UAG codons in *M. acetivorans* are readily found. The *pyl* genes themselves contain an example of a UAG codon serving as stop rather than sense codon. The *pylB* gene has a TAA codon ending the open reading frame in *M. barkeri*, but a TAG codon corresponds to the same position in *M. mazei*, *M. acetivorans*, and *M. burtonii*, and when the *M. acetivorans* gene is expressed in *E. coli* with TAG changed to TAA, the protein is functional in pyrrolysine biosynthesis [66]. A family of transposase genes derived from a *Bacillus* insertion element found in *M. acetivorans* and *M. mazei* [12, 56, 58] may also prove an interesting story when the functionality of these genes is investigated. Each representative has a conserved in-frame amber codon. One of the *M. acetivorans* transposase genes with an amber codon is most similar to several from *M. mazei*, indicating possible transfer of this transposase gene between species. The same family of transposases is found in *M. burtonii*, but in spite of the presence of *pyl* genes in this organism, these highly similar genes lack the amber codon [128].

8. Conclusion

Selenocysteine and pyrrolysine are powerful examples of the versatility inherent in the genetic code. They further provide examples of how precedent, though valuable, is not always the best predictor in scientific investigation, and that unpredicted paths can often be found as solutions for apparently similar phenomena. Selenocysteine was first thought to be another example of a posttranslationally modified amino acid, and was later found to be a genetically encoded amino acid made and incorporated into protein in a way unlike the canonical twenty amino acids. Pyrrolysine was subsequently thought to be most likely analogous to selenocysteine, yet it is biosynthesized and ligated to tRNA in a manner much more reminiscent of the common twenty amino acids. The story of each of these residues illustrates the immense information and opportunity found in the sequenced genomes, but also provides a reminder of the obligate pairing of prediction with experimental investigation.

Acknowledgments

M. Rother acknowledges the support from V. Müller, University of Frankfurt. Work in M. Rother's laboratory is supported by grants from the Deutsche Forschungsgemeinschaft (through SFB 579). The work of J. Krzycki's laboratory is supported by National Institutes of Health Grant GM070663 and Department of Energy Grant DE-FG0202-91ER200042.

References

[1] A. Kaji, H. Kaji, and G. D. Novelli, "A soluble amino acid incorporating system," *Biochemical and Biophysical Research Communications*, vol. 10, no. 5, pp. 406–409, 1963.

[2] F. Wold, "In vivo chemical modification of proteins (post-translational modification)," *Annual Review of Biochemistry*, vol. 50, pp. 783–814, 1981.

[3] J. E. Cone, R. M. Del Rio, J. N. Davis, and T. C. Stadtman, "Chemical characterization of the selenoprotein component of clostridial glycine reductase: identification of selenocysteine as the organoselenium moiety," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 73, no. 8, pp. 2659–2663, 1976.

[4] I. Chambers, J. Frampton, P. Goldfarb, N. Affara, W. McBain, and P. R. Harrison, "The structure of the mouse glutathione peroxidase gene: the selenocysteine in the active site is encoded by the 'termination' codon, TGA," *The EMBO journal*, vol. 5, no. 6, pp. 1221–1227, 1986.

[5] F. Zinoni, A. Birkmann, T. C. Stadtman, and A. Böck, "Nucleotide sequence and expression of the selenocysteine-containing polypeptide of formate dehydrogenase (formate-hydrogen-lyase-linked) from *Escherichia coli*," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 83, no. 13, pp. 4650–4654, 1986.

[6] A. Böck, K. Forchhammer, J. Heider et al., "Selenocysteine: the 21st amino acid," *Molecular Microbiology*, vol. 5, no. 3, pp. 515–520, 1991.

[7] S. A. Burke, S. L. Lo, and J. A. Krzycki, "Clustered genes encoding the methyltransferases of methanogenesis from monomethylamine," *Journal of Bacteriology*, vol. 180, no. 13, pp. 3432–3440, 1998.

[8] L. Paul, D. J. Ferguson Jr., and J. A. Krzycki, "The trimethylamine methyltransferase gene and multiple dimethylamine methyltransferase genes of *Methanosarcina barkeri* contain in-frame and read-through amber codons," *Journal of Bacteriology*, vol. 182, no. 9, pp. 2520–2529, 2000.

[9] B. Hao, W. Gong, T. K. Ferguson, C. M. James, J. A. Krzycki, and M. K. Chan, "A new UAG-encoded residue in the structure of a methanogen methyltransferase," *Science*, vol. 296, no. 5572, pp. 1462–1466, 2002.

[10] B. Hao, G. Zhao, P. T. Kang et al., "Reactivity and chemical synthesis of L-pyrrolysine—the 22nd genetically encoded amino acid," *Chemistry and Biology*, vol. 11, no. 9, pp. 1317–1324, 2004.

[11] G. Srinivasan, C. M. James, and J. A. Krzycki, "Pyrrolysine encoded by UAG in archaea: charging of a UAG-decoding specialized tRNA," *Science*, vol. 296, no. 5572, pp. 1459–1462, 2002.

[12] Y. Zhang, P. V. Baranov, J. F. Atkins, and V. N. Gladyshev, "Pyrrolysine and selenocysteine use dissimilar decoding strategies," *Journal of Biological Chemistry*, vol. 280, no. 21, pp. 20740–20751, 2005.

[13] J. Yuan, P. O'Donoghue, A. Ambrogelly et al., "Distinct genetic code expansion strategies for selenocysteine and pyrrolysine are reflected in different aminoacyl-tRNA formation systems," *FEBS Letters*, vol. 584, pp. 342–349, 2009.

[14] J. B. Jones and T. C. Stadtman, "*Methanococcus vannielii*: culture and effects of selenium and tungsten on growth," *Journal of Bacteriology*, vol. 130, no. 3, pp. 1404–1406, 1977.

[15] W. J. Jones, J. A. Leigh, C. R. Woese, R. S. Wolfe, and F. Mayer, "*Methanococcus jannaschii* sp. nov., an extremely thermophilic methanogen from a submarine hydrothermal vent," *Archives of Microbiology*, vol. 136, no. 4, pp. 254–261, 1983.

[16] W. B. Whitman, E. Ankwanda, and R. S. Wolfe, "Nutrition and carbon metabolism of *Methanococcus voltae*," *Journal of Bacteriology*, vol. 149, no. 3, pp. 852–863, 1982.

[17] P. Bousquet, P. Ciais, J. B. Miller et al., "Contribution of anthropogenic and natural sources to atmospheric methane variability," *Nature*, vol. 443, no. 7110, pp. 439–443, 2006.

[18] R. K. Thauer, A.-K. Kaster, H. Seedorf, W. Buckel, and R. Hedderich, "Methanogenic archaea: ecologically relevant differences in energy conservation," *Nature Reviews Microbiology*, vol. 6, no. 8, pp. 579–591, 2008.

[19] J. G. Ferry, "Enzymology of the fermentation of acetate to methane by Methanosarcina thermophila," *Biofactors*, vol. 6, no. 1, pp. 25–35, 1997.

[20] U. Deppenmeier and V. Müller, "Life close to the thermodynamic limit: how methanogenic archaea conserve energy," in *Bioenergetics: Energy Conservation and Conversion*, G. Schäfer and H. S. Penefsky, Eds., vol. 45, pp. 123–152, Springer, Heidelberg, Germany, 2008.

[21] R. K. Thauer, "Biochemistry of methanogenesis: a tribute to Marjory Stephenson," *Microbiology*, vol. 144, no. 9, pp. 2377–2406, 1998.

[22] D. R. Boone, W. B. Whitman, and P. Rouviere, "Diversity and taxonomy of methanogens," in *Methanogenesis*, J. G. Ferry, Ed., pp. 35–80, Chapman & Hall, New York, NY, USA, 1993.

[23] J. B. Jones and T. C. Stadtman, "Selenium-dependent and selenium-independent formate dehydrogenase of *Methanococcus vannielii*. Separation of the two forms and characterization of the purified selenium independent form," *Journal of Biological Chemistry*, vol. 256, no. 2, pp. 656–663, 1981.

[24] J. B. Jones, G. L. Dilworth, and T. C. Stadtman, "Occurrence of selenocysteine in the selenium-dependent formate dehydrogenase of *Methanococcus vannielii*," *Archives of Biochemistry and Biophysics*, vol. 195, no. 2, pp. 255–260, 1979.

[25] J. A. Vorholt, M. Vaupel, and R. K. Thauer, "A selenium-dependent and a selenium-independent formylmethanofuran dehydrogenase and their transcriptional regulation in the hyperthermophilic *Methanopyrus kandleri*," *Molecular Microbiology*, vol. 23, no. 5, pp. 1033–1042, 1997.

[26] S. Halboth and A. Klein, "*Methanococcus voltae* harbors four gene clusters potentially encoding two [NiFe] and two [NiFeSe] hydrogenases, each of the cofactor F_{420}-reducing or F_{420}-non-reducing types," *Molecular and General Genetics*, vol. 233, no. 1-2, pp. 217–224, 1992.

[27] O. Sorgenfrei, D. Linder, M. Karas, and A. Klein, "A novel very small subunit of a selenium containing [NiFe] hydrogenase of *Methanococcus voltae* is postranslationally processed by cleavage at a defined position," *European Journal of Biochemistry*, vol. 213, no. 3, pp. 1355–1358, 1993.

[28] R. Wilting, S. Schorling, B. C. Persson, and A. Böck, "Selenoprotein synthesis in *Archaea*: identification of an mRNA element of *Methanococcus jannaschii* probably directing selenocysteine insertion," *Journal of Molecular Biology*, vol. 266, no. 4, pp. 637–641, 1997.

[29] T. Stock, M. Selzer, and M. Rother, "*In vivo* requirement of selenophosphate for selenoprotein synthesis in archaea," *Molecular Microbiology*, vol. 75, no. 1, pp. 149–160, 2010.

[30] G. V. Kryukov and V. N. Gladyshev, "The prokaryotic selenoproteome," *EMBO Reports*, vol. 5, no. 5, pp. 538–543, 2004.

[31] D. E. Graham, N. Kyrpides, I. J. Anderson, R. Overbeek, and W. B. Whitman, "Genome of *methanocaldococcus (methanococcus) jannaschii*," *Methods in Enzymology*, vol. 330, pp. 40–123, 2001.

[32] A. Böck and O. Kandler, "Antibiotic sensitivity of archaebacteria," in *Archaebacteria*, C. R. Woese and R. S. Wolfe, Eds., vol. 8, pp. 525–544, Academic Press, New York, NY, USA, 1985.

[33] G. Schmid and A. Böck, "Immunological comparison of ribosomal proteins from archaebacteria," *Journal of Bacteriology*, vol. 147, no. 2, pp. 282–288, 1981.

[34] T. C. Stadtman, "Methane fermentation," *Annual Review of Microbiology*, vol. 21, pp. 121–142, 1967.

[35] A. Böck, M. Rother, M. Leibundgut, and N. Ban, "Selenium metabolism in prokaryotes," in *Seleniumml: Its Molecular Biology and Role in Human Health*, D. L. Hatfield, M. J. Berry, and V. N. Gladyshev, Eds., pp. 9–28, Springer, New York, NY, USA, 2nd edition, 2006.

[36] M. Rother, A. Resch, R. Wilting, and A. Böck, "Selenoprotein synthesis in archaea," *Biofactors*, vol. 14, no. 1–4, pp. 75–83, 2001.

[37] T. Stock and M. Rother, "Selenoproteins in Archaea and Gram-positive bacteria," *Biochimica et Biophysica Acta*, vol. 1790, pp. 1520–1532, 2009.

[38] L. D. Eirich, "Proposed structure for coenzyme F_{420} from *methanobacterium*," *Biochemistry*, vol. 17, no. 22, pp. 4583–4593, 1978.

[39] O. Sorgenfrei, S. Müller, M. Pfeiffer, I. Sniezko, and A. Klein, "The [NiFe] hydrogenases of *Methanococcus voltae*: genes, enzymes and regulation," *Archives of Microbiology*, vol. 167, no. 4, pp. 189–195, 1997.

[40] C. J. Bult, O. White, G. J. Olsen et al., "Complete genome sequence of the Methanogenic archaeon, *Methanococcus jannaschii*," *Science*, vol. 273, no. 5278, pp. 1058–1073, 1996.

[41] E. Setzke, R. Hedderich, S. Heiden, and R. K. Thauer, "H_2: heterodisulfide oxidoreductase complex from *Methanobacterium thermoautotrophicum*. Composition and properties," *European Journal of Biochemistry*, vol. 220, no. 1, pp. 139–148, 1994.

[42] G. Kulkarni, D. M. Kridelbaugh, A. M. Guss, and W. W. Metcalf, "Hydrogen is a preferred intermediate in the energy-conserving electron transport chain of *Methanosarcina barkeri*," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 106, no. 37, pp. 15915–15920, 2009.

[43] P. V. Welander and W. W. Metcalf, "Mutagenesis of the C1 oxidation pathway in *Methanosarcina barkeri*: new insights into the Mtr/Mer bypass pathway," *Journal of Bacteriology*, vol. 190, no. 6, pp. 1928–1936, 2008.

[44] P. van der Meijden, H. J. Heythuysen, A. Pouwels, F. Houwen, C. van der Drift, and G. D. Vogels, "Methyltransferases involved in methanol conversion by *Methanosarcina barkeri*," *Archives of Microbiology*, vol. 134, no. 3, pp. 238–242, 1983.

[45] K. Sauer, U. Harms, and R. K. Thauer, "Methanol: coenzyme M methyltransferase from *Methanosarcina barkeri* purification, properties and encoding genes of the corrinoid protein MT1," *European Journal of Biochemistry*, vol. 243, no. 3, pp. 670–677, 1997.

[46] S. A. Burke and J. A. Krzycki, "Reconstitution of monomethylamine:coenzyme M methyl transfer with a corrinoid protein and two methyltransferases purified from *Methanosarcina barkeri*," *Journal of Biological Chemistry*, vol. 272, no. 26, pp. 16570–16577, 1997.

[47] D. J. Ferguson Jr., N. Gorlatova, D. A. Grahame, and J. A. Krzycki, "Reconstitution of dimethylamine:coenzyme M methyl transfer with a discrete corrinoid protein and two methyltransferases purified from *Methanosarcina barkeri*," *Journal of Biological Chemistry*, vol. 275, no. 37, pp. 29053–29060, 2000.

[48] D. J. Ferguson Jr. and J. A. Krzycki, "Reconstitution of trimethylamine-dependent coenzyme M methylation with the trimethylamine corrinoid protein and the isozymes of methyltransferase II from *Methanosarcina barkeri*," *Journal of Bacteriology*, vol. 179, no. 3, pp. 846–852, 1997.

[49] D. J. Ferguson Jr., J. A. Krzycki, and D. A. Grahame, "Specific roles of methylcobamide:coenzyme M methyltransferase isozymes in metabolism of methanol and methylamines in *Methanosarcina barkeri*," *Journal of Biological Chemistry*, vol. 271, no. 9, pp. 5189–5194, 1996.

[50] T. Ferguson, J. A. Soares, T. Lienard, G. Gottschalk, and J. A. Krzycki, "RamA, a protein required for reductive activation of corrinoid-dependent methylamine methyltransferase reactions in methanogenic archaea," *Journal of Biological Chemistry*, vol. 284, no. 4, pp. 2285–2295, 2009.

[51] L. Paul and J. A. Krzycki, "Sequence and transcript analysis of a novel *Methanosarcina barkeri* methyltransferase II homolog and its associated corrinoid protein homologous to methionine synthase," *Journal of Bacteriology*, vol. 178, no. 22, pp. 6599–6607, 1996.

[52] C. L. Drennan, S. Huang, J. T. Drummond, R. G. Matthews, and M. L. Ludwig, "How a protein binds B_{12}: A 3.0 Å X-ray structure of B_{12}-binding domains of methionine synthase," *Science*, vol. 266, no. 5191, pp. 1669–1674, 1994.

[53] U. Harms and R. K. Thauer, "Methylcobalamin:coenzyme M methyltransferase isoenzymes MtaA and MtbA from *Methanosarcina barkeri*. Cloning, sequencing and differential transcription of the encoding genes, and functional overexpression of the *mtaA* gene in *Escherichia coli*," *European Journal of Biochemistry*, vol. 235, no. 3, pp. 653–659, 1996.

[54] C. H. Hagemeier, M. Krüer, R. K. Thauer, E. Warkentin, and U. Ermler, "Insight into the mechanism of biological methanol activation based on the crystal structure of the methanol-cobalamin methyltransferase complex," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 103, no. 50, pp. 18917–18922, 2006.

[55] J. C. Evans, D. P. Huddler, M. T. Hilgers, G. Romanchuk, R. G. Matthews, and M. L. Ludwig, "Structures of the N-terminal modules imply large domain motions during catalysis by methionine synthase," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 101, no. 11, pp. 3729–3736, 2004.

[56] J. E. Galagan, C. Nusbaum, A. Roy et al., "The genome of *M. acetivorans* reveals extensive metabolic and physiological diversity," *Genome Research*, vol. 12, no. 4, pp. 532–542, 2002.

[57] D. L. Maeder, I. Anderson, T. S. Brettin et al., "The *Methanosarcina barkeri* genome: comparative analysis with *Methanosarcina acetivorans* and *Methanosarcina mazei* reveals extensive rearrangement within methanosarcinal genomes," *Journal of Bacteriology*, vol. 188, no. 22, pp. 7922–7931, 2006.

[58] U. Deppenmeier, A. Johann, T. Hartsch et al., "The genome of *Methanosarcina mazei*: evidence for lateral gene transfer between bacteria and archaea," *Journal of Molecular Microbiology and Biotechnology*, vol. 4, no. 4, pp. 453–461, 2002.

[59] A. Goodchild, N. F. W. Saunders, H. Ertan et al., "A proteomic determination of cold adaptation in the Antarctic archaeon, *Methanococcoides burtonii*," *Molecular Microbiology*, vol. 53, no. 1, pp. 309–321, 2004.

[60] M. A. Pritchett and W. W. Metcalf, "Genetic, physiological and biochemical characterization of multiple methanol methyltransferase isozymes in *Methanosarcina acetivorans* C2A," *Molecular Microbiology*, vol. 56, no. 5, pp. 1183–1194, 2005.

[61] J. A. Soares, L. Zhang, R. L. Pitsch et al., "The residue mass of L-pyrrolysine in three distinct methylamine methyltransferases," *Journal of Biological Chemistry*, vol. 280, no. 44, pp. 36962–36969, 2005.

[62] K. Veit, C. Ehlers, and R. A. Schmitz, "Effects of nitrogen and carbon sources on transcription of soluble methyltransferases in *Methanosarcina mazei* strain Gö1," *Journal of Bacteriology*, vol. 187, no. 17, pp. 6147–6154, 2005.

[63] T. J. Williams, D. W. Burg, H. Ertan et al., "Global proteomic analysis of the insoluble, soluble, and supernatant fractions of the psychrophilic archaeon *Methanococcoides burtonii* part II: the effect of different methylated growth substrates," *Journal of Proteome Research*, vol. 9, no. 2, pp. 653–663, 2010.

[64] S. Asakawa, K. Sauer, W. Liesack, and R. K. Thauer, "Tetramethylammoniumml:coenzyme M methyltransferase system from *Methanococcoides* sp," *Archives of Microbiology*, vol. 170, no. 4, pp. 220–226, 1998.

[65] C. M. James, T. K. Ferguson, J. F. Leykam, and J. A. Krzycki, "The amber codon in the gene encoding the monomethylamine methyltransferase isolated from *Methanosarcina Barkeri* is translated as a sense codon," *Journal of Biological Chemistry*, vol. 276, no. 36, pp. 34252–34258, 2001.

[66] D. G. Longstaff, R. C. Larue, J. E. Faust et al., "A natural genetic code expansion cassette enables transmissible biosynthesis and genetic encoding of pyrrolysine," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 104, no. 3, pp. 1021–1026, 2007.

[67] A. Böck, M. Thanbichler, M. Rother, and A. Resch, "Selenocysteine," in *Aminoacyl-tRNA Synthetases*, M. Ibba, C. S. Francklyn, and S. Cusack, Eds., pp. 320–327, Landes Bioscience, Georgetown, Wash, USA, 2005.

[68] F. Zinoni, J. Heider, and A. Böck, "Features of the formate dehydrogenase mRNA necessary for decoding of the UGA codon as selenocysteine," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 87, no. 12, pp. 4660–4664, 1990.

[69] J. Heider, C. Baron, and A. Böck, "Coding from a distance: dissection of the mRNA determinants required for the incorporation of selenocysteine into protein," *EMBO Journal*, vol. 11, no. 10, pp. 3759–3766, 1992.

[70] C. Baron, J. Heider, and A. Böck, "Interaction of translation factor SELB with the formate dehydrogenase H selenopolypeptide mRNA," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 90, no. 9, pp. 4181–4185, 1993.

[71] M. Thanbichler, A. Böck, and R. S. Goody, "Kinetics of the interaction of translation factor SelB from *Escherichia coli* with guanosine nucleotides and selenocysteine insertion sequence RNA," *Journal of Biological Chemistry*, vol. 275, no. 27, pp. 20458–20466, 2000.

[72] M. Thanbichler and A. Böck, "Functional analysis of prokaryotic SELB proteins," *Biofactors*, vol. 14, no. 1–4, pp. 53–59, 2001.

[73] N. Fischer, A. Paleskava, K. B. Gromadski et al., "Towards understanding selenocysteine incorporation into bacterial proteins," *Biological Chemistry*, vol. 388, no. 10, pp. 1061–1067, 2007.

[74] B. A. Carlson, X.-M. Xu, G. V. Kryukov et al., "Identification and characterization of phosphoseryl-tRNA$^{[Ser]Sec}$ kinase," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 101, no. 35, pp. 12848–12853, 2004.

[75] J. T. Kaiser, K. Gromadski, M. Rother, H. Engelhardt, M. V. Rodnina, and M. C. Wahl, "Structural and functional investigation of a putative archaeal selenocysteine synthase," *Biochemistry*, vol. 44, no. 40, pp. 13315–13327, 2005.

[76] J. Yuan, S. Palioura, J. C. Salazar et al., "RNA-dependent conversion of phosphoserine forms selenocysteine in eukaryotes and archaea," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 103, no. 50, pp. 18923–18927, 2006.

[77] X.-M. Xu, B. A. Carlson, H. Mix et al., "Biosynthesis of selenocysteine on its tRNA in eukaryotes," *PLoS Biology*, vol. 5, no. 1, article no. e4, pp. 0096–0105, 2007.

[78] K. Forchhammer and A. Böck, "Selenocysteine synthase from *Escherichia coli*: analysis of the reaction sequence," *Journal of Biological Chemistry*, vol. 266, no. 10, pp. 6324–6328, 1991.

[79] E. Aeby, S. Palioura, M. Pusnik et al., "The canonical pathway for selenocysteine insertion is dispensable in trypanosomes," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 106, no. 13, pp. 5088–5092, 2009.

[80] O. M. Ganichkin, X.-M. Xu, B. A. Carlson et al., "Structure and catalytic mechanism of eukaryotic selenocysteine synthase," *Journal of Biological Chemistry*, vol. 283, no. 9, pp. 5849–5865, 2008.

[81] Y. Araiso, S. Palioura, R. Ishitani et al., "Structural insights into RNA-dependent eukaryal and archaeal selenocysteine formation," *Nucleic Acids Research*, vol. 36, no. 4, pp. 1187–1199, 2008.

[82] Y. Araiso, R. L. Sherrer, R. Ishitani, J. M. L. Ho, D. Söll, and O. Nureki, "Structure of a tRNA-dependent kinase essential for selenocysteine decoding," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 106, no. 38, pp. 16215–16220, 2009.

[83] S. Palioura, R. L. Sherrer, T. A. Steitz, D. Soil, and M. Simonovic, "The human SepSecS-tRNA^{Sec} complex reveals the mechanism of selenocysteine formation," *Science*, vol. 325, no. 5938, pp. 321–325, 2009.

[84] Y. Itoh, S. Chiba, S.-I. Sekine, and S. Yokoyama, "Crystal structure of human selenocysteine tRNA," *Nucleic Acids Research*, vol. 37, no. 18, pp. 6259–6268, 2009.

[85] R. S. Glass, W. P. Singh, W. Jung, Z. Veres, T. D. Scholz, and T. C. Stadtman, "Monoselenophosphate: synthesis, characterization, and identity with the prokaryotic biological selenium donor, compound SePX," *Biochemistry*, vol. 32, no. 47, pp. 12555–12559, 1993.

[86] A. Ehrenreich, K. Forchhammer, P. Tormay, B. Veprek, and A. Böck, "Selenoprotein synthesis in *E. coli*. Purification and characterisation of the enzyme catalysing selenium activation," *European Journal of Biochemistry*, vol. 206, no. 3, pp. 767–773, 1992.

[87] J. Lu, L. Zhong, M. E. Lönn, R. F. Burk, K. E. Hill, and A. Holmgren, "Penultimate selenocysteine residue replaced by cysteine in thioredoxin reductase from selenium-deficient rat liver," *FASEB Journal*, vol. 23, no. 8, pp. 2394–2402, 2009.

[88] J. Yuan, M. J. Hohn, R. L. Sherrer, S. Palioura, D. Su, and D. Söll Dieter, "A tRNA-dependent cysteine biosynthesis enzyme recognizes the selenocysteine-specific tRNA in *Escherichia coli*," *FEBS Letters*, vol. 584, no. 13, pp. 2857–2861, 2010.

[89] W. T. Self, R. Pierce, and T. C. Stadtman, "Cloning and heterologous expression of a *Methanococcus vannielii* gene encoding a selenium-binding protein," *IUBMB Life*, vol. 56, no. 8, pp. 501–507, 2004.

[90] K. G. Patteson, N. Trivedi, and T. C. Stadtman, "*Methanococcus vannielii* selenium-binding protein (SeBP): chemical reactivity of recombinant SeBP produced in *Escherichia coli*," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 102, no. 34, pp. 12029–12034, 2005.

[91] M. Rother, R. Wilting, S. Commans, and A. Böck, "Identification and characterisation of the selenocysteine-specific translation factor SelB from the archaeon *Methanococcus jannaschii*," *Journal of Molecular Biology*, vol. 299, no. 2, pp. 351–358, 2000.

[92] D. Fagegaltier, N. Hubert, K. Yamada, T. Mizutani, P. Carbon, and A. Krol, "Characterization of mSelB, a novel mammalian elongation factor for selenoprotein translation," *EMBO Journal*, vol. 19, no. 17, pp. 4796–4805, 2000.

[93] R. M. Tujebajeva, P. R. Copeland, X.-M. Xu et al., "Decoding apparatus for eukaryotic selenocysteine insertion," *EMBO Reports*, vol. 1, no. 2, pp. 158–163, 2000.

[94] L. Chavatte, B. A. Brown II, and D. M. Driscoll, "Ribosomal protein L30 is a component of the UGA-selenocysteine recoding machinery in eukaryotes," *Nature Structural and Molecular Biology*, vol. 12, no. 5, pp. 408–416, 2005.

[95] P. R. Copeland, J. E. Fletcher, B. A. Carlson, D. L. Hatfield, and D. M. Driscoll, "A novel RNA binding protein, SBP2, is required for the translation of mammalian selenoprotein mRNAs," *EMBO Journal*, vol. 19, no. 2, pp. 306–314, 2000.

[96] C. Allmang, L. Wurth, and A. Krol, "The selenium to selenoprotein pathway in eukaryotes: more molecular partners than anticipated," *Biochimica et Biophysica Acta*, vol. 1790, pp. 1415–1423, 2009.

[97] J. E. Squires and M. J. Berry, "Eukaryotic selenoprotein synthesis: mechanistic insight incorporating new factors and new functions for old factors," *IUBMB life*, vol. 60, no. 4, pp. 232–235, 2008.

[98] C. Buettner, J. W. Harney, and P. R. Larsen, "The 3′-untranslated region of human type 2 lodothyronine deiodinase mRNA contains a functional selenocysteine insertion sequence element," *Journal of Biological Chemistry*, vol. 273, no. 50, pp. 33374–33378, 1998.

[99] G. W. Martin III, J. W. Harney, and M. J. Berry, "Selenocysteine incorporation in eukaryotes: insights into mechanism and efficiency from sequence, structure, and spacing proximity studies of the type 1 deiodinase SECIS element," *RNA*, vol. 2, no. 2, pp. 171–182, 1996.

[100] A. I. Slesarev, K. V. Mezhevaya, K. S. Makarova et al., "The complete genome of hyperthermophile *Methanopyrus kandleri* AV19 and monophyly of archaeal methanogens," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 99, no. 7, pp. 4644–4649, 2002.

[101] E. L. Hendrickson, R. Kaul, Y. Zhou et al., "Complete genome sequence of the genetically tractable hydrogenotrophic methanogen *Methanococcus maripaludis*," *Journal of Bacteriology*, vol. 186, no. 20, pp. 6956–6969, 2004.

[102] D. L. Tumbula and W. B. Whitman, "Genetics of *Methanococcus*: possibilities for functional genomics in archaea," *Molecular Microbiology*, vol. 33, no. 1, pp. 1–7, 1999.

[103] M. Rother and W. W. Metcalf, "Genetic technologies for Archaea," *Current Opinion in Microbiology*, vol. 8, no. 6, pp. 745–751, 2005.

[104] C. Polycarpo, A. Ambrogelly, B. Ruan et al., "Activation of the pyrrolysine suppressor tRNA requires formation of a ternary complex with class I and class II Lysyl-tRNA synthetases," *Molecular Cell*, vol. 12, no. 2, pp. 287–294, 2003.

[105] A. Mahapatra, G. Srinivasan, K. B. Richter et al., "Class I and class II lysyl-tRNA synthetase mutants and the genetic encoding of pyrrolysine in *Methanosarcina* spp," *Molecular Microbiology*, vol. 64, no. 5, pp. 1306–1318, 2007.

[106] J. A. Krzycki, "The direct genetic encoding of pyrrolysine," *Current Opinion in Microbiology*, vol. 8, no. 6, pp. 706–712, 2005.

[107] S. K. Blight, R. C. Larue, A. Mahapatra et al., "Direct charging of tRNACUA with pyrrolysine *in vitro* and *in vivo*," *Nature*, vol. 431, no. 7006, pp. 333–335, 2004.

[108] C. Polycarpo, A. Ambrogelly, A. Bérubé et al., "An aminoacyl-tRNA synthetase that specifically activates pyrrolysine," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 101, no. 34, pp. 12450–12454, 2004.

[109] T. Fekner, X. Li, M. M. Lee, and M. K. Chan, "A pyrrolysine analogue for protein click chemistry," *Angewandte Chemie*, vol. 48, no. 9, pp. 1633–1635, 2009.

[110] T. Mukai, T. Kobayashi, N. Hino, T. Yanagisawa, K. Sakamoto, and S. Yokoyama, "Adding l-lysine derivatives to the genetic code of mammalian cells with engineered pyrrolysyl-tRNA synthetases," *Biochemical and Biophysical Research Communications*, vol. 371, no. 4, pp. 818–822, 2008.

[111] P. R. Chen, D. Groff, J. Guo et al., "A facile system for encoding unnatural amino acids in mammalian cells," *Angewandte Chemie*, vol. 48, no. 22, pp. 4052–4055, 2009.

[112] R. Ishii, O. Nureki, T. Yanagisawa, R. Fukunaga, and S. Yokoyama, "Crystallization and preliminary X-ray crystallographic analysis of the catalytic domain of pyrrolysyl-tRNA synthetase from the methanogenic archaeon *Methanosarcina mazei*," *Acta Crystallographica Section F*, vol. 62, no. 10, pp. 1031–1033, 2006.

[113] J. M. Kavran, S. Gundllapalli, P. O'Donoghue, M. Englert, D. Söll, and T. A. Steitz, "Structure of pyrrolysyl-tRNA synthetase, an archaeal enzyme for genetic code innovation," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 104, no. 27, pp. 11268–11273, 2007.

[114] T. Yanagisawa, R. Ishii, R. Fukunaga, T. Kobayashi, K. Sakamoto, and S. Yokoyama, "Crystallographic studies on multiple conformational states of active-site loops in pyrrolysyl-tRNA synthetase," *Journal of Molecular Biology*, vol. 378, no. 3, pp. 634–652, 2008.

[115] K. Nozawa, P. O'Donoghue, S. Gundllapalli et al., "Pyrrolysyl-tRNA synthetase-tRNA^{Pyl} structure reveals the molecular basis of orthogonality," *Nature*, vol. 457, no. 7233, pp. 1163–1167, 2009.

[116] W.-T. Li, A. Mahapatra, D. G. Longstaff et al., "Specificity of pyrrolysyl-tRNA synthetase for pyrrolysine and pyrrolysine analogs," *Journal of Molecular Biology*, vol. 385, no. 4, pp. 1156–1164, 2009.

[117] T. Yanagisawa, R. Ishii, R. Fukunaga, T. Kobayashi, K. Sakamoto, and S. Yokoyama, "Multistep engineering of pyrrolysyl-tRNA synthetase to genetically encode N_{ε}-(o-azidobenzyloxycarbonyl) lysine for site-specific protein modification," *Chemistry and Biology*, vol. 15, no. 11, pp. 1187–1197, 2008.

[118] H. Neumann, S. Y. Peak-Chew, and J. W. Chin, "Genetically encoding N_{ε}-acetyllysine in recombinant proteins," *Nature Chemical Biology*, vol. 4, no. 4, pp. 232–234, 2008.

[119] S. Herring, A. Ambrogelly, S. Gundllapalli, P. O'Donoghue, C. R. Polycarpo, and D. Söll, "The amino-terminal domain of pyrrolysyl-tRNA synthetase is dispensable in vitro but required for in vivo activity," *FEBS Letters*, vol. 581, no. 17, pp. 3197–3203, 2007.

[120] A. Théobald-Dietrich, M. Frugier, R. Giegé, and J. Rudinger-Thirion, "Atypical archaeal tRNA pyrrolysine transcript behaves towards EF-Tu as a typical elongator tRNA," *Nucleic Acids Research*, vol. 32, no. 3, pp. 1091–1096, 2004.

[121] J. A. Krzycki, "Function of genetically encoded pyrrolysine in corrinoid-dependent methylamine methyltransferases," *Current Opinion in Chemical Biology*, vol. 8, no. 5, pp. 484–491, 2004.

[122] P. A. Frey, A. D. Hegeman, and F. J. Ruzicka, "The radical SAM superfamily," *Critical Reviews in Biochemistry and Molecular Biology*, vol. 43, no. 1, pp. 63–88, 2008.

[123] Y. Zhang and V. N. Gladyshev, "High content of proteins containing 21st and 22nd amino acids, selenocysteine and pyrrolysine, in a symbiotic deltaproteobacterium of gutless worm *Olavius algarvensis*," *Nucleic Acids Research*, vol. 35, no. 15, pp. 4952–4963, 2007.

[124] D. G. Longstaff, S. K. Blight, L. Zhang, K. B. Green-Church, and J. A. Krzycki, "*In vivo* contextual requirements for UAG translation as pyrrolysine," *Molecular Microbiology*, vol. 63, no. 1, pp. 229–241, 2007.

[125] O. Namy, J.-P. Rousset, S. Napthine, and I. Brierley, "Reprogrammed genetic decoding in cellular gene expression," *Molecular Cell*, vol. 13, no. 2, pp. 157–168, 2004.

[126] A. Théobald-Dietrich, R. Giegé, and J. Rudinger-Thirion, "Evidence for the existence in mRNAs of a hairpin element responsible for ribosome dependent pyrrolysine insertion into proteins," *Biochimie*, vol. 87, no. 9-10, pp. 813–817, 2005.

[127] L. Paul, *Analysis of the genes encoding the enzymes initiating methanogenesis from methylthols, trimethylamine, and dimethylamine in Methanosarcina barkeri*, MS. Ph.D., Ohio State University, Columbus, Ohio, USA, 1999.

[128] J. A. Krzycki, "UAG translation as pyrrolysine," in *Recoding: Expansion of Decoding Rules Enriches Gene Expression*, J. F. Atkins and R. F. Gesteland, Eds., vol. 24, pp. 53–77, Springer, New York, NY, USA, 2010.

[129] E. Alkalaeva, B. Eliseev, A. Ambrogelly et al., "Translation termination in pyrrolysine-utilizing archaea," *FEBS Letters*, vol. 583, no. 21, pp. 3455–3460, 2009.

[130] M. Ibba and D. Söll, "Aminoacyl-tRNAs: setting the limits of the genetic code," *Genes and Development*, vol. 18, no. 7, pp. 731–738, 2004.

[131] M. Li, Y. Huang, and Y. Xiao, "A method for identification of selenoprotein genes in archaeal genomes,," *Genomics, Proteomics and Bioinformatics*, vol. 7, no. 1-2, pp. 62–70, 2009.

[132] D. E. Fomenko, W. Xing, B. M. Adair, D. J. Thomas, and V. N. Gladyshev, "High-throughput identification of catalytic redox-active cystein residues," *Science*, vol. 315, no. 5810, pp. 387–389, 2007.

[133] M. Rother, I. Mathes, F. Lottspeich, and A. Böck, "Inactivation of the selB gene in *Methanococcus maripaludis*: effect on synthesis of selenoproteins and their sulfur-containing homologs," *Journal of Bacteriology*, vol. 185, no. 1, pp. 107–114, 2003.

[134] C. Buettner, J. W. Harney, and P. R. Larsen, "The role of selenocysteine 133 in catalysis by the human type 2 iodothyronine deiodinase," *Endocrinology*, vol. 141, no. 12, pp. 4606–4612, 2000.

[135] M. J. Axley, A. Böck, and T. C. Stadtman, "Catalytic properties of an *Escherichia coli* formate dehydrogenase mutant in which sulfur replaces selenium," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 88, no. 19, pp. 8450–8454, 1991.

[136] S. Gromer, L. Johansson, H. Bauer et al., "Active sites of thioredoxin reductases: why selenoproteins?" *Proceedings of the National Academy of Sciences of the United States of America*, vol. 100, no. 22, pp. 12618–12623, 2003.

[137] R. E. Huber and R. S. Criddle, "Comparison of the chemical properties of selenocysteine and selenocystine with their sulfur analogs," *Archives of Biochemistry and Biophysics*, vol. 122, no. 1, pp. 164–173, 1967.

[138] F. Zinoni, A. Birkmann, W. Leinfelder, and A. Böck, "Cotranslational insertion of selenocysteine into formate dehydrogenase from *Escherichia coli* directed by a UGA codon," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 84, no. 10, pp. 3156–3160, 1987.

[139] G. Bulaj, T. Kortemme, and D. P. Goldenberg, "Ionization-reactivity relationships for cysteine thiols in polypeptides," *Biochemistry*, vol. 37, no. 25, pp. 8965–8972, 1998.

[140] S. Toppo, L. Flohé, F. Ursini, S. Vanin, and M. Maiorino, "Catalytic mechanisms and specificities of glutathione peroxidases: variations of a basic scheme," *Biochimica et Biophysica Acta*, vol. 1790, pp. 1486–1500, 2009.

[141] S. Müller, H. Senn, B. Gsell, W. Vetter, C. Baron, and A. Böck, "The formation of diselenide bridges in proteins by incorporation of selenocysteine residues: biosynthesis and characterization of $(Se)_2$-thioredoxin," *Biochemistry*, vol. 33, no. 11, pp. 3404–3412, 1994.

[142] N. Metanis, E. Keinan, and P. E. Dawson, "Synthetic seleno-glutaredoxin 3 analogues are highly reducing oxidoreductases with enhanced catalytic efficiency," *Journal of the American Chemical Society*, vol. 128, no. 51, pp. 16684–16691, 2006.

[143] E. S.J. Arnér, "Selenoproteins-What unique properties can arise with selenocysteine in place of cysteine?" *Experimental Cell Research*, vol. 316, no. 8, pp. 1296–1303, 2010.

[144] E. L. Ruggles and R. J. Hondal, "Differing views of the role of selenium in thioredoxin reductase," *Amino Acids*.

[145] T. Nauser, S. Dockheer, R. Kissner, and W. H. Koppenol, "Catalysis of electron transfer by selenocysteine," *Biochemistry*, vol. 45, no. 19, pp. 6038–6043, 2006.

[146] S. H. Zinder, "Physiological ecology of methanogens," in *Methanogenesis*, J. G. Ferry, Ed., pp. 128–206, Chapman & Hall, New York, NY, USA, 1993.

[147] Y. Zhang, H. Romero, G. Salinas, and V. N. Gladyshev, "Dynamic evolution of selenocysteine utilization in bacteria: a balance between selenoprotein loss and evolution of selenocysteine from redox active cysteine residues," *Genome Biology*, vol. 7, no. 10, 2006.

[148] H. Romero, Y. Zhang, V. N. Gladyshev, and G. Salinas, "Evolution of selenium utilization traits," *Genome biology*, vol. 6, no. 8, p. R66, 2005.

[149] A. Böck and M. Rother, "A pseudo-SECIS element in *Methanococcus voltae* documents evolution of a selenoprotein into a sulphur-containing homologue," *Archives of Microbiology*, vol. 183, no. 2, pp. 148–150, 2005.

[150] H. L. Ehrlich and D. K. Newman, "Geomicrobiology of selenium and tellurium," in *Geomicrobiology*, pp. 527–536, CRC Press, Boca Raton, Fla, USA, 2009.

[151] G. J. Rapp, "Seleniumml: element and geochemistry," in *The Encyclopedia of Geochemistry and Environmental Sciences*, R. W. Fairbridge, Ed., vol. IVA, pp. 1079–1080, Van Nostrand Reinhold, New York, NY, USA, 1972.

[152] J. F. Stolz and R. S. Oremland, "Bacterial respiration of arsenic and selenium," *FEMS Microbiology Reviews*, vol. 23, no. 5, pp. 615–627, 1999.

[153] J. F. Stolz, P. Basu, J. M. Santini, and R. S. Oremland, "Arsenic and selenium in microbial metabolism," *Annual Review of Microbiology*, vol. 60, pp. 107–130, 2006.

[154] C. A. Woolfolk and H. R. Whiteley, "Reduction of inorganic compounds with molecular hydrogen by Micrococcus lactilyticus. I. Stoichiometry with compounds of arsenic, selenium, tellurium, transition and other elements," *Journal of bacteriology*, vol. 84, pp. 647–658, 1962.

[155] A. Mahapatra, A. Patel, J. A. Soares et al., "Characterization of a *Methanosarcina acetivorans* mutant unable to translate UAG as pyrrolysine," *Molecular Microbiology*, vol. 59, no. 1, pp. 56–66, 2006.

[156] A. M. Krishnakumar, D. Sliwa, J. A. Endrizzi, E. S. Boyd, S. A. Ensign, and J. W. Peters, "Getting a handle on the role of coenzyme M in alkene metabolism," *Microbiology and Molecular Biology Reviews*, vol. 72, no. 3, pp. 445–456, 2008.

[157] S. F. Altschul, W. Gish, W. Miller, E. W. Myers, and D. J. Lipman, "Basic local alignment search tool," *Journal of Molecular Biology*, vol. 215, no. 3, pp. 403–410, 1990.

[158] J. Rétey, "Discovery and role of methylidene imidazolone, a highly electrophilic prosthetic group," *Biochimica et Biophysica Acta*, vol. 1647, no. 1-2, pp. 179–184, 2003.

[159] M. G. Abad, B. S. Rao, and J. E. Jackman, "Template-dependent 3′–5′ nucleotide addition is a shared feature of tRNAHis guanylyltransferase enzymes from multiple domains of life," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 107, no. 2, pp. 674–679, 2010.

[160] I. U. Heinemann, P. O'Donoghue, C. Madinger et al., "The appearance of pyrrolysine in $tRNA^{His}$ guanylyltransferase by neutral evolution," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 106, no. 50, pp. 21103–21108, 2009.

Hindawi Publishing Corporation
Archaea
Volume 2010, Article ID 820681, 9 pages
doi:10.1155/2010/820681

Review Article

Protein Acetylation in Archaea, Bacteria, and Eukaryotes

Jörg Soppa

Institute for Molecular Biosciences, Goethe University, Max-von-Laue-StraBe 9, 60438 Frankfurt, Germany

Correspondence should be addressed to Jörg Soppa, soppa@bio.uni-frankfurt.de

Received 1 June 2010; Accepted 22 July 2010

Academic Editor: Jerry Eichler

Proteins can be acetylated at the alpha-amino group of the N-terminal amino acid (methionine or the penultimate amino acid after methionine removal) or at the epsilon-amino group of internal lysines. In eukaryotes the majority of proteins are N-terminally acetylated, while this is extremely rare in bacteria. A variety of studies about N-terminal acetylation in archaea have been reported recently, and it was revealed that a considerable fraction of proteins is N-terminally acetylated in haloarchaea and *Sulfolobus*, while this does not seem to apply for methanogenic archaea. Many eukaryotic proteins are modified by differential internal acetylation, which is important for a variety of processes. Until very recently, only two bacterial proteins were known to be acetylation targets, but now 125 acetylation sites are known for *E. coli*. Knowledge about internal acetylation in archaea is extremely limited; only two target proteins are known, only one of which—Alba—was used to study differential acetylation. However, indications accumulate that the degree of internal acetylation of archaeal proteins might be underestimated, and differential acetylation has been shown to be essential for the viability of haloarchaea. Focused proteomic approaches are needed to get an overview of the extent of internal protein acetylation in archaea.

1. Introduction

Many different forms of posttranslational modifications of proteins have been characterized. Posttranslational modifications can influence many different features of proteins, for example, their folding, activity, stability, antigenicity, intracellular localization, and interaction with other proteins or with nucleic acids. The fraction of posttranslationally modified proteins and thus the importance of posttranslational modification is generally believed to be very different for eukaryotes—having a high fraction of modified proteins—and prokaryotes, which are thought to harbor only very few modified proteins.

For eukaryotes it is thought that acetylation is the most common covalent modification out of 200 types that have been reported [1]. It has also been argued that acetylation is a regulatory modification of the same importance as phosphorylation [2]. The arguments were that acetylation, like phosphorylation, affects many different proteins, can have a variety of consequences, and can regulate key cellular processes in response to extracellular signals [2]. Nevertheless, the wealth of experimental data on protein phosphorylation in eukaryotes is much higher than on acetylation, and in addition, it was typically generated earlier.

In stark contrast to eukaryotic proteins, until very recently only very few bacterial proteins were known to be acetylated. It has long been thought that this would also be true for archaeal proteins, and thus several years ago it was summarized that the available results underscore the conviction that N-terminal acetylation is fundamentally different in eukaryotes compared to archaea and bacteria [3]. The belief that protein acetylation plays an insignificant role in Archaea also held true in the archaeal community for example, in a review about "posttranslation protein modification in Archaea" only about 1% of the text was dedicated to protein acetylation [4].

However, results obtained during recent years have revealed that this belief is far from being true and that—in contrast—N-terminal protein acetylation adds to the ever growing number of biological functions that combine eukaryotes and archaea to the exclusion of the bacteria. Earlier recognized examples of similarity between eukaryotic and archaeal systems are the proteins and DNA elements of the transcription initiation machinery [5, 6], the presence of

histones [7], the translation initiation factor repertoire [8], and replication proteins [9].

However, the information about protein acetylation in archaea is still pretty limited. Therefore, this paper plays a dual role on the one hand it will summarize the results about protein acetylation in archaea, and on the other hand it has the hope to nudge further research in this fascinating and mainly unexplored area. Two different types of protein acetylation are known, which will be discussed sequentially, that is, (1) the acetylation of the alpha-amino group of the N-terminal amino acid of proteins and (2) the acetylation of the epsilon-amino group of internal lysine residues. The former reaction is irreversible, and acetylated proteins stay modified until their degradation, in contrast to the latter, which is reversible and its biological role is thought to be the differential regulation of proteins, for example, the DNA-binding affinity of histones, the most extensively studied example of internally acetylated eukaryotic proteins.

2. N-Terminal Protein Acetylation in Eukaryotes and in Bacteria

The occurrence of N-terminal acetylation of proteins in eukaryotes was discovered very early; the first example was described in 1958. In the next decades many examples were found, typically when protein sequencing via the standard method of that time, Edman degradation, was unsuccessful because the N-terminus was "blocked" and alternative methods were developed to unravel the nature of the chemical modification of the alpha amino group. In a review in 1985, already more than 300 examples were listed, which represented more than 100 different proteins, many of which had been analyzed from several species [10]. Recently, techniques have been developed to specifically enrich N-terminal peptides and use mass spectrometric techniques for large-scale analyses of the correct start codon, methionine removal, and covalent modification of the alpha amino group [11, 12]. Examples for their application are the determination of more than 900 N-termini from human HELA cells and of more than 1200 N-termini from *Drosophila melanogaster* proteins [13, 14]. The new large-scale studies underscore that knowledge about N-terminal protein acetylation in eukaryotes reviewed earlier [1, 3] still holds true and give it a higher statistical validation. The identity of the N-terminal amino acid is determined by the activity of the methionine aminopeptidase (MAP). In fact the methionine is removed by MAP from the vast majority of eukaryotic proteins, and thus most proteins start with the second (penultimate) amino acid of the genomically encoded open reading frame. Eukaryotes contain three different N-terminal acetyltransferases (Nat) termed NatA, NatB, and NatC, which have different substrate specificities. They acetylate a high fraction of proteins, for example, around 60% in yeast, 30% in *D. melanogaster,* and more than 80% in humans [14, 15]. NatA acetylates proteins beginning with the N-terminal amino acids Ser, Ala, Gly, or Thr. Therefore NatA depends on the prior action of MAP. As the N-terminal methionine is removed from the majority of proteins, NatA is the major player responsible for N-terminal acetylation. It could be shown that NatA is highly conserved and a yeast NatA mutant could be rescued by human NatA, which acetylated in yeast nearly the same set of proteins as the endogenous NatA, albeit only partially [15]. NatB is specific for proteins starting with methionine and a bulky hydrophobic amino acid in the second (penultimate) position, while NatC is specific for the N-termini Met-Glu and Met-Asp. NatB and NatC are therefore independent of methionine removal by MAP. N-terminal acetylation in eukaryotes occurs cotranslationally, when the amino terminus leaves the ribosome and is complete when the first 40–50 amino acids have been synthesized [3, 16]. In spite of the fact that N-terminal protein acetylation in eukaryotes is known for more than 50 years, the biological function of acetylation of a major fraction of proteins is still not really known. For several proteins, differences between the acetylated and nonacetylated form have been described, including differential activity, biological half life, or thermal stability [3, 17], but a general role of acetylation has not yet been found. An obvious function that has repeatedly been discussed is stabilization of proteins from proteolytic degradation, and the finding that acetylated proteins are strongly overrepresented among the most abundant proteins [18] is in line with this view. However, recently also the opposite has been proposed, namely, that N-terminal acetylation creates specific degradation signals [19]. Irrespective of the clarification of these opposing views, it is clear that N-terminal acetylation is of great significance *in vivo*, because single deletion mutants of all three genes encoding the three Nats in yeast are highly impaired. Characterization of the human homologs has revealed that depletion of hNatB results in a disruption of normal cell cycle progression, while depletion of hNatC induces apoptosis and cell death [20, 21].

In stark contrast to the high fraction of N-terminally acetylated proteins in eukaryotes, only very few proteins were found to be N-terminally acetylated in bacteria. Only five *E. coli* proteins are known to be N-terminally acetylated, three of them being the ribosomal proteins S5, S12, and S18. In contrast to eukaryotes, acetylation in *E. coli* occurs posttranslationally. Acetylation of S12 has been shown to stablize the ribosomal stalk complex [22]. Three proteins with Nat activity are known, RimI, RimJ, and RimL. However, at least RimJ has a dual function; a native RimJ as well as an acetylation-deficient variant could suppress the different phenotypes of a ribosomal protein S5 (G28D) mutation, that is, cold-sensitivity, anomalous ribosomal profiles and mRNA misreading. RimJ was found to be associated with the pre-30S subunit, and it was proposed that it is a ribosome assembly factor in addition to being an N-terminal acetyl transferase [23].

3. N-Terminal Protein Acetylation in the Archaea

Already 20 years ago it was reported that a few ribosomal proteins of haloarchaea are acetylated at their N-terminus [24–26]. However, this did not increase the interest in

protein acetylation in archaea because on the one hand it perfectly fitted to the bacterial paradigm and thus seemed to underscore that protein acetylation is a rare event also in archaea, and on the other hand only extremely few reports about the acetylation of archaeal proteins appeared. The reason why the degree of N-terminal acetylation was overlooked for so long is probably that the acetylation typically is only partial (see below) and that therefore the un-acetylated fraction allowed successful N-terminal sequencing, and therefore ample "proof" was generated that archaeal proteins are blocked in only extreme rare cases.

One large-scale proteomics study with *Halobacterium salinarum* and *Natronomonas pharaonis* completely changed the picture and currently dominates the view about the degree of N-terminal acetylation in "the archaea." It should be noted that this is no exception but that the current view about protein acetylation in "the eukaryotes" and "the bacteria" is also based on studies with only one or very few species. The above-mentioned large-scale proteomic study let to the identification of 606 N-terminal peptides from *H. salinarum* and 328 N-terminal peptides from *N. pharaonis*, adding up to a sum of 934 N-termini. On the one hand, the results were used to enhance the reliability of start codon assignments, which is not trivial in high G+C-rich genomes, and to reveal the extent and the rules of N-terminal methionine cleavage in haloarchaea [27].

On the other hand, the results were used to specifically address the question of the degree of N-terminal protein acetylation in haloarchaea [28]. It turned out that the N-terminal methionine was cleaved from about two-thirds of all haloarchaeal proteins. Cleavage occurred when the penultimate amino acid was small (glycine, alanine, proline, valine, serine, threonine). The substrate specificity for haloarchaeal methionine aminopeptidases (MAPs) matches the specificity of bacterial and eukaryotic MAPs, and thus the biochemistry of methionine removal appears to be universally conserved. Surprisingly, it was found that N-terminal acetylation is not uncommon in haloarchaea, but that 14% to 19% of the proteins are N-terminally acetylated.

This is in stark contrast to *E. coli* and reminds more of the situation in eukaryotes; albeit the fraction of acetylated proteins is somewhat smaller. Acetylation occurred nearly exclusively after methionine removal, and only serine and alanine as penultimate amino acids were acetylated [28]. It was found that also the antepenultimate position (position three of the open reading frame and position two in the protein after methionine removal) has a strong influence on protein acetylation. Acetylation is favoured when serine, alanine, and glycine are in the antepenultimate position, while, in contrast, aspartic acid and glutamic acid in this position strongly interfere with acetylation. Therefore, while the degree of acetylation somewhat resembles that of eukaryotic proteins, the substrate specificities of the respective NATs are different as eukaryotic proteins are regularly acetylated when acidic amino acids are in the antepenultimate position [3]. The acetylation pattern is most similar to the substrate specificity of the yeast NatA enzyme, while NatB- and NatC-like activities are missing in haloarchaea.

Very few exceptions from the "haloarchaeal acetylation rules" were observed, most of which were conserved between both species. Examples are the alpha subunit of the proteasome (see below), the beta subunit of prefoldin and a hypothetical protein. This indicates that in addition to one or more general Nats, also one or very few additional acetyl transferases with a very high substrate specificity exist in haloarchaea.

It was also shown that the acetylation efficiency for the majority of acetylated proteins is not 100%, but that it differed protein specifically from 13% to 100%. This could be an indication that N-terminal protein acetylation in archaea does not occur cotranslationally, like in eukaryotes, but posttranslatinally. While the genes for putative Nats have been detected in haloarchaeal genomes, as yet no experimental data about the molecular mechanism of N-terminal acetylation are available.

A few additional proteomic studies enable first estimations about how general these results are true for "the archaea." A proteomic study with a third haloarchaeon, *Haloferax volcanii*, representing a third genus, mostly underscored the results of the study discussed above [29]. N-terminal peptides of 236 proteins were identified, and the initial methionine was removed in 70% of all cases. 29% of all proteins were N-terminally acetylated, a fraction that is somewhat higher than for the other two haloarchaeal species. One explanation could be that the degree of acetylation is higher in *Haloferax* than in *Halobacterium* and *Nitrosomonas*; an alternative explanation could be that the fraction of identified proteins was lower for *H. volcanii* and that the degree of acetylation is higher for abundant proteins (like in eukaryotes). Nevertheless, the study showed that significant N-terminal acetylation occurs in at least three different haloarchaeal genera and thus can probably be generalized to all haloarchaea.

The situation seems to be different for another group of the Euryarchaeota, the methanogenic archaea. Two proteomic studies are available for *Methanococcus jannaschii*, which identified 72 proteins and 963 proteins, respectively [30, 31]. Only a single acetylation site at an internal lysine was reported [30], while N-terminal acetylation is not mentioned at all. Of course one explanation could be that it was overlooked based on the—at that time—general belief that it does not occur in archaea anyhow. However, I find it more probable to assume that it would have been detected, like in haloarchaea, if it would also occur in methanogenic archaea. Therefore, the two proteomics studies as well as many studies about individual proteins can be taken as an indication that N-terminal acetylation in methanogenic archaea is very rare or does not occur.

A very limited proteomic survey of N-terminal acetylation was performed with *Sulfolobus solfataricus*, a species belonging to the kingdom of Crenarchaeota [32]. Of the 26 N-terminal peptides identified, 17 were acetylated. Although no "per cent" values can be calculated due to the low absolute numbers of proteins, the fraction of N-terminally acetylated proteins appears to be much higher than for haloarchaea. Clearly a more general study is needed to clarify if these values are of statistical significance. If this would turn out

to be true, *Sulfolobales* and maybe Crenarchaeota as a whole would be more alike eukaryotes than all hitherto tested species of Euryarchaeota. For both Cren- and Euryarchaeota it was found that acetylation occurred at penultimate serine and alanine residues, and for both groups partial acetylation was observed.

The *S. solfataricus* genome was found to contain one gene encoding a putative Nat (sso0209), which was named ssArd1 based on a sequence identity of 37% with the human Ard1, a homologue of yeast NatA. The protein was heterologously produced, and it was confirmed that it can acetylate the N-terminus of Alba, a DNA-binding protein of *Sulfolobus* [32]. A variety of Alba mutants as well as additional *Sulfolobus* proteins were used to characterize the substrate specificity of ssArd1. Surprisingly it was found that in addition to N-terminal Ser and Ala (a NatA-like activity) also the N-termini Met-Glu and Met-Leu were acetylated (NatC- and NatB-like activity). Therefore, it was concluded that the situation in *Sulfolobus* represents an ancestral state with a single Nat, which is not part of a protein complex and which has a broader substrate specificity compared to the eukaryotic Nats. Eukaryotic Nats have later experienced gene duplications and have evolved further into specialized proteins [32]. The fact that some proteins with an N-terminal Ala are not acetylated *in vitro* and *in vivo* was taken as a first indication that N-terminal acetylation in *Sulfolobus* might occur posttranslationally and the degree of folding of the N-terminus determines whether a protein is a substrate for ssArd1.

The exceptional acetylation of the N-terminus Met-Gln was not only found for the alpha subunit of the proteasome in *H. salinarum* and *N. pharaonis* (see above), but also in *Haloferax volcanii* [33]. In the latter species the penultimate, Gln was mutated to several other amino acids, and the consequences for methionine removal and acetylation *in vivo* were characterized [34]. As expected, the introduction of an Ala at the penultimate position resulted in total methionine removal and Ala acetylation, indicating that the protein had been switched from a specific into the default haloarchaeal acetylation pathway. Unexpectedly the N-terminal methionine was not removed in the two mutants Q2S and Q2V, in contrast to the usual substrate specificity of the haloarchaeal MAP.

Both mutants (Q2S, Q2V) were acetylated at the methionine, indicating that the presumed specific acetylase can tolerate Ser and Val at the penultimate position. Three other mutations (Q2D, Q2P, Q2T) resulted in protein mixtures comprised of methionine cleaved and uncleaved, acetylated and unacetylated forms, indicating that these variants are partial substrates for MAP and specific/default Nat [34]. Clearly, further research is needed to identify the presumed specific Nat and unravel its molecular mechanism. While it does not belong to the topic of this paper, it is interesting to note that the different single amino acid mutants affected the phenotype of the cells, for example, osmotolerance, thermotolerance, or growth rate, underscoring the importance of the proteasome for the physiology of haloarchaea [34].

Also the yeast proteasome, which consists of seven different alpha and seven different beta subunits, is target for N-terminal acetylation. In contrast to haloarchaea, no specific Nat is required, but acetylation is performed by the defaults Nats. However, the situation is rather complex: NatA, NatB, and NatC are all required and responsible for the N-terminal acetylation of a specific subset of subunits [35].

4. Internal Protein Acetylation in Eukaryotes and Bacteria

In eukaryotes many proteins are differentially acetylated at the epsilon-amino group of internal lysines. Only very little is known about internal protein acetylation in archaea (see below), therefore the overview about internal acetylation in eukaryotes will be kept rather short. By far the most studied eukaryotic target proteins for internal acetylation are the histones, which in addition to acetylation can be posttranslationally modified by phosphorylation, methylation, ubiquitination, and ADP ribosylation [36]. Acetylation shields the charge of the lysine amino group and therefore decreases the binding affinity to DNA; therefore differential acetylation is a means for the regulation of gene expression via differential DNA compaction. Several families of histone acetyl transferases (HATs) and histone deacetylases (HDACs) exist in eukaryotes (e.g., [37, 38]). The names HAT and HDAC are used even when the enzymes have also or even only other targets than histones. Special interest has been given to the Sir2 subfamily of protein deacetylases (also called Sirtuins), which are highly conserved and occur not only in eukaryotes, but also in bacteria and archaea. They are NAD-dependent deacetylases and have been shown to be involved not only in a variety of gene regulatory pathways, but also metabolism, cell motility, multicellular development in social amoeba, longevity in response to caloric restriction, and different kinds of cancers [39–42]. Therefore especially HDACs have been tested as possible targets for anticancer treatments, and HDAC inhibitors have in fact entered Phase I clinical trials [43]. Today many different proteins in addition to histones are known to be regulated by differential acetylation [17, 44] including the cytoskeleton protein tubulin [45, 46]. Acetylation can have an influence on transcriptional regulation, pre-mRNA splicing, protein stability, protein interactions, cell cycle, circadian rhythm, and others [17].

Until very recently it was believed that in bacteria internal acetylation hardly occurs at all. Only two proteins were known to be acetylated at an internal lysine, that is, the chemotaxis protein CheY and the acetyl-CoA synthase. A protein acetyl transferase (PAT) and the deacetylase CobB (belonging to the Sir2 family) have been identified, which regulate the acetylation level of both proteins [47–51]. However, an affinity enrichment of acetylated peptides from *E. coli* led to the identification of 125 acetylation sites in 85 proteins, indicating that internal acetylation is much more common in bacteria than previously thought [52]. The proteins belong to a variety of functional classes, for example, protein synthesis, carbohydrate metabolism, nucleotide metabolism, and TCA cycle. 83 of the 125 acetylation sites were exclusively modified during stationary phase

and deacetylated when stationary phase *E. coli* cells were inoculated into fresh medium. Therefore, it has been argued that the main biological function of internal acetylation in *E. coli* might be the downregulation of protein activities during phases of starvation [52]

5. Differential Internal Protein Acetylation Is Essential at Least for Haloarchaea

One genetic approach has been performed to elucidate the importance of internal protein acetylation for the haloarchaeal species *H. volcanii* [53]. The genome of *H. volcanii* was found to contain three genes for protein acetylases and two genes for protein deacetylases. All three acetylases belong to the Gcn5 family of acetylases and have been named Pat1, Pat2 (Hvo_1756 and Hvo_1821), and Elp3 (Hvo_2888). One of the deacetylases (Hvo_2194) belongs to the Sir2 subfamily, while the second (Hvo_0522) belongs to the HdaI family. It was attempted to construct single deletion mutants of all five genes. Four of the five mutants could be generated and grew indistinguishably from the wild type. However it turned out to be impossible to generate the *hdaI* deletion mutant, indicating that the deacetylase HdaI is essential for *H. volcanii*. This was experimentally proven by the ability to delete the chromosomal *hdaI* gene in a strain that carried a copy of the gene on a plasmid. As the *hdaI* gene overlaps and is cotranscribed with the gene encoding the histone, it is reasonable to assume that the histone is one substrate for HdaI. It remains to be clarified whether differential histone acetylation is essential for *H. volcanii* or whether the acetylation of other target proteins of HdaI is responsible for the phenotype. While single mutants of all three acetylase genes could be obtained, the *pat2 elp3* double mutant could not be generated. Therefore, the two genes are synthetically lethal indicating that they have overlapping substrate specificities and that the acetylation of at least one protein is essential for *H. volcanii*. The fact that the ability for reversible internal protein acetylation is essential for *H. volcanii* underscores the importance for this posttranslational modification at least for this archaeal species. As genetic techniques for other archaeal species have been developed [54, 55], it will be interesting to clarify whether this is also true for additional species and can be generalized to many or all Archaea.

6. Internal Protein Acetylation in the Archaea: ALBA and a Little Bit More

The first archaeal protein with an internally acetylated lysine was reported as early as 1978; it was a 2Fe-2S ferredoxin from *H. salinarum* (at that time named *H. halobium*), which was monoacetylated on lysine 118 near the C-terminus [56]. Shortly thereafter it was reported that the homologous 2Fe-2S ferredoxin from *H. marismortui* (at that time named "*Halobacterium* from the Dead Sea") was acetylated at the equivalent position [57]. However, these observations did not trigger subsequent interest in differential protein acetylation in archaea and are today known by only few researchers.

Obvious candidates as targets for internal acetylation are the archaeal histones. It has long been argued that archaeal histones lack the N-terminal domain of eukaryotic histones, which are heavily posttranslationally modified and can therefore not be acetylated. However, meanwhile it has been found that also positions in the conserved histone core domain of eukaryotic histones are acetylated [58]. However, a proteomic approach that specifically addressed the question of histone acetylation in *Methanococcus jannaschii* and *Methanosarcina acetivorans* came to the conclusion that histone acetylation does not occur in either of the two species [59]. Cotranscription of the histone gene with the gene of a deacetylase in *H. volanii* indicates that this might be different in haloarchaea and candidate lysines, which are acetylated in eukaryotes, are conserved in haloarchaea [53], but experimental proof is still missing.

In the meantime another chromatin protein became the second known target for internal protein acetylation in archaea. It is an *S. solfataricus* protein that had originally been named Sso10b and was renamed "Alba" (acetylation lowers binding affinity) [60]. In this case the observation triggered an intensive characterization of differential acetylation and the responsible enzymes. It was shown that Alba carries two acetyl groups; on the one hand it is N-terminally acetylated, and on the other hand the epsilon-amino group of lysine 16 is acetylated [60]. The *Sulfolobus* member of the Sir2 (Sirtuin) protein family can deacetylate Alba in a NAD-dependent manner. The nonacetylated Alba could repress transcription in an *in vitro* transcription assay, in contrast to the acetylated protein, and the different activities of the two forms were verified after deacetylation of the native protein *in vitro* with Sir2 [60]. The sequence of the *Salmonella* protein acetyl transferase (PAT) was used to identify the *Sulfolobus* homologue, and it was shown that *Sulfolobus* PAT can acetylate Alba *in vitro* with a very high efficiency at the native target amino acid, lysine 16 [61]. The acetylation was shown to reduce the affinity of Alba for double-stranded DNA as well as RNA by a factor of two [61]. The rather moderate influence of the acetylation state of Alba on DNA binding has led to the proposal that it therefore seems probable that the specific acetylation of lysine 16 functions as a modular signal to other proteins rather than as a direct modulation of DNA binding affinity [62]. One interaction partner of Alba is the DNA helicase MCM. It has been shown that Alba strongly inhibits the activity of MCM in an *in vitro* helicase assay. Acetylation of Alba reduced this antagonistic activity of Alba, notably at concentrations at which acetylated Alba bound DNA, excluding an indirect effect [63].

While the molecular details of the biological consequences of the different activities of Alba remain to be clarified, it seems to be clear that the acetylation state of Alba influences the degree of "chromatin packaging" in *Sulfolobus*, analogous to the differential acetylation of eukaryotic histones. Structures of Alba from several species are available, and its binding to DNA has been modeled [64–66]. The structure of the Alba acetylase PAT has also been reported [67], and further structures are on their

way [68, 69]. In addition, the structure of the deacetylase Sir2 has been determined both in complex with NAD and with an artificial substrate, an acetylated peptide derived from the human protein p53 [70, 71]. Therefore, the archaeal chromatin protein Alba together with its cognate acetylase and deacetylase is by far the best characterized example for the relevance and mechanism of internal acetylation in archaea, and it has the potential to enable the understanding of the molecular mechanism in the near future.

However, it has been proposed that Alba is not the only acetylated protein in *Sulfolobus* and might not even be the major substrate for PAT and Sir2, because the binding affinity between PAT and Alba is quite low compared to cognate acetylase/protein target pairs in eukaryotes and because PAT is encoded in the genome of species that do not contain Alba [61]. Another argument for the presence of more internally acetylated proteins than currently known is that many archaea encode not only a single but several different protein acetylases and deacetylases. In *H. volcanii,* the protein levels of the protein acetylase Pat1 and the deacetylase Sir2 (both nonessential) have been quantified using specific antisera. It was revealed that Pat1 is constitutively present in the cells while Sir2 is downregulated in stationary phase (Hering and Soppa, unpublished results). This indicates that the acetylation level of proteins increases in stationary phase, reminiscent of the situation that has been described for *E. coli* (see above).

The current situation concerning internal protein acetylation in archaea resembles the situation in bacteria before the first focused large scale study aiming at the identification of acetylated peptides was reported, which increased the number of proven internally acetylated bacterial proteins from two to about ninety. It seems safe to predict that also in archaea many different proteins are differentially acetylated at the epsilon amino groups of internal lysines. Therefore, focused large scale approaches to identify differentially acetylated archaeal proteins are badly needed.

7. Novel Approaches to Investigate Protein Acetylation

During recent years several bioinformatic approaches have been developed for a genome-wide prediction of N-terminal acetylation or internal protein acetylation [72–75]. Currently it is unclear whether the rules obtained from eukaryotic proteins can also be used to predict acetylation in archaeal proteomes. For N-terminal acetylation one approach has already been modified using the *H. salinarum* and *N. pharaonis* dataset described above [18]. For internal protein acetylation, benchmarking of bioinformatic programs is another reason why large-scale studies of acetylated lysines are needed. The approach of affinity isolation of peptides with acetylated lysines and their identification by mass spectrometry [52] can and should also be established with a few archaeal species. For the experimental characterization of the influence of acetylation, it would be desirable to be able to compare acetylated and nonacetylated protein variants *in vivo* and *in vitro*. For N-terminal acetylation in *Drosophila,* the (X)PX-rule (proline at first or second position inhibits acetylation) was used to express mutated genes in cells and flies and study the functional relevance of N-terminal acetylation [14]. The authors propose that the (X)PX rule could be applied universally and could be used for equivalent approaches with many other species. For internal acetylation, lysines have often been replaced by other amino acids, but of course in this way not only the acetylation but also the functionality of the lysine is lost. To circumvent this problem, a system has been established for *E. coli* that enables the acetylation of internal lysines in recombinant proteins at any desired position [76].

8. Conclusion and Outlook

Knowledge about N-terminal protein acetylation in archaea has increased tremendously in recent years. Several large scale studies are available, and it became obvious that a considerable fraction of proteins is N-terminally acetylated in haloarchaea and in *Sulfolobus*. The situation seems to be different in methanogenic archaea for which N-terminal acetylation has not been mentioned despite the availability of proteomic studies.

Knowledge about internal acetylation in archaea is still very limited. Only two targets are known, and only one of which has been experimentally characterized (cognate HAT and HDAC, functional consequences). However, there are ample indications that the number of internally acetylated proteins is highly underestimated, including the presence of Pat and Sir2 in species devoid of Alba, the presence of several genes for HATs and HDACs in archaeal genomes, essentiality of HdaI and Pat2/Elp3 in *H. volcanii,* and differential regulation of Sir2 in the same species. The current situation concerning internal acetylation of archaeal proteins resembles the situation concerning *E. coli* proteins before 2008, when the number of known internal acetylation sites was raised from two to 125 due to one proteomic study. Similarly, the role of small noncoding RNAs (sRNAs) for archaeal physiology was unknown a few years ago and today their presence has been proven for any species that was looked at. It can be predicted that differential internal protein acetylation in archaea is another treasure that is waiting to be lifted. It will add an additional layer of complexity to the increasing regulatory network in archaea.

References

[1] B. Polevoda and F. Sherman, "The diversity of acetylated proteins," *Genome Biology*, vol. 3, no. 5, Article ID reviews0006, 2002.

[2] T. Kouzarides, "Acetylation: a regulatory modification to rival phosphorylation?" *EMBO Journal*, vol. 19, no. 6, pp. 1176–1179, 2000.

[3] B. Polevoda and F. Sherman, "N-terminal acetyltransferases and sequence requirements for N-terminal acetylation of eukaryotic proteins," *Journal of Molecular Biology*, vol. 325, no. 4, pp. 595–622, 2003.

[4] J. Eichler and M. W. W. Adams, "Posttranslational protein modification in *Archaea*," *Microbiology and Molecular Biology Reviews*, vol. 69, no. 3, pp. 393–425, 2005.

[5] S. D. Bell and S. P. Jackson, "Mechanism and regulation of transcription in archaea," *Current Opinion in Microbiology*, vol. 4, no. 2, pp. 208–213, 2001.

[6] J. Soppa, "Basal and regulated transcription in Archaea," *Advances in Applied Microbiology*, vol. 50, pp. 171–217, 2001.

[7] J. N. Reeeve, K. A. Bailey, W.-T. Li, F. Marc, K. Sandman, and D. J. Soares, "Archaeal histones: structures, stability and DNA binding," *Biochemical Society Transactions*, vol. 32, no. 2, pp. 227–230, 2004.

[8] P. Londei, "Evolution of translational initiation: new insights from the archaea," *FEMS Microbiology Reviews*, vol. 29, no. 2, pp. 185–200, 2005.

[9] E. R. Barry and S. D. Bell, "DNA replication in the archaea," *Microbiology and Molecular Biology Reviews*, vol. 70, no. 4, pp. 876–887, 2006.

[10] H. P. Driessen, W. W. de Jong, G. I. Tesser, and H. Bloemendal, "The mechanism of N-terminal acetylation of proteins," *CRC Critical Reviews in Biochemistry*, vol. 18, no. 4, pp. 281–325, 1985.

[11] W. Dormeyer, S. Mohammed, B. Van Breukelen, J. Krijgsveld, and A. J. R. Heck, "Targeted analysis of protein termini," *Journal of Proteome Research*, vol. 6, no. 12, pp. 4634–4645, 2007.

[12] P. Van Damme, J. Van Damme, H. Demol, A. Staes, J. Vandekerckhove, and K. Gevaert, "A review of COFRADIC techniques targeting protein N-terminal acetylation," *BMC Proceedings*, vol. 3, supplement 6, article S6, 2009.

[13] X. Zhang, J. Ye, and P. Højrup, "A proteomics approach to study *in vivo* protein Nα-modifications," *Journal of Proteomics*, vol. 73, no. 2, pp. 240–251, 2009.

[14] S. Goetze, E. Qeli, C. Mosimann et al., "Identification and functional characterization of N-terminally acetylated proteins in *Drosophila melanogaster*," *PLoS Biology*, vol. 7, no. 11, Article ID e1000236, 2009.

[15] T. Arnesen, P. Van Damme, B. Polevoda et al., "Proteomics analyses reveal the evolutionary conservation and divergence of N-terminal acetyltransferases from yeast and humans," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 106, no. 20, pp. 8157–8162, 2009.

[16] M. Gautschi, S. Just, A. Mun et al., "The yeast Nα-acetyltransferase NatA is quantitatively anchored to the ribosome and interacts with nascent polypeptides," *Molecular and Cellular Biology*, vol. 23, no. 20, pp. 7403–7414, 2003.

[17] S. Spange, T. Wagner, T. Heinzel, and O. H. Krämer, "Acetylation of non-histone proteins modulates cellular signalling at multiple levels," *International Journal of Biochemistry and Cell Biology*, vol. 41, no. 1, pp. 185–198, 2009.

[18] A. Martinez, J. A. Traverso, B. Valot et al., "Extent of N-terminal modifications in cytosolic proteins from eukaryotes," *Proteomics*, vol. 8, no. 14, pp. 2809–2831, 2008.

[19] C.-S. Hwang, A. Shemorry, and A. Varshavsky, "N-terminal acetylation of cellular proteins creates specific degradation signals," *Science*, vol. 327, no. 5968, pp. 973–977, 2010.

[20] K. K. Starheim, T. Arnesen, D. Gromyko, A. Ryningen, J. E. Varhaug, and J. R. Lillehaug, "Identification of the human Nα-acetyltransferase complex B (hNatB): a complex important for cell-cycle progression," *Biochemical Journal*, vol. 415, no. 2, pp. 325–331, 2008.

[21] K. K. Starheim, D. Gromyko, R. Evjenth et al., "Knockdown of human Nα-terminal acetyltransferase complex C leads to p53-dependent apoptosis and aberrant human Arl8b localization," *Molecular and Cellular Biology*, vol. 29, no. 13, pp. 3569–3581, 2009.

[22] Y. Gordiyenko, S. Deroo, M. Zhou, H. Videler, and C. V. Robinson, "Acetylation of L12 increases interactions in the *Escherichia coli* ribosomal stalk complex," *Journal of Molecular Biology*, vol. 380, no. 2, pp. 404–414, 2008.

[23] B. Roy-Chaudhuri, N. Kirthi, T. Kelley, and G. M. Culver, "Suppression of a cold-sensitive mutation in ribosomal protein S5 reveals a role for RimJ in ribosome biogenesis," *Molecular Microbiology*, vol. 68, no. 6, pp. 1547–1559, 2008.

[24] T. Hatakeyama and T. Hatakeyama, "Amino acid sequences of the ribosomal proteins HL30 and HmaL5 from the archaebacterium *Halobacterium marismortui*," *Biochimica et Biophysica Acta*, vol. 1039, no. 3, pp. 343–347, 1990.

[25] M. Kimura, E. Arndt, T. Hatakeyama, T. Hatakeyama, and J. Kimura, "Ribosomal proteins in halobacteria," *Canadian Journal of Microbiology*, vol. 35, no. 1, pp. 195–199, 1989.

[26] S. Klussmann, P. Franke, U. Bergmann, S. Kostka, and B. Wittmann-Liebold, "N-terminal modification and amino-acid sequence of the ribosomal protein HmaS7 from *Haloarcula marismortui* and homology studies to other ribosomal proteins," *Biological Chemistry Hoppe-Seyler*, vol. 374, no. 5, pp. 305–312, 1993.

[27] M. Aivaliotis, K. Gevaert, M. Falb et al., "Large-scale identification of N-terminal peptides in the halophilic archaea *Halobacterium salinarum* and *Natronomonas pharaonis*," *Journal of Proteome Research*, vol. 6, no. 6, pp. 2195–2204, 2007.

[28] M. Falb, M. Aivaliotis, C. Garcia-Rizo et al., "Archaeal N-terminal protein maturation commonly involves N-terminal acetylation: a large-scale proteomics survey," *Journal of Molecular Biology*, vol. 362, no. 5, pp. 915–924, 2006.

[29] P. A. Kirkland, M. A. Humbard, C. J. Daniels, and J. A. Maupin-Furlow, "Shotgun proteomics of the haloarchaeon *Haloferax volcanii*," *Journal of Proteome Research*, vol. 7, no. 11, pp. 5033–5039, 2008.

[30] A. J. Forbes, S. M. Patrie, G. K. Taylor, Y.-B. Kim, L. Jiang, and N. L. Kelleher, "Targeted analysis and discovery of posttranslational modifications in proteins from methanogenic archaea by top-down MS," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 101, no. 9, pp. 2678–2683, 2004.

[31] W. Zhu, C. I. Reich, G. J. Olsen, C. S. Giometti, and J. R. Yates III, "Shotgun proteomics of *Methanococcus jannaschii* and insights into methanogenesis," *Journal of Proteome Research*, vol. 3, no. 3, pp. 538–548, 2004.

[32] D. T. Mackay, C. H. Botting, G. L. Taylor, and M. F. White, "An acetylase with relaxed specificity catalyses protein N-terminal acetylation in *Sulfolobus solfataricus*," *Molecular Microbiology*, vol. 64, no. 6, pp. 1540–1548, 2007.

[33] M. A. Humbard, S. M. Stevens Jr., and J. A. Maupin-Furlow, "Posttranslational modication of the 20S proteasomal proteins of the archaeon *Haloferax volcanii*," *Journal of Bacteriology*, vol. 188, no. 21, pp. 7521–7530, 2006.

[34] M. A. Humbard, G. Zhou, and J. A. Maupin-Furlow, "The N-terminal penultimate residue of 20S proteasome α1 influences its Nα acetylation and protein levels as well as growth rate and stress responses of *Haloferax volcanii*," *Journal of Bacteriology*, vol. 191, no. 12, pp. 3794–3803, 2009.

[35] Y. Kimura, M. Takaoka, S. Tanaka et al., "Nα-acetylation and proteolytic activity of the yeast 20 S proteasome," *Journal of Biological Chemistry*, vol. 275, no. 7, pp. 4635–4639, 2000.

[36] E. Bártová, J. Krejci, A. Harničarová, G. Galiová, and S. Kozubek, "Histone modifications and nuclear architecture: a review," *Journal of Histochemistry and Cytochemistry*, vol. 56, no. 8, pp. 711–721, 2008.

[37] K. K. Lee and J. L. Workman, "Histone acetyltransferase complexes: one size doesn't fit all," *Nature Reviews Molecular Cell Biology*, vol. 8, no. 4, pp. 284–295, 2007.

[38] C. Hildmann, D. Riester, and A. Schwienhorst, "Histone deacetylases—an important class of cellular regulators with a variety of functions," *Applied Microbiology and Biotechnology*, vol. 75, no. 3, pp. 487–497, 2007.

[39] R. A. Frye, "Phylogenetic classification of prokaryotic and eukaryotic Sir2-like proteins," *Biochemical and Biophysical Research Communications*, vol. 273, no. 2, pp. 793–798, 2000.

[40] R. Marmorstein, "Structure and chemistry of the Sir2 family of NAD^+-dependent histone/protein deactylases," *Biochemical Society Transactions*, vol. 32, no. 6, pp. 904–909, 2004.

[41] G. Blander and L. Guarente, "The Sir2 family of protein deacetylases," *Annual Review of Biochemistry*, vol. 73, pp. 417–435, 2004.

[42] R. Sawarkar, S. S. Visweswariah, W. Nellen, and V. Nanjundiah, "Histone deacetylases regulate multicellular development in the social amoeba *Dictyostelium discoideum*," *Journal of Molecular Biology*, vol. 391, no. 5, pp. 833–848, 2009.

[43] H. Hess-Stumpp, "Histone deacetylase inhibitors and cancer: from cell biology to the clinic," *European Journal of Cell Biology*, vol. 84, no. 2-3, pp. 109–121, 2005.

[44] X.-J. Yang and E. Seto, "Lysine acetylation: codified crosstalk with other posttranslational modifications," *Molecular Cell*, vol. 31, no. 4, pp. 449–461, 2008.

[45] B. J. North, B. L. Marshall, M. T. Borra, J. M. Denu, and E. Verdin, "The human Sir2 ortholog, SIRT2, is an NAD^+-dependent tubulin deacetylase," *Molecular Cell*, vol. 11, no. 2, pp. 437–444, 2003.

[46] Y. Zhang, N. Li, C. Caron et al., "HDAC-6 interacts with and deacetylates tubulin and microtubules *in vivo*," *EMBO Journal*, vol. 22, no. 5, pp. 1168–1179, 2003.

[47] R. Barak, K. Prasad, A. Shainskaya, A. J. Wolfe, and M. Eisenbach, "Acetylation of the chemotaxis response regulator CheY by acetyl-CoA synthetase purified from *Escherichia coli*," *Journal of Molecular Biology*, vol. 342, no. 2, pp. 383–401, 2004.

[48] V. J. Starai and J. C. Escalante-Semerena, "Identification of the protein acetyltransferase (Pat) enzyme that acetylates acetyl-CoA synthetase in *Salmonella enterica*," *Journal of Molecular Biology*, vol. 340, no. 5, pp. 1005–1012, 2004.

[49] J. G. Gardner, F. J. Grundy, T. M. Henkin, and J. C. Escalante-Semerena, "Control of acetyl-coenzyme A synthetase (AcsA) activity by acetylation/deacetylation without NAD^+ involvement in *Bacillus subtilis*," *Journal of Bacteriology*, vol. 188, no. 15, pp. 5460–5468, 2006.

[50] K. Zhao, X. Chai, and R. Marmorstein, "Structure and substrate binding properties of CobB, a Sir2 homolog protein deacetylase from *Escherichia coli*," *Journal of Molecular Biology*, vol. 337, no. 3, pp. 731–741, 2004.

[51] R. Li, J. Gu, Y.-Y. Chen et al., "CobB regulates *Escherichia coli* chemotaxis by deacetylating the response regulator CheY," *Molecular Microbiology*, vol. 76, no. 5, pp. 1162–1174, 2010.

[52] B. J. Yu, J. A. Kim, J. H. Moon, S. E. Ryu, and J.-G. Pan, "The diversity of lysine-acetylated proteins in *Escherichia coli*," *Journal of Microbiology and Biotechnology*, vol. 18, no. 9, pp. 1529–1536, 2008.

[53] N. Altman-Price and M. Mevarech, "Genetic evidence for the importance of protein acetylation and protein deacetylation in the halophilic archaeon *Haloferax volcanii*," *Journal of Bacteriology*, vol. 191, no. 5, pp. 1610–1617, 2009.

[54] T. Allers and M. Mevarech, "Archaeal genetics—the third way," *Nature Reviews Genetics*, vol. 6, no. 1, pp. 58–73, 2005.

[55] M. Rother and W. W. Metcalf, "Genetic technologies for Archaea," *Current Opinion in Microbiology*, vol. 8, no. 6, pp. 745–751, 2005.

[56] T. Hase, S. Wakabayashi, H. Matsubara, et al., "Complete amino acid sequence of *Halobacterium halobium* ferredoxin containing an N(ε)-acetyllysine residue," *Journal of Biochemistry*, vol. 83, no. 6, pp. 1657–1670, 1978.

[57] T. Hase, S. Wakabayashi, H. Matsubara, M. Mevarech, and M. M. Werber, "Amino acid sequence of 2Fe-2S ferredoxin from an extreme halophile, *Halobacterium* of the Dead Sea," *Biochimica et Biophysica Acta*, vol. 623, no. 1, pp. 139–145, 1980.

[58] L. Zhang, E. E. Eugeni, M. R. Parthun, and M. A. Freitas, "Identification of novel histone post-translational modifications by peptide mass fingerprinting," *Chromosoma*, vol. 112, no. 2, pp. 77–86, 2003.

[59] A. J. Forbes, S. M. Patrie, G. K. Taylor, Y.-B. Kim, L. Jiang, and N. L. Kelleher, "Targeted analysis and discovery of posttranslational modifications in proteins from methanogenic archaea by top-down MS," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 101, no. 9, pp. 2678–2683, 2004.

[60] S. D. Bell, C. H. Botting, B. N. Wardleworth, S. P. Jackson, and M. F. White, "The interaction of Alba, a conserved archaeal chromatin protein, with Sir2 and its regulation by acetylation," *Science*, vol. 296, no. 5565, pp. 148–151, 2002.

[61] V. L. Marsh, S. Y. Peak-Chew, and S. D. Bell, "Sir2 and the acetyltransferase, Pat, regulate the archaeal chromatin protein, Alba," *Journal of Biological Chemistry*, vol. 280, no. 22, pp. 21122–21128, 2005.

[62] C. Jelinska, M. J. Conroy, C. J. Craven et al., "Obligate heterodimerization of the archaeal Alba2 protein with Alba1 provides a mechanism for control of DNA packaging," *Structure*, vol. 13, no. 7, pp. 963–971, 2005.

[63] V. L. Marsh, A. T. McGeoch, and S. D. Bell, "Influence of chromatin and single strand binding proteins on the activity of an archaeal MCM," *Journal of Molecular Biology*, vol. 357, no. 5, pp. 1345–1350, 2006.

[64] B. N. Wardleworth, R. J. M. Russell, S. D. Bell, G. L. Taylor, and M. F. White, "Structure of Alba: an archaeal chromatin protein modulated by acetylation," *EMBO Journal*, vol. 21, no. 17, pp. 4654–4662, 2002.

[65] K. Zhao, X. Chai, and R. Marmorstein, "Structure of a Sir2 substrate, Alba, reveals a mechanism for deacetylation-induced enhancement of DNA binding," *Journal of Biological Chemistry*, vol. 278, no. 28, pp. 26071–26077, 2003.

[66] K. Hada, T. Nakashima, T. Osawa, H. Shimada, Y. Kakuta, and M. Kimura, "Crystal structure and functional analysis of an archaeal chromatin protein alba from the hyperthermophilic archaeon *Pyrococcus horikoshii* OT3," *Bioscience, Biotechnology and Biochemistry*, vol. 72, no. 3, pp. 749–758, 2008.

[67] M. M. Brent, A. Iwata, J. Carten, K. Zhao, and R. Marmorstein, "Structure and biochemical characterization of protein acetyltransferase from *Sulfolobus solfataricus*," *Journal of Biological Chemistry*, vol. 284, no. 29, pp. 19412–19419, 2009.

[68] S. Biarrotte-Sorin and C. Mayer, "Cloning, purification, crystallization and preliminary crystallographic analysis of a hypothetical acetyltransferase from *Pyrococcus furiosus*," *Acta Crystallographica Section F*, vol. 61, no. 3, pp. 269–270, 2005.

[69] C.-C. Cho, C.-W. Luo, and C.-H. Hsu, "Crystallization and preliminary X-ray diffraction analysis of PAT, an acetyltransferase from *Sulfolobus solfataricus*," *Acta Crystallographica Section F*, vol. 64, no. 11, pp. 1049–1051, 2008.

[70] J. Min, J. Landry, R. Sternglanz, and R.-M. Xu, "Crystal structure of a SIR2 homolog-NAD complex," *Cell*, vol. 105, no. 2, pp. 269–279, 2001.

[71] J. L. Avalos, I. Celic, S. Muhammad, M. S. Cosgrove, J. D. Boeke, and C. Wolberger, "Structure of a Sir2 enzyme bound to an acetylated p53 peptide," *Molecular Cell*, vol. 10, no. 3, pp. 523–535, 2002.

[72] A. Li, Y. Xue, C. Jin, M. Wang, and X. Yao, "Prediction of Nε-acetylation on internal lysines implemented in Bayesian Discriminant Method," *Biochemical and Biophysical Research Communications*, vol. 350, no. 4, pp. 818–824, 2006.

[73] Y.-D. Cai and L. Lu, "Predicting N-terminal acetylation based on feature selection method," *Biochemical and Biophysical Research Communications*, vol. 372, no. 4, pp. 862–865, 2008.

[74] T. Meinnel and C. Giglione, "Tools for analyzing and predicting N-terminal protein modifications," *Proteomics*, vol. 8, no. 4, pp. 626–649, 2008.

[75] A. Basu, K. L. Rose, J. Zhang et al., "Proteome-wide prediction of acetylation substrates," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 106, no. 33, pp. 13785–13790, 2009.

[76] H. Neumann, S. Y. Peak-Chew, and J. W. Chin, "Genetically encoding Nε-acetyllysine in recombinant proteins," *Nature Chemical Biology*, vol. 4, no. 4, pp. 232–234, 2008.

Hindawi Publishing Corporation
Archaea
Volume 2010, Article ID 106341, 6 pages
doi:10.1155/2010/106341

Research Article

Extensive Lysine Methylation in Hyperthermophilic Crenarchaea: Potential Implications for Protein Stability and Recombinant Enzymes

Catherine H. Botting,[1] Paul Talbot,[1] Sonia Paytubi,[1,2] and Malcolm F. White[1]

[1] *Centre for Biomolecular Sciences, University of St Andrews, Fife KY16 9ST, UK*
[2] *Departamento de Microbiologia, Facultat de Biologia, Universitat de Barcelona, Avenida. Diagonal 645, 08028 Barcelona, Spain*

Correspondence should be addressed to Catherine H. Botting, cb2@st-and.ac.uk and Malcolm F. White, mfw2@st-and.ac.uk

Received 20 May 2010; Accepted 1 July 2010

Academic Editor: Julie Maupin-Furlow

In eukarya and bacteria, lysine methylation is relatively rare and is catalysed by sequence-specific lysine methyltransferases that typically have only a single-protein target. Using RNA polymerase purified from the thermophilic crenarchaeum *Sulfolobus solfataricus*, we identified 21 methyllysines distributed across 9 subunits of the enzyme. The modified lysines were predominantly in α-helices and showed no conserved sequence context. A limited survey of the *Thermoproteus tenax* proteome revealed widespread modification with 52 methyllysines in 30 different proteins. These observations suggest the presence of an unusual lysine methyltransferase with relaxed specificity in the crenarchaea. Since lysine methylation is known to enhance protein thermostability, this may be an adaptation to a thermophilic lifestyle. The implications of this modification for studies and applications of recombinant crenarchaeal enzymes are discussed.

1. Introduction

Lysine methylation is found in all three domains of life. In bacteria, this posttranslational modification is restricted to a handful of ribosomal and flagellar proteins [1, 2]. In eukaryotes, lysine methylation is also restricted to a subset of proteins and catalysed by highly specific methyltransferases that can generate mono-, di-, and trimethylated lysines (reviewed in [3]). The most well-known examples are the histone proteins, where lysine methylation is carried out by sequence specific SET family methyltransferases using an S-adenosyl methionine (SAM) cofactor. These modifications result in changes in protein : protein interactions, chromatin structure and gene expression (reviewed in [4]). A limited number of other eukaryal proteins including notably the large subunit of Rubisco [5] are also subject to lysine methylation, though the function of these modifications is often not known [6]. More recently, proteome-wide studies of lysine methylation in the mouse brain [7] and *Saccharomyces cerevisiae* [8] have been added to the list of modified proteins. The latter study generated preliminary evidence for 25 monomethylated and 20 dimethylated lysines from a set of 2600 yeast proteins. The methylated proteins tended to have a higher abundance and longer half-life than average and included 11 ribosomal proteins [8].

The euryarchaeon *Methanosarcina mazei* encodes a clear SET-domain protein that has been shown to methylate a single lysine in the archaeal chromatin protein MC1, suggesting that mechanisms to modulate chromatin by posttranslational modification pre-date the divergence of the archaeal and eukaryal domains [9]. However, the distribution of this archaeal SET domain protein is limited to a few methanogens. Lysine methylation has also been noted in a handful of enzymes from the crenarchaeal Sulfolobales. Six methylated lysines were identified in glutamate dehydrogenase (GDH) from the hyperthermophile *Sulfolobus solfataricus* [10] leading the authors to speculate on a role in protein thermostability. Intriguingly, no lysine methylation was observed in the orthologous protein from the hyperthermophilic euryarchaeon *Pyrococcus furiosus* [11].

For *S. solfataricus* β-glycosidase, 5 lysines were methylated with an average of 3-4 modifications per protein [12]. The recombinant, unmodified protein purified from *Escherichia coli* showed an increase in denaturation and aggregation events compared to the native version, supporting the hypothesis that lysine methylation improves protein stability in extremophiles [12]. The extent of methylation of the *Sulfolobus* chromatin protein Sso7d was shown to increase with increasing growth temperature [13], again consistent with a potential role for methylation in thermostability.

Since the limited evidence available suggested that lysine methylation in the crenarchaea might follow a different pattern to that observed in other lineages, we decided to investigate this phenomenon further by mass spectrometry. Here we show that lysine methylation is common in the crenarchaea and is catalysed by an unknown methyltransferase that lacks sequence specificity but is likely influenced by the local structure of the protein target. This has implications for the molecular biology of the crenarchaea, in particular the physical properties of crenarchaeal proteins and their recombinant counterparts.

2. Methods

2.1. Archaeal Biomass. *S. solfataricus* P2 biomass was a gift from Neil Raven (CAMR, Porton Down, UK). *Thermoproteus tenax* biomass was a gift from Bettina Siebers (University of Duisburg-Essen, Germany).

2.2. Purification of RNA Polymerase (RNAP). RNAP was purified from *S. solfataricus* P2 biomass as described previously [14] by heparin and gel-filtration chromatography with the addition of an anion-exchange column (GE Healthcare MonoQ 5/5 column) as a final polishing step. Partly purified RNAP following gel filtration was loaded onto this column in buffer A (20 mM Tris-HCl pH 8.0, 50 mM NaCl, 1 mM DTT, 1 mM EDTA) and eluted with a linear gradient of buffer B (buffer A plus 1 M NaCl). Fractions containing pure RNAP were analysed by SDS-PAGE, pooled, and stored until required.

2.3. Mass Spectrometry. The purified RNA polymerase complex (5 μL, 10 pmoles/μL) was dialysed into 50 mM ammonium bicarbonate pH 8.0 using a membrane filter (Millipore, Billerica, MA) and trypsin (0.5 μL, 0.1 μg, Promega, Madison, WI) or clostripain (ArgC) (0.5 μL, 0.1 μg, Promega) was added. The samples were incubated at 37°C overnight.

After acidification (0.1% trifluoroacetic acid (TFA)), the peptides were separated using a Dionex UltiMate 3000 nanoLC (Dionex, Sunnyvale, CA) equipped with a PepMap100 C18 300 μm × 5 mm trap and 75 μm × 15 cm column (Dionex), using a 3.5 hr gradient of increasing acetonitrile concentration, containing 0.05% TFA (5%–35% acetonitrile in 3 hours, 35%–50% in a further 30 minutes, followed by 95% acetonitrile to clean the column). The eluent was spotted onto a MALDI target plate, along with α-cyano-4-hydroxycinnamic acid matrix solution (2 mg/mL in 70% acetonitrile, 0.1% TFA matrix solution) using a Dionex Probot spotter.

The nLC-MALDI MSMS runs were analysed using an Applied Biosystems 4800 MALDI TOF/TOF Analyser (Applied Biosystems, Foster City, CA) equipped with a Nd:YAG 355 nm laser in a plate-wide-data dependent manner. All spots were initially analysed in positive MS mode in the range 800 to 4000 m/z by averaging 1000 laser spots. The MS ions that satisfied the precursor criteria (200 ppm fraction-to-fraction precursor exclusion, S/N ratio >20) were selected for subsequent MSMS from the spot where the MS ion gave the highest counts, with up to 5 MSMS being acquired from each spot, selecting the strongest precursor ion first. MSMS spectra were acquired with a maximum of 3000 laser shots or until the accumulated spectrum reached an S/N ratio of 35 for 10 peaks. All MSMS data were acquired using 1 keV collision energy.

Peak lists were extracted from the MSMS spectra and analysed using the Mascot 2.1 search engine (Matrix Science, London, UK), against a database containing the *S. solfataricus* proteome. The data were searched with tolerances of 100 ppm for the precursor ions and 0.5 Da for the fragment ions, trypsin as the cleavage enzyme, assuming up to one missed cleavage, carbamidomethyl modification of cysteines as a fixed modification and mono and dimethylation of lysine, acetylation and methionine oxidation selected as a variable modification. The matches were then manually validated.

For ESI analysis, the acidified digested sample was separated using an UltiMate nanoLC (Dionex) equipped with a PepMap C18 trap and column, using a 210-minute gradient of increasing acetonitrile concentration, containing 0.1% formic acid (5%–35% acetonitrile in 180 min, 35%–50% in a further 30 minutes, followed by 95% acetonitrile to clean the column). The eluent was sprayed into a Q-Star XL tandem mass spectrometer (Applied Biosystems) and analysed in Information-Dependent Acquisition (IDA) mode, performing 1 second of MS followed by 3 seconds of MSMS analyses of the 2 most intense peaks seen by MS. These masses are then excluded from analysis for the next 60 seconds. MS/MS data for doubly and triply charged precursor ions were converted to centroid data, without smoothing, using the Analyst QS1.1 mascot.dll data import filter with default settings. The MS/MS data file generated was analysed as above with tolerances of 0.2 Da for the precursor and fragment ions.

Specific components of the polymerase complex were analysed from SDS-PA gel bands. The gel bands were excised and cut into 1 mm cubes. These were then subjected to in-gel digestion, using a ProGest Investigator in-gel digestion robot (Genomic Solutions, Ann Arbor, MI) using standard protocols [15]. Briefly, the gel cubes were destained by washing with acetonitrile and subjected to reduction, with dithiothreitol, and alkylation, with iodoacetamide, before digestion with trypsin or clostripain at 37°C. The peptides were extracted with 10% formic acid and analysed as described above. For the nLC-ESI MSMS analysis, shorter gradients (60 or 90 minutes) were used. For MALDI analysis, the digest solution (0.5 μL) was applied to the MALDI target

along with α-cyano-4-hydroxycinnamic acid matrix (0.5 μL, 10 mg/mL in 50 : 50 acetonitrile: 0.1% TFA) and 0.1% TFA (0.5 μL) and allowed to dry. MALDI analysis was performed using the 4800 MALDI TOF/TOF Analyser. The spot was analysed in MS mode as described above, and then the most intense peptides (up to 15) were selected for MSMS analysis and acquired as described above. The combined MS and MSMS data were analysed as described above, using GPS Explorer (Applied Biosystems) to interface with the Mascot search engine.

T. Tenax samples were digested similarly in solution or from SDS-PA gel bands and analysed by nLC-ESI MSMS on the QStar XL as described above.

3. Results

3.1. Identification of Methylated Lysines in S. solfataricus RNA Polymerase. We had previously purified archaeal RNA polymerase (RNAP), a complex enzyme consisting of 12-13 subunits, from *S. solfataricus* for transcription studies [14]. As crystal structures of RNAP are available [16, 17] and significant quantities of the protein could be purified, it represented an ideal system to investigate the extent of lysine methylation in this organism. RNAP was purified to homogeneity from 50 g of *S. solfataricus* strain P2 cells by sequential chromatography on heparin and Superdex 200 gel-filtration columns as described previously [14, 18]. A final chromatography step using a MonoQ column yielded pure RNAP as shown in Figure 1.

RNAP was digested in solution using trypsin or clostripain (ArgC), and methylated peptides were identified by either LC-MALDI or nLC-ESI MS/MS. Subunits RpoD, E′, G, F, H, L, and K were also analysed individually by excising the relevant bands from a gel following SDS-PAGE and subjecting them to in-gel digestion and analysis by MALDI MS and MSMS or nLC-ESI MSMS. The Mascot search algorithm was used to identify potential sites of methylation. Subsequent manual validation centred particularly on whether the monomethyllysine immonium related ion at 98 m/z (immonium ion—17) and similarly the dimethyllysine immonium-related ion at 112 m/z were present, as well as assessing the confidence of the assignment of the specific site of modification [19]. Lysine methylation typically prevented cleavage by trypsin, providing further evidence of modification. In total, 21 monomethylated lysine residues were identified, from 9 of the RNAP subunits (Table 1, see Supplementary Table 1 at supplementary material available online at doi:10.1155/2010/106341). In several cases, the modified lysines were also detected in unmodified form, suggesting that the degree of methylation is not 100% and that a stochastic mixture of proteins with variable levels of methylation exists *in vivo*. In addition to lysine methylation, N-terminal peptides modified by acetylation were detected for four subunits: RpoA′ (Sso0225, N-terminal seq. SEKNIK), RpoB′ (Sso3254, ASNLTI), RpoD (Sso0071, SINLLH) and RpoF (Sso0751, SSVYIV). Only one unmodified N-terminal peptide was identified, that for RpoE′ (Sso0415, MYKLIK). These observations support

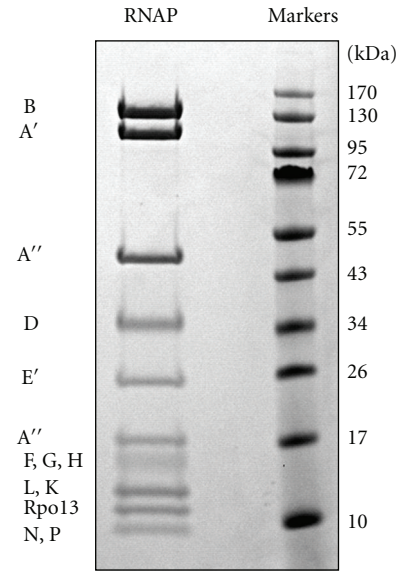


Figure 1: SDS-PAGE showing the purified RNA polymerase from *S. solfataricus*. In addition to all the expected subunits, a proteolytic product of the A″ subunit was observed at 17 kDa.

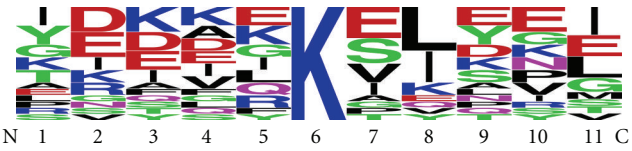


Figure 2: Sequence analysis of methylation sites in RNAP subunits. The frequency of each amino acid in an eleven-residue window centred on the modified lysine shows that there is no clear sequence specificity associated with methylation (generated by WebLogo [21]).

previous work reporting acetylation of N-terminal serine and alanine residues in *S. solfataricus*, catalysed by the acetylase ssArd1 [20].

3.2. Sequence and Structural Context of Lysine Methylation. The sequence context of methylation sites was analysed by plotting the frequency of each amino acid residue found at the 5 positions N- and C-terminal to the modified lysine (Figure 2). There was no detectable bias in amino acid residue or type (charged, hydrophobic, etc.) at any of the positions. This suggests that lysine methylation is catalysed by a methyltransferase that is largely sequence independent. To determine whether methylation occurred in a structural context, we examined the local structure of each of the 17 modified lysines visible in the RNAP crystal structures (Figure 3). All of the modified lysines were present on the surface of the protein and would be accessible to a modifying enzyme after protein folding had occurred. We noticed a striking correlation between structure and the site of modification, which was almost exclusively in a helical region of the protein (Figure 4). Exceptions were a lysine near the end of a β-turn in subunit RpoH and

Table 1: Methylation sites identified in *S. solfataricus* RNA polymerase subunits[†].

	Protein	Subunit	Residue	Peptide (methyllysine in bold)
1	Sso0225	RpoA′	395	**K**ELASTLAPGYIIER
2	Sso0225		659	**K**EIYNEIDR
3	Sso0227	RpoB′-C	12	IVE**K**TLYEMGVVPVEEVIR
4	Sso0227		311	GYKG**K**EYYR
5	Sso0227		349	FLQEF**K**ELSPEQAKR
6	Sso0071	RpoD	115	DI**K**SEDPSVVPISGDIPIVLLGTNQK
7	Sso0415	RpoE′	20	IPPNEFG**K**PLNEIALNELR
8	Sso0415		131	GIIFGE**K**SKKVIQKGDKVR
9	Sso0415		133	GIIFGEKS**K**KVIQKGDKVR
10	Sso0415		171	QPYLG**K**LEWITQTKK
11	Sso0415		179	LEWITQT**K**K
12	Sso0751	RpoF	54	CDAESAQ**K**VIEELSNIVSR
13	Sso0751		102	TYTSEDIQ**K**IIDIIR
14	Sso5468	RpoH	30	HEVLNIDEAY**K**ILK
15	Sso5468		68	**K**SQLYGEVVSYR
16	Sso5577	RpoL	71	DALL**K**AIENIR
17	Sso5577		88	GMTSHYIDEI**K**GLTK
18	Sso5865	RpoP	19	TFTDEQL**K**VLPGVR
19	Sso0396	Rpo13	66	**K**LFEDNYK
20	Sso0396		98	KAKKAVSKKVKKTKK**K**EKSVEG
21	Sso0396		100	KAKKAVSKKVKKTKKKE**K**SVEG

Monomethylated lysines are shown in bold. All lysines are conserved in the *S. shibatae* RNAP sequence with the exception of the lysine in peptide 19, which is an arginine in *S. shibatae* Rpo13.

[†]Supplementary Table 1 contains more detail and evidence supporting the identification of each modification site.

one in a hairpin turn in subunit RpoE′. Previous studies of lysines methylated in *S. solfataricus* β-glycosidase revealed two sites in α-helices and three in β-turns [12]. Together, these data suggest that lysine methylation in *S. solfataricus* is catalysed posttranslationally by a methyltransferase that is sequence independent but shows some structural specificity.

3.3. Lysine Methylation Is Prevalent in Thermoproteus tenax. Given the extensive methylation observed for RNAP subunits and β-glycosidase purified from *S. solfataricus* and the lack of any sequence dependence, lysine methylation in *S. solfataricus* is likely to be a general phenomenon involving many cellular proteins. To determine whether lysine methylation was specific to the *Sulfolobales* or more widespread, we investigated proteins from the crenarchaeon *Thermoproteus tenax*. In all, 90 proteins were represented by 3 or more tryptic peptides. Of these, 30 showed strong evidence for lysine methylation (52 sites in total, 6 of which were di-methylated; Supplementary Table 2) whilst about half the proteins showed no methylation and the remainder had weak or equivocal evidence for methylation. However, since the peptides identified did not provide complete sequence coverage, it is reasonable to assume that a proportion (and potentially a significant proportion) of these proteins do carry some lysine methylation *in vivo*. In other words, the incomplete peptide coverage leads to a potentially high false-negative rate but a low false-positive rate. These proteins had been partly purified from *T. tenax* for experiments aimed at detecting novel single-stranded DNA-binding proteins [22] and could be considered to represent a reasonably random sampling of the total soluble proteome. Although a bias towards more highly expressed proteins is likely, the methylated proteins have a wide variety of known or predicted functions. As with RNAP, there was no detectable sequence motif close to methylation sites.

The widespread methylation observed from our limited sampling of these two species together with the absence of any sequence-specific modification leads us to predict that extensive lysine methylation may be a feature of the crenarchaea. In contrast, proteomic studies in the euryarchaea *Haloferax volcanii* [23], *Thermococcus gammatolerans* [24], *Pyrococcus furiosus* [25], and *Methanococcus maripaludis* [26] have not reported any evidence of extensive lysine methylation.

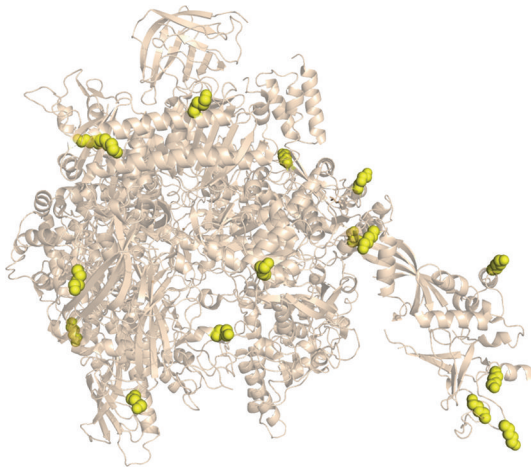


Figure 3: Mapping lysine methylation sites on the RNA polymerase crystal structure. The structure of RNAP from *S. shibatae* is shown [16] with methylated lysines listed in Table 1 indicated in yellow. This figure was generated using Pymol (The PyMOL Molecular Graphics System, Version 1.2r3pre, Schrödinger, LLC.)

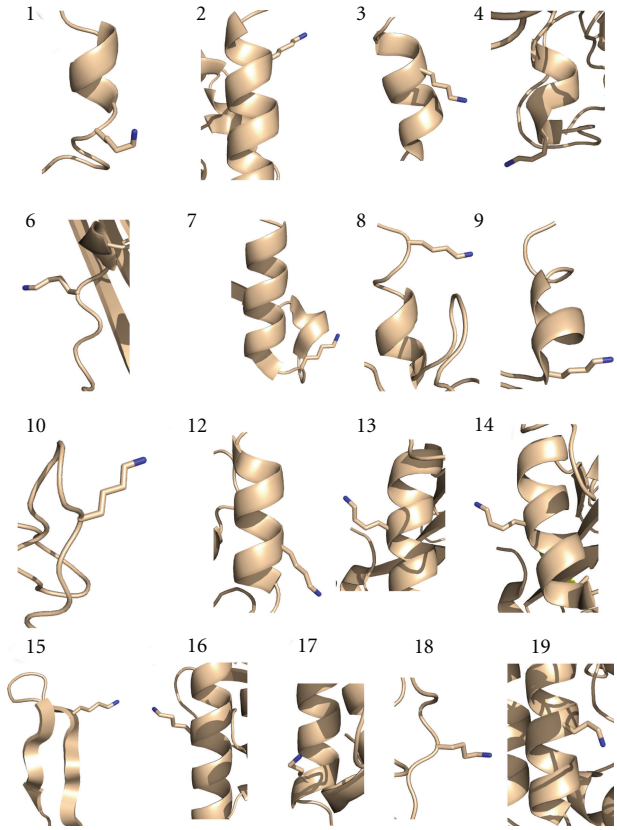


Figure 4: Structural context of methylation sites in RNAP subunits. Numbering is taken from Table 1. Fifteen of the sites are situated in α-helices whilst site 7 is in a hairpin turn and site 15 is situated in the β-sheet region of a β-hairpin. All structures are from *S. shibatae* RNAP structure (PDB 2WAQ) with the exception of 8, 9 and 10 which are from the *S. solfataricus* RNAP structure (PDB 2PMZ) as the two structures differ in this region of the E′ subunit.

4. Discussion

The data presented here expand previous work on lysine methylation in the *Sulfolobales* and point to the existence of a novel sequence-independent lysine methylase that is distinct from the canonical SET-domain methylases prevalent in eukarya. Indeed, no SET-domain protein is detectable in any crenarchaeal genome using standard protein similarity searches. This enzyme appears to target lysines in alpha helices and near hairpin turns, suggesting that modification is posttranslational and has some structural specificity. It is of course possible that lysine methylation is catalysed by multiple enzymes rather than a single one. Our limited sampling of methylated proteins in *S. solfataricus* and *T. tenax* suggest that this modification may be quite prevalent in the crenarchaea, with a significant proportion of proteins carrying one or more methyllysine residues. It will be important to extend these studies to provide a more exhaustive survey of crenarchaeal proteomes. Whilst the data currently available for euryarchaea suggests that lysine methylation is not widespread, it will be interesting to determine whether this modification is also found in the thaumarchaea [27] or is an adaptation specific to a hyperthermophilic lifestyle.

Methylation of lysine increases the pKa of the side chain, allowing stronger ionic interactions to be made, and changes the hydropathy and solubility of proteins [12]. There are many examples in the literature where lysine methylation has resulted in improved stability of a variety of proteins and peptides. For example, methylation of human amyloid peptide β(25-35) reduced aggregation and toxicity [28]. Methylation of the protease bovine trypsin has been shown to increase the thermostability and reduce autolysis, resulting in a more useful enzyme for trypsinolysis [29]. Reductive (nonenzymatic) methylation is being used extensively in crystallography to improve crystallisability of proteins, a process that does not generally affect enzyme activity [30]. Our data suggest that crenarchaeal proteins expressed in recombinant systems may have altered physical properties due to a lack of lysine methylation compared to the situation *in vivo*. Potentially, this could affect their efficacy in biotechnological and other applications. The identification of the as-yet unknown crenarchaeal lysine methyltransferase, currently underway in our laboratory, could allow specific modification of lysine residues in a wide variety of proteins to improve stability and solubility.

Acknowledgments

The authors thank the St Andrews Mass Spectrometry Facility, which is funded by the Wellcome Trust, and the expert technical help from Mr. Alex Houston and Dr. Sally Shirran.

References

[1] F. N. Chang, "A sensitive method for the quantitative determination of minor bases in ribosomal RNA," *Analytical Biochemistry*, vol. 63, no. 2, pp. 371–379, 1975.

[2] W. K. Paik and S. Kim, "Protein methylation," *Science*, vol. 174, no. 4005, pp. 114–119, 1971.

[3] W. K. Paik, D. C. Paik, and S. Kim, "Historical review: the field of protein methylation," *Trends in Biochemical Sciences*, vol. 32, no. 3, pp. 146–152, 2007.

[4] C. Martin and Y. Zhang, "The diverse functions of histone lysine methylation," *Nature Reviews Molecular Cell Biology*, vol. 6, no. 11, pp. 838–849, 2005.

[5] R. C. Trievel, E. M. Flynn, R. L. Houtz, and J. H. Hurley, "Mechanism of multiple lysine methylation by the SET domain enzyme Rubisco LSMT," *Nature Structural Biology*, vol. 10, no. 7, pp. 545–552, 2003.

[6] R. L. Houtz, R. Magnani, N. R. Nayak, and L. M. A. Dirk, "Co- and post-translational modifications in Rubisco: unanswered questions," *Journal of Experimental Botany*, vol. 59, no. 7, pp. 1635–1645, 2008.

[7] H. Iwabata, M. Yoshida, and Y. Komatsu, "Proteomic analysis of organ-specific post-translational lysine-acetylation and -methylation in mice by use of anti-acetyllysine and -methyllysine mouse monoclonal antibodies," *Proteomics*, vol. 5, no. 18, pp. 4653–4664, 2005.

[8] C. N. Pang, E. Gasteiger, and M. R. Wilkins, "Identification of arginine- and lysine-methylation in the proteome of *Saccharomyces cerevisiae* and its functional implications," *BMC Genomics*, vol. 11, article 92, 2010.

[9] K. L. Manzur and M.-M. Zhou, "An archaeal SET domain protein exhibits distinct lysine methyltransferase activity towards DNA-associated protein MC1-α," *FEBS Letters*, vol. 579, no. 17, pp. 3859–3865, 2005.

[10] B. Maras, V. Consalvi, R. Chiaraluce et al., "The protein sequence of glutamate dehydrogenase from Sulfolobus solfataricus, a thermoacidophilic archaebacterium. Is the presence of N-ε-methyllysine related to thermostability?" *European Journal of Biochemistry*, vol. 203, no. 1-2, pp. 81–87, 1992.

[11] B. Maras, S. Valiante, R. Chiaraluce et al., "The amino acid sequence of glutamate dehydrogenase from *Pyrococcus furiosus*, a hyperthermophilic archaebacterium," *Journal of Protein Chemistry*, vol. 13, no. 2, pp. 253–259, 1994.

[12] F. Febbraio, A. Andolfo, F. Tanfani et al., "Thermal stability and aggregation of Sulfolobus solfataricus β-glycosidase are dependent upon the N-ε-methylation of specific lysyl residues: critical role of in vivo post-translational modifications," *Journal of Biological Chemistry*, vol. 279, no. 11, pp. 10185–10194, 2004.

[13] H. Baumann, S. Knapp, T. Lundbäck, R. Ladenstein, and T. Härd, "Solution structure and DNA-binding properties of a thermostable protein from the archaeon Sulfolobus solfataricus," *Nature Structural Biology*, vol. 1, no. 11, pp. 808–819, 1994.

[14] S. Paytubi and M. F. White, "The crenarchaeal DNA damage-inducible transcription factor B paralogue TFB3 is a general activator of transcription," *Molecular Microbiology*, vol. 72, no. 6, pp. 1487–1499, 2009.

[15] A. Shevchenko, M. Wilm, O. Vorm, and M. Mann, "Mass spectrometric sequencing of proteins from silver-stained polyacrylamide gels," *Analytical Chemistry*, vol. 68, no. 5, pp. 850–858, 1996.

[16] Y. Korkhin, U. M. Unligil, O. Littlefield et al., "Evolution of complex RNA polymerases: the complete archaeal RNA polymerase structure," *PLoS Biology*, vol. 7, no. 5, article e102, 2009.

[17] A. Hirata, B. J. Klein, and K. S. Murakami, "The X-ray crystal structure of RNA polymerase from Archaea," *Nature*, vol. 451, no. 7180, pp. 851–854, 2008.

[18] S. Paytubi and M. F. White, "The crenarchaeal DNA damage-inducible transcription factor B paralogue TFB3 is a general activator of transcription," *Molecular Microbiology*, vol. 72, no. 6, pp. 1487–1499, 2009.

[19] T. A. Couttas, M. J. Raftery, G. Bernardini, and M. R. Wilkins, "Immonium ion scanning for the discovery of post-translational modifications and its application to histones," *Journal of Proteome Research*, vol. 7, no. 7, pp. 2632–2641, 2008.

[20] D. T. Mackay, C. H. Botting, G. L. Taylor, and M. F. White, "An acetylase with relaxed specificity catalyses protein N-terminal acetylation in Sulfolobus solfataricus," *Molecular Microbiology*, vol. 64, no. 6, pp. 1540–1548, 2007.

[21] G. E. Crooks, G. Hon, J.-M. Chandonia, and S. E. Brenner, "WebLogo: a sequence logo generator," *Genome Research*, vol. 14, no. 6, pp. 1188–1190, 2004.

[22] X. Luo, U. Schwarz-Linek, C. H. Botting, R. Hensel, B. Siebers, and M. F. White, "CC1, a novel crenarchaeal DNA binding protein," *Journal of Bacteriology*, vol. 189, no. 2, pp. 403–409, 2007.

[23] P. A. Kirkland, M. A. Humbard, C. J. Daniels, and J. A. Maupin-Furlow, "Shotgun proteomics of the haloarchaeon haloferax volcanii," *Journal of Proteome Research*, vol. 7, no. 11, pp. 5033–5039, 2008.

[24] Y. Zivanovic, J. Armengaud, A. Lagorce et al., "Genome analysis and genome-wide proteomics of *Thermococcus gammatolerans*, the most radioresistant organism known amongst the Archaea," *Genome Biology*, vol. 10, no. 6, article R70, 2009.

[25] A. M. Lee, J. R. Sevinsky, J. L. Bundy, A. M. Grunden, and J. L. Stephenson Jr., "Proteomics of *Pyrococcus furiosus*, a hyperthermophilic archaeon refractory to traditional methods," *Journal of Proteome Research*, vol. 8, no. 8, pp. 3844–3851, 2009.

[26] Q. Xia, T. Wang, E. L. Hendrickson, T. J. Lie, M. Hackett, and J. A. Leigh, "Quantitative proteomics of nutrient limitation in the hydrogenotrophic methanogen Methanococcus maripaludis," *BMC Microbiology*, vol. 9, article 149, 2009.

[27] C. Brochier-Armanet, B. Boussau, S. Gribaldo, and P. Forterre, "Mesophilic crenarchaeota: proposal for a third archaeal phylum, the Thaumarchaeota," *Nature Reviews Microbiology*, vol. 6, no. 3, pp. 245–252, 2008.

[28] E. Hughes, R. M. Burke, and A. J. Doig, "Inhibition of toxicity in the β-amyloid peptide fragment β-(25-35) using N-methylated derivatives. A general strategy to prevent amyloid formation," *Journal of Biological Chemistry*, vol. 275, no. 33, pp. 25109–25115, 2000.

[29] M. Šebela, T. Štosová, J. Havliš et al., "Thermostable trypsin conjugates for high-throughput proteomics: synthesis and performance evaluation," *Proteomics*, vol. 6, no. 10, pp. 2959–2963, 2006.

[30] Y. Kim, P. Quartey, H. Li et al., "Large-scale evaluation of protein reductive methylation for improving protein crystallization," *Nature Methods*, vol. 5, no. 10, pp. 853–854, 2008.

Hindawi Publishing Corporation
Archaea
Volume 2010, Article ID 842639, 14 pages
doi:10.1155/2010/842639

Review Article

Iron-Sulfur World in Aerobic and Hyperthermoacidophilic Archaea *Sulfolobus*

Toshio Iwasaki

Department of Biochemistry and Molecular Biology, Nippon Medical School, Sendagi, Bunkyo-ku, Tokyo 113-8602, Japan

Correspondence should be addressed to Toshio Iwasaki, tiwasaki@nms.ac.jp

Received 17 May 2010; Accepted 26 July 2010

Academic Editor: J. Soppa

The general importance of the Fe-S cluster prosthetic groups in biology is primarily attributable to specific features of iron and sulfur chemistry, and the assembly and interplay of the Fe-S cluster core with the surrounding protein is the key to in-depth understanding of the underlying mechanisms. In the aerobic and thermoacidophilic archaea, zinc-containing ferredoxin is abundant in the cytoplasm, functioning as a key electron carrier, and many Fe-S enzymes are produced to participate in the central metabolic and energetic pathways. *De novo* formation of intracellular Fe-S clusters does not occur spontaneously but most likely requires the operation of a SufBCD complex of the SUF machinery, which is the only Fe-S cluster biosynthesis system conserved in these archaea. In this paper, a brief introduction to the buildup and maintenance of the intracellular Fe-S world in aerobic and hyperthermoacidophilic crenarchaeotes, mainly *Sulfolobus*, is given in the biochemical, genetic, and evolutionary context.

1. Introduction

The structure of a metal site in metalloenzymes critically influences the fine-tuning of redox and/or catalytic activities in biology [1–3], and the substitution and/or displacement events at the local metal-binding site(s) in a protein might have greatly enhanced their capabilities of conducting a wide range of unique redox chemistry in biological electron transfer conduits which often use a limited number of basic protein scaffolds. Iron-sulfur (Fe-S) cluster prosthetic groups, consisting of nonheme iron and acid-labile inorganic sulfide atoms, are functionally highly versatile and may be among the most ancient *modular* metallocofactors to sustain biologically and evolutionary indispensable processes in the early days of primitive life on earth, such as electron transfer, substrate binding/activation in the iron/sulfur storage, hydrogen and nitrogen metabolisms, anaerobic respiration, and photosynthesis—some of the most complicated reactions in the chemistry of life processes [1, 2, 4, 5]. Among protein-bound Fe-S redox sites, which usually contain terminal sulfur ligands from cysteinyl groups, the mononuclear Fe atom in a tetrahedral environment of S ligands is the simplest form, as seen in the rubredoxin family. Other major forms are polynuclear clusters, such as those containing [2Fe-2S], [3Fe-4S], [4Fe-4S], or [8Fe-7S] core units, found in a variety of ferredoxins and complex Fe-S metalloenzymes. In addition to their electron transfer roles, Fe-S proteins are also known to participate in substrate binding/activation, environmental sensing and gene regulation [2, 5–7], and more recently are suggested to be potentially involved in several human diseases [8, 9]. The physiological importance of the Fe-S clusters is largely attributable to specific features of iron and sulfur chemistry, and the biogenesis and interplay of the Fe-S cluster core with the surrounding protein is the key to in-depth understanding of the underlying mechanisms at atomic resolution.

The archaeal domain contains organisms having the most extraordinary optimal growth conditions, with members flourishing at the extremes of pH, temperature, and salinity. The majority of thermophilic archaea are anaerobic organisms, because oxygen is often scarce in their habitats [10–13]. Biochemical and genetic evidence indicates that one of the characteristic features in the central metabolic pathways of *both* anaerobic and more unusual aerobic archaea is the involvement of small modular Fe-S proteins called ferredoxins in electron transport, and many Fe-S

enzymes are produced as well in the cells [4, 14–17]. The cytoplasm of chemoheterotrophically-grown aerobic and thermoacidophilic archaea such as *Sulfolobus tokodaii* and *Thermoplasma acidophilum* is in fact remarkably enriched with bacterial-type ferredoxin(s), containing at least ~10 times more than some aerobic and thermophilic "fast-clock" bacteria such as *Thermus thermophilus* HB8 (unpublished results).

The variation of a common theme in the ferredoxin-dependent pathways of anaerobic and aerobic archaea is striking, especially considering the early days of living organisms on earth, during which the atmosphere became progressively oxidative due to the emergence of photosynthesis by cyanobacterial ancestors that leads to the prevalence of environmental iron in the trivalent state. Under these conditions *microaerobic* archaeal ancestors had to adapt to the circumstances where, in some cases, the concentration of soluble iron compounds is *below* their physiological demands. Diminishing iron levels posed a serious challenge for early *aerobic* archaea. Additionally, the polynuclear Fe-S cluster prosthetic groups contain "acid-labile" sulfide bridges, which are inherently unstable at very acidic conditions [2, 5, 18]. The stunning capability of some contemporary *aerobic and thermoacidophilic* archaea to grow at extremely low pH with intact Fe-S clusters [19–22] has implicit meaning in that the intracellular Fe-S world must be protected not only by scavenging reactive oxygen species but also by balancing their intracellular pH at an acceptable value in the face of a huge proton gradient. This short review provides a brief introduction to the buildup and maintenance of the intracellular Fe-S world in *aerobic and thermoacidophilic* archaea, mainly *Sulfolobus*. The properties of Fe-S proteins from *anaerobic and hyperthermophilic* archaea have been extensively reviewed elsewhere by others [4, 23–26].

2. Zinc-Containing Ferredoxins Are Abundant in the Aerobic and Thermoacidophilic Archaeal Cells

The physiological significance of bacterial-type ferredoxins in the aerobic and thermoacidophilic archaea, such as *Sulfolobus* and *Thermoplasma*, was first recognized by Kerscher et al. [15], who demonstrated that ferredoxins are abundant in the cytoplasm and serve as an effective electron acceptor of a coenzyme A-acylating 2-oxoacid:ferredoxin oxidoreductase. It is a key Fe-S enzyme of the oxidative tricarboxylic acid cycle and of coenzyme A-dependent pyruvate oxidation in aerobic archaea [15, 16, 27–30]. This oxidation takes the place in a NAD^+-dependent 2-oxoacid dehydrogenase multienzyme complex in most aerobic and mesophilic bacteria and eukarya [15, 30]. Many 2-oxoacid:ferredoxin oxidoreductase paralogs have been identified in hyperthermophilic archaea and bacteria, some of which participate in the ferredoxin-dependent peptide fermentation pathways [24, 31]. It remains to be established how enzymatically reduced ferredoxin is reoxidized in aerobic and thermoacidophilic archaea.

The X-ray crystal structure of the A_2-type pyruvate:ferredoxin oxidoreductase from *Desulfovibrio africanus* has been determined by Chabrière et al. [32, 33] and shown to contain one thiamine pyrophosphate, one Mg^{2+}, and three [4Fe-4S] clusters as prosthetic groups per protomer. The ab-/a_2b_2-type homologs from aerobic archaea inherently lack the Fe-S subunit/domain called δ, which harbors two [4Fe-4S] clusters [30, 34], presumably as an evolutionary consequence of one protein adaptation strategy occurring under permanently oxidative conditions. Likewise, the superfamily of archaeal and bacterial 2-oxoacid:ferredoxin oxidoreductases have different molecular sizes and subunit compositions and exhibit highly mosaic structures with respect to their domain/subunit arrangements. This implies that they might have evolved through multiple gene duplication, fusion, and reorganization events of primordial smaller fragments [17, 30, 31]. Notably, many other Fe-S enzyme complexes in biology seem to have evolved by modular evolution in an analogous way [35–38], through which the representative superfamily has become functionally divergent to meet the physiological demands.

Major ferredoxins from chemoheterotrophically-grown aerobic and thermoacidophilic archaea such as *Sulfolobus* and *Thermoplasma*, which serve as electron acceptors of 2-oxoacid:ferredoxin oxidoreductases, are characterized by relatively higher molecular masses for bacterial-type ferredoxins (~12–16 kDa) because of a long N-terminal extension region and central loop attached to the ferredoxin core fold [15, 16, 39–44]. Unlike small [4Fe-4S] ferredoxins from some anaerobic and hyperthermophilic archaea such as *Pyrococcus furiosus* [4, 25, 45] and *Thermococcus profundus* [46], they harbor one each of low-potential $[3Fe\text{-}4S]^{1+,0}$ and $[4Fe\text{-}4S]^{2+,1+}$ clusters. The most unusual feature of these ferredoxins is the presence of an isolated zinc center [17, 41, 43, 44, 47–49], and hence they are called the "zinc-containing ferredoxins" (Figure 1).

The isolated zinc site was first identified by the 2.00-Å structure of the *S. tokodaii* ferredoxin (PDB code, 1xer.pdb) in conjunction with the metal content analysis [41, 47, 48]. It is tetrahedrally coordinated with three histidine imidazole groups (contributed by His16, His19, and His34 in the N-terminal extension region, consisting of three β-strands and one α-helix) and one carboxylate group (contributed by Asp76 in the ferredoxin core fold). This zinc site is buried within the molecule (about 5 Å deep from the protein surface), in the boundary between the N-terminal extension and the cluster-binding ferredoxin core fold, connecting these together (Figure 1(a)). Subsequently, the zinc K-edge X-ray absorption spectroscopic analysis has shown the presence of an isolated and structurally conserved zinc center in *S. tokodaii* and *T. acidphilum* zinc-containing ferredoxins. This center is tetrahedrally coordinated with (most likely) three histidine imidazoles and one carboxylate, with the average Zn–N bond distance of 2.01 Å and the Zn–O bond distance in the range 1.89–1.94 Å [43]. These values are very similar to the average crystallographic Zn–N and Zn–O bond distance of 1.96 Å and 1.90 Å, respectively, in the *S. tokodaii* zinc-containing ferredoxin structure [47, 48]. The sequence

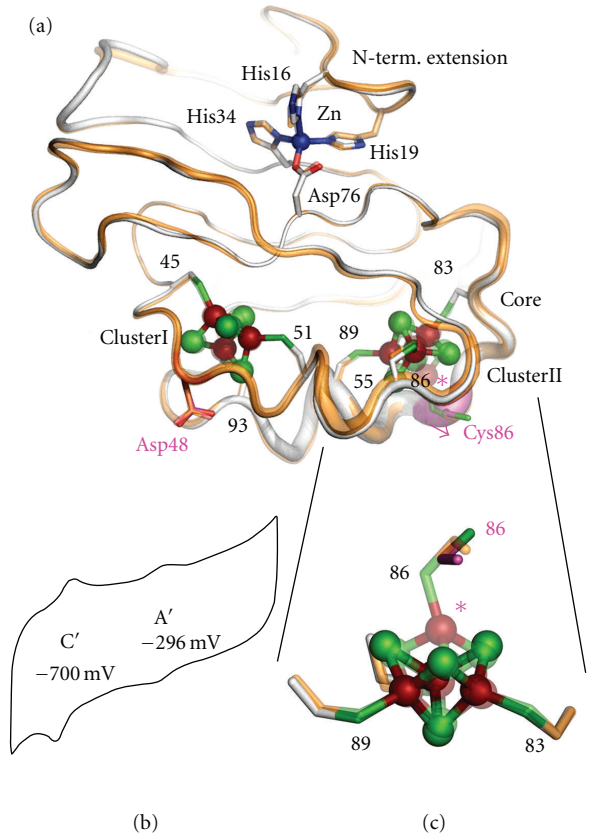


Figure 1: Comparative structures by superposition of archaeal zinc-containing ferredoxins from *S. tokodaii* (6Fe form, mostly gray, and pink for Asp48 and Cys86; 1xer.pdb [47, 48]) and *A. ambivalens* (7Fe form, transparent orange; 2vkr.pdb [49]), drawn in *B*-factor putty mode with *PyMOL* <http://pymol.sourceforge.net/> (a), and their close-up view by superposition of the cluster II site (c). In panels (a) and (c), key residues are labeled; pink asterisk indicates the special iron of the cluster II, which is missing in the 6Fe form (1xer.pdb). Typical fast-scan film voltammogram (at 400 mV·s^{-1}) of the 6Fe form of zinc-containing ferredoxin purified from *S. tokodaii* [17, 44] (measured with a pyrolytic graphite "edge" (PGE) electrode in 5 mM each of MES/PIPES/HEPES buffer, pH 7.0, containing 100 mM NaCl and 0.2 mg/ml poly-L-lysine (Sigma) as a protomer [16]; Couple A′ (for [3Fe-4S]$^{1+/0}$), $E_{1/2}$ = −296 mV (versus NHE); Couple C′ (for [3Fe-4S]$^{0/2-}$), $E_{1/2}$ = −700 mV (versus NHE); note that wave couple B for [4Fe-4S]$^{2+/1+}$ ($E_{1/2}$ = −530 mV versus NHE) [16] was undetectable in the cyclic voltammogram) [T.I. and K. Tanaka, unpublished results] (b).

comparisons suggest that three histidine residues in the N-terminal extension region and one conserved aspartate in the ferredoxin core fold are strictly conserved in all archaeal zinc-containing ferredoxins [17, 41, 43] (see Figure 1(a)), which suggests that they probably serve as ligands to the isolated zinc center. Although the isolated zinc site apparently contributes in part to ferredoxin thermal stability [50–52], zinc-lacking isoforms, for example, from *Sulfolobus metallicus* [53] and *Acidianus ambivalens* [54], have devised a natural strategy that accounts for an enhanced thermal stability without using the zinc site. It remains unknown whether another metal such as iron could replace the mononuclear zinc site of zinc-containing ferredoxin, when the archaeal cells are grown under zinc-limited conditions. Alternatively, ferredoxin(s) without zinc may functionally replace a zinc-containing ferredoxin under these conditions.

The overall protein fold of archaeal zinc-containing ferredoxins is largely asymmetric due to the presence of a long N-terminal extension and the insertion of central loop region (Figure 1(a)). Nevertheless, the ferredoxin core fold shows the strict conservation of a *pseudo*-two-fold symmetry with respect to the local two Fe-S cluster binding sites, as typically found for the bacterial-type 8Fe-containing dicluster ferredoxins harboring two [4Fe-4S]$^{2+,1+}$ clusters [17, 44]. It seems reasonable to postulate that early zinc-containing ferredoxins might have evolved as an 8Fe-containing two-electron carrier without zinc, to which the N-terminal extension and central loop regions were attached in the later stage of modular evolution. Interestingly, zinc-containing ferredoxins exhibit limited distribution in the archaeal domain (such as the Sulfolobales, Halobacteriales, and Thermoplasmatales) at the genomic sequence level, and up to now have been purified exclusively from the *aerobic and thermoacidophilic* archaea such as *Sulfolobus* and *Thermoplasma* [17]. It is emphasized that, in thermophilic euryarchaeotes, zinc-containing ferredoxin has been isolated as a major ferredoxin from the Thermoplasmatales but not Halobacteriales, an unexpected result based on the universal 16S rRNA-based phylogenetic tree [41, 43]. Analogous observation has been reported for the functionally equivalent ferredoxins of extremely halophilic and aerobic euryarchaeotes [14, 55], which contain a single plant-type [2Fe-2S] cluster and exhibit the sequence similarity to those of extremely halophilic cyanobacteria [56, 57]. It should be noted that the bacterial-type and (usually more oxygen-tolerant) plant-type ferredoxins are totally unrelated at the primary and tertiary structural levels. These observations lend credence for possible phylogenetic distribution of these archaeal ferredoxin genes driven in part by horizontal (lateral) gene transfer in the extreme environments.

Biochemical and biophysical data showed that *all* archaeal native zinc-containing ferredoxins contain one [3Fe-4S]$^{1+,0}$ cluster (cluster I) and one [4Fe-4S]$^{2+,1+}$ cluster (cluster II) [16, 39–44]. In *S. tokodaii* zinc-containing ferredoxin, cluster I (E_m = −280 mV) is selectively reduced by the cognate 2-oxoacid:ferredoxin oxidoreductase during the steady-state turnover in the presence of 2-oxoglutarate and coenzyme A, while the bulk of cluster II (E_m = −530 mV) remains in the oxidized state [16]. This suggests that the cluster I plays a crucial redox role in the physiological electron transfer. Crystal structures of *S. tokodaii* (1xer.pdb) [47, 48] and *A. ambivalens* (2vkr.pdb) [49] zinc-containing ferredoxins indicate that the [3Fe-4S] cluster I is bound to the polypeptide chain by three cysteinyl residues, Cys45, Cys51, and Cys93 (Figure 1(a)). Residue Asp48 (a potential ligand for a fourth site, if the cluster I were a [4Fe-4S] cluster) is not bound and its carboxyl $O_\delta 1$ connects to the side chain O_γ and the main chain amide NH of Ser50 by hydrogen bonds, away from the [3Fe-4S] cluster I. It should be added that [4Fe-4S] ferredoxins from anaerobic and hyperthermophilic archaea

such as *P. furiosus* [25, 45] and *T. profundus* [46] contain a conserved aspartate residue at the equivalent position, serving as a ligand to the oxygen-labile $[4Fe\text{-}4S]^{2+,1+}$ cluster. It has been reported that a one-electron reduced $[3Fe\text{-}4S]^{0}$ cluster I of the Sulfolobales zinc-containing ferredoxins undergoes a one-proton uptake reaction, and that further two-electron hyper-reduction, which also involves uptake of protons, reversibly produces a stable, hyper-reduced $[3Fe\text{-}4S]^{2-}$ species containing the formal equivalent of three ferrous ions [16, 39, 40, 42] (see Figure 1(b)).

An unexpected result from the structure of *S. tokodaii* zinc-containing ferredoxin (1xer.pdb) [47, 48] was that the cluster II is converted to a cuboidal [3Fe-4S] cluster, ligated by only three cysteinyl residues, Cys55, Cys83, and Cys89, in the lattice (Figures 1(a) and 1(c)). The side chain of Cys86, a potential ligand for a fourth site, is not bound but apparently rotated toward the solvent, away from the cluster II. Additionally, the electron density for Cys86 is much lower than those of other cysteinyl ligand residues (1xer.pdb). Given the *pseudo*-two-fold symmetry of a ferredoxin core fold of bacterial-type ferredoxins, the structure indicates that, whenever a [3Fe-4S] cluster is present (and regardless of the cluster site I or II), the missing corner (Fe) of the cube is associated with either replacement (e.g., $Cys^{II} \rightarrow Asp$, as observed for archaeal zinc-containing ferredoxins) or tilting away to the solvent of the second cysteine residue (Cys^{II}) in the -Cys^{I}-XaaXaa-Cys^{II}-XaaXaa-Cys^{III}-XaaXaaXaa-Cys^{IV}-(Pro)- motif [17, 44] (Figure 1(c)). More recently, the structure of the 2.01-Å structure of *A. ambivalens* zinc-containing ferredoxin, harboring one [3Fe-4S] cluster and one [4Fe-4S] cluster, was reported (PDB code, 1vkr.pdb) [49], confirming this concept (Figures 1(a) and 1(c)).

The presence of two [3Fe-4S] clusters is very unusual in the bacterial-type dicluster ferredoxins. Biochemical and spectroscopic analyses of *S. tokodaii* zinc-containing ferredoxin showed that the 6Fe-containing species, harboring two [3Fe-4S] clusters in the lattice (Figure 1(a)), is an artifact of the crystallization procedures at pH 5; it also represents a stable intermediate produced by mild oxidative degradation of the cluster II site that occurs very slowly in solution at pH 5 *in vitro* and does not degrade Fe-S clusters to the point of an apoprotein [44]. This raises the question of how the intact Fe-S clusters of zinc-containing ferredoxins, abundant in the cells, are maintained *in vivo*, given that the intracellular pH value of aerobic and thermoacidophilic archaea is estimated around pH 5.6–6.5 [21, 58–60]. One likely possibility is that damaged Fe-S clusters are rapidly repaired *in vivo* as in *Escherichia coli* [61], but nothing is known to date about the mechanism of the archaeal Fe-S cluster repair system.

Contemporary aerobic and anaerobic archaea apparently inherited the intracellular Fe-S world from their *anaerobic* ancestors, which can be explained by the emergence of Fe-S clusters as central catalysts of metabolism from when life had evolved in an anaerobic environment. The stunning capability of some *aerobic and thermoacidophilic* archaea to grow at extremely low pH [19–22] has therefore implicit meaning in that the intracellular Fe-S world must be protected not only by scavenging reactive oxygen species (e.g., see [62–67]) but also by balancing the intracellular pH at an acceptable value (typically 5.6–6.5 in *Sulfolobus* and *Thermoplasma* [21, 58–60]) in the face of a huge proton gradient ($\Delta pH = pH_{in} - pH_{out}$). The ΔpH across the cytoplasmic membrane of these archaea is intrinsically linked to the cellular energetics [21, 58, 68], because it is the primary contributor to the proton motive force (PMF)

$$PMF = \Delta\Psi - 2.3\left(\frac{RT}{F}\right)\Delta pH \ (mV), \tag{1}$$

where ΔΨ is the membrane potential generated by the transport of electrical charge, R the gas constant, T the absolute temperature (K), and *F* the Faraday constant (the effect of 1 unit pH difference is 2.3(RT/F), which equals 59 mV at 25°C and 70 mV at 80°C). However, the influx of protons through the archaeal A_oA_1-ATP synthase upon ATP production [22, 60, 69] intensifies cellular protonation, and therefore need to be balanced by extrusion using the cognate proton translocating systems to remove excess protons from the cytoplasm (otherwise, this would simply result in rapid acidification of the cytoplasm and dissipate any ΔpH formation across the membrane [21, 58–60, 68]). Thus, the balance between the proton permeability across the membrane (kept *very low* in thermoacidoacidophiles), the proton influx through the energetic and transport systems, and the rate of outward proton pumping determines whether an archaeal cell can sustain an appropriate PMF. The mechanistic detail of how this thin-line balance could be achieved has not been clearly understood.

A mechanism used by some acidophilic archaea such as *Thermoplasma* to reduce the proton influx is the generation of an inside positive ΔΨ [59], which is opposite to the inside negative ΔΨ in mammalian mitochondria. It is suggested that the reversed ΔΨ is generated by a difference in electrical potential (Donnan potential) formed between a greater influx of cations (such as potassium ions) and the outward flux of protons [19, 21, 59]. This is in line with our preliminary study on the aerobic respiratory chain of *T. acidophilum*, which contains at least cytochrome *bd* as a terminal oxidase (unpublished results) that is usually not a proton pump. However, this concept is apparently not applicable to *Sulfolobus*, where the inside negative ΔΨ is rather low (about −20 to −40 mV at 45°C) and the PMF is largely composed of a ΔpH of greater than 2 units [58, 60].

As reviewed elsewhere [68, 70–75], the *Sulfolobus* species have the unusual terminal oxidase segments of the aerobic respiratory chain, consisting mainly of only *a*- and *b*-type cytochromes, which most likely fulfill the role as an effective proton pump *in vivo* and preserve the cognate Fe-S world. Primary dehydrogenases, some of which are complex Fe-S enzymes, provide the reducing equivalents (from the respiratory substrates such as succinate, NADH, and sulfide) to the central caldariellaquinone pool in the tetraetherlipid membrane [35, 68, 71, 74, 75]. It should be commented here; however, that most of key biochemical/genetic characterization of the *Sulfolobus* respiratory chains (e.g., [68, 71, 72, 76–78]) were carried out before the availability of a variety of the genome-wide sequence information [74, 75, 79, 80] and the mechanistically insightful 3D structures of

the terminal segments of the tractable respiratory complexes from bacteria and eukaryal mitochondria (reviewed in [81–87]). In retrospect, many experimental data in the literature from the pregenomic era were discussed in seemingly oversimplified ways, perhaps inspired by an idea that an archaeal aerobic respiratory chain might be "primitive and simple". The archaeal respiratory chains may be in fact archaic, but not so primitive as they had seemed two decades ago. For instance, the *S. tokodaii* genomic sequence [74] shows at least seven paralogous genes coding for the putative quinol/cytochrome oxidase subunit I superfamily, two of which are homologs of SoxB [76] and SoxM [77] of *S. acidocaldarius*; of course, not all of these proteins may be spontaneously expressed to function as true respiratory terminal oxidases (some of them may be induced only under specific growth conditions [88–90] and/or serve as a putative oxygen sensor(s) for aerotaxis [91]). Additionally, while the terminal oxidase supercomplexes of *S. acidocaldarius* (SoxABCD and SoxM supercomplexes [68, 76, 77]) and *S. tokodaii* [71] (presumably SoxABCD-like supercomplex as estimated from the similarity of the redox potentials of heme A_S centers [68]) have been shown to mimic a genetic and functional fusion of mitochondrial respiratory complexes III and IV, the number of the redox centers resolved spectroscopically in the literature is insufficient to explain the proposed intramolecular electron transfer mechanism, particularly in the light of a modified Q-cycle scheme of respiratory complex III, which is characterized by bifurcation of electron transfer between two different acceptor chains that allows coupling to proton transfer [85–87, 92]. Thus, the *Sulfolobus* aerobic respiratory chain in a mechanistic context is still in its infancy compared with the mitochondrial and bacterial tractable model systems, and needs to be explored in future studies.

In the thermoacidophilic archaea, the transmembrane ΔpH-driven secondary transporters for peptides, sugars, and inorganic compounds are preferred over primary (ABC) transporter systems [19–21], which is not surprising given a permanent huge ΔpH across the membrane. Available genomic sequences of the *Sulfolobus* species [74, 75, 79, 93] suggest the presence of metal transporter homologs [20, 22, 94, 95], some of which may be involved in trafficking iron ions for the biogenesis of Fe-S proteins. Very little is known to date about *in vivo* iron-trafficking and homeostasis systems in these archaea (e.g., [88–90]), and further genome-wide "omics" approaches in a functional context may bring a better understanding of these mechanisms.

3. Formation of Intracellular Fe-S Clusters Does Not Occur Spontaneously but Requires Specific Biosynthetic Pathways

In contemporary bacteria and eukarya, the *de novo* Fe-S cluster biogenesis and maturation *in vivo* have been shown to require specific enzymes in the carefully regulated Fe-S cluster biosynthesis systems [5, 7–9, 96–101], while spontaneous assembly of the Fe-S clusters does occur *in vitro*. At least three types of the Fe-S cluster biosynthesis systems (ISC (*i*ron *s*ulfur *c*luster), SUF (mobilization of *suf* ur), and NIF (*ni*trogen *f*ixation)) are known, with significant variations in terms of the phylogenetic distribution [7, 99–101]. For example, in *Escherichia coli* the ISC pathway [102–104] is the major system for *in vivo* Fe-S cluster assembly, compared to the SUF pathway [98, 105]. In the eukaryal domain [7, 8], ISC homologs are found to be localized largely in mitochondria, while SUF homologs are found in some chloroplasts. It is therefore possible to postulate that the mitochondrial ISC system originated from the endosymbiotic bacterial ancestor and the plastid SUF system from the cyanobacterial ancestor. In these tractable model organisms, the regulation of biological Fe-S cluster assembly is further complicated by the involvement of other accessory proteins required for the *in vivo* function [7, 8, 99, 101, 106, 107], and is not fully understood.

The common concept of the three *de novo* Fe-S cluster biosynthesis systems is that *in vivo* cluster assembly requires at least (i) cysteine desulfurases (such as NifS, IscS and SufS) [105, 108–113] and (ii) Fe-S cluster scaffold proteins (such as IscU, IscA, SufU, and likely SufBCD) with the capacity to construct transient [2Fe-2S] or [4Fe-4S] clusters and then transfer Fe-S clusters to target apo-proteins [114–120] (as schematically illustrated in Figure 2). While pyridoxal phosphate-containing cysteine desulfurases utilize L-cysteine for mobilization of S for Fe-S core formation, there is as yet no consensus concerning immediate iron donor for Fe-S cluster assembly. The ISC machinery has been most closely investigated, and bacterial genomic sequence analyses showed the relatively conserved gene arrangement as *iscR-iscS-iscU-iscA-hscB-hscA-fdx* [102, 109, 121], where IscR is a transcriptional regulatory protein, HscA/HscB DnaK/J-type heat-shock proteins, and Fdx an adrenodoxin-like [2Fe-2S] ferredoxin.

The importance of the SUF machinery in the Fe-S cluster biosynthesis function was clarified in *E. coli* by construction of different combinations for altered expression of the ISC and SUF operons [98, 122, 123]. Disruption of the *E. coli suf* operon does not cause any major defects, whereas the loss of both the ISC and SUF systems leads to synthetic lethality. The components of the *suf* operon has been shown to be preferred for Fe-S cluster biosnthesis under oxidative stress conditions [124, 125] and during iron starvation [122] although the ISC and SUF systems are principally interexchangeable, especially in an anaerobic environment [123]. In *E. coli* and *Thermotoga maritima*, the *suf* gene cluster is arranged as *sufA-sufB-sufC-sufD-sufS-sufE1* and *sufC-sufB-sufD-sufS-sufU*, respectively, [98, 105] (Figure 2, bottom). In some hyperthermophilic archaea and bacteria, the SUF system has been proposed to be the sole pathway for cluster assembly [98, 126]. This implies that some components of the hyperthermophile SUF-related system might represent a primordial pathway for the Fe-S cluster biogenesis.

Although aerobic and anaerobic archaea produce numerous Fe-S proteins, the major components of the bacterial and eukaryal Fe-S cluster biosynthesis systems are not universally conserved in archaea. In *S. tokodaii* [74] (and some other archaeal species), only the *sufB*, *sufC*, and *sufD* genes

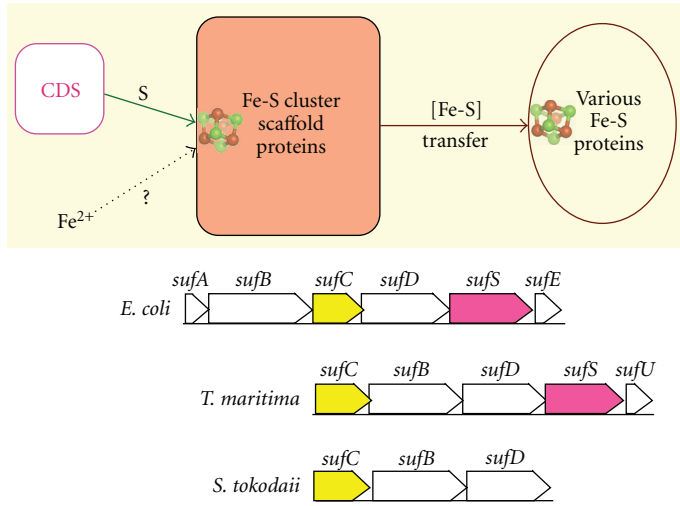


FIGURE 2: Schematic illustration of the cysteine desulfurase (CDS)-mediated, transient Fe-S cluster assembly on Fe-S cluster scaffold proteins and subsequent cluster transfer to various target apoproteins [7, 99, 116] (top), and the organization of the *suf* gene clusters annotated in the *E. coli*, *T. maritima*, and *S. tokodaii* genomic sequences (bottom).

are conserved, which are arranged as the *sufC*(ST1201)-*sufB*(ST1200)-*sufD*(ST1199) gene cluster (Figure 2, bottom). SufB and SufD are paralogs and form a water-soluble, unorthodox ATP-binding cassette-like complex with SufC that has intrinsic ATPase activity [113, 127]. No *sufA* homolog could be identified in this archaeal genomic sequence [74, 126]. This is in line with a detection of archaeal SufBCD complex by the native proteome approach from native biomass using *P. furiosus* [128]. Recently, the *E. coli* $SufBC_{(2)}D$ complex has been shown to function as a novel Fe-S scaffold machine and interacts with SufA for the Fe-S cluster transfer [119, 129], and formation of the oxygen-labile [4Fe-4S] cluster was characterized by *in vitro* reconstitution of $SufBC_2D$ under anaerobic conditions [120]. These findings strongly argue for the archaeal $SufBC_{(2)}D$ complex functioning as a possible Fe-S scaffold machine.

While SufA is absent in most archaea [126] (Figure 2, bottom), the homologs of the bacterial *apbC* [130, 131] and eukaryotic *NBP35* [107, 132] genes, coding for Fe-S cluster carrier proteins, are conserved in some archaea [133] (ST0174 in the *S. tokodaii* genomic sequence [74]). ApbC from *Salmonella enterica* is a homodimeric ATPase which can bind an Fe-S (presumably [4Fe-4S]) cluster and activate yeast apo-isopropylmalate dehydratase (apo-Leu1) *in vitro*, in an ATP-independent manner [130, 131], and the *S. enterica* strains defective in *apbC* (*mrp* in *E. coli*) showed that alterned thiamine biosynthesis are impaired in Fe-S cluster metabolism [134]. Likewise, the eukaryal ApbC homologs Cfd1 and Nbp35 form the extramitochondrial homotetrameric complex, and bind labile [4Fe-4S] clusters (after *in vitro* reconstitution), which can be transferred to target Fe-S apoproteins but only when other CIA (*c*ytosolic *i*ron-sulfur protein *a*ssembly) proteins Nar1 and Cia1 co-exist [107]. The archaeal ApbC/NBP35 homolog shows similar properties as *S. enterica* ApbC [133], and is a putative candidate for an Fe-S cluster shuttle that delivers a preassembled Fe-S cluster to a recipient apoprotein, although nothing is known to date about its interplay with the cognate SUF system.

A missing piece in the SUF system of aerobic and thermoacidophilic archaea is a cysteine desulfurase (IscS/SufS/CsdA) homolog (see Figure 2, bottom). For example, an archaeal SufS homolog was recently identified from *Haloferax volcanii* [135] and a possible CsdA (but not SufS) homolog is found in the genomic sequence of *Aeropyrum pernix* K1 (APE2023 [136]), but they are poorly conserved in *S. tokodaii* (presumably ST2140, tentatively annotated as a hypothetical isopenicillin *N* epimerase [74]). Thus, an alternative possibility is still open for novel cysteine desulfurases in these archaeal SUF systems. There are very few genetic and biochemical studies (e.g., [128, 133]) on the archaeal Fe-S cluster biosynthesis system so far, and further development of the genetic manipulation systems is needed to verify these hypotheses in a functional context.

4. Geometric Tolerance of the Cluster Binding Loop Region and the Thiophilicity with Iron Ions Respect to the Fe-S Cluster Recognition

As described briefly in the preceding section, the *de novo* Fe-S cluster biosynthesis, which is catalyzed and regulated by a number of specific enzymes, can be divided into two major steps (Figure 2). The first step is a transient *de novo* Fe-S cluster assembly on a scaffold protein requiring sulfur and iron donors. In the second step, the transient Fe-S cluster is dislocated from the scaffold protein, followed by transfer and insertion into recipient apoproteins, either during or shortly after the apoprotein generation and before the folding into its native-like conformation. A question of how the required (and rather ill-defined) binding site of a recipient protein-matrix, often categorized as the "binding

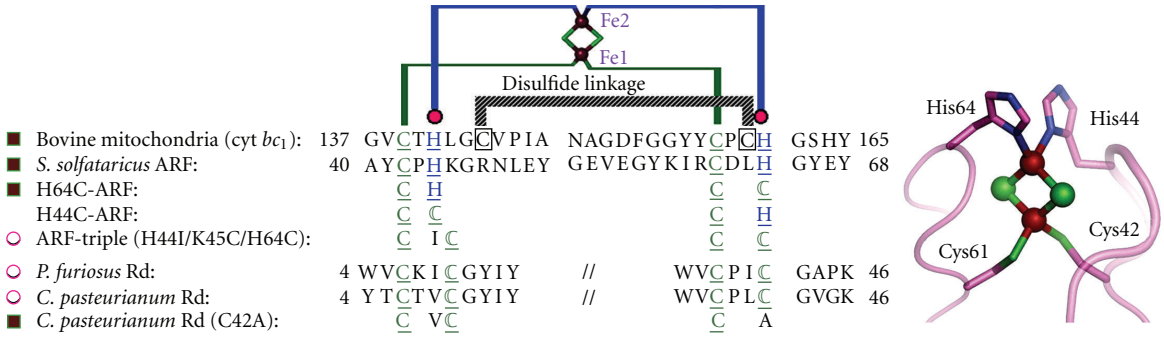


FIGURE 3: Multiple sequence alignment of the metal-binding sites of selected Rieske-type proteins and rubredoxins (Rd). The cluster-binding motif of *S. solfataricus* ARF is characteristic of Rieske-type ferredoxins involved in bacterial multicomponent oxygenases, containing two histidyl (blue) and two cysteinyl (green) ligands, and lacks two conserved cysteine residues (black) that serve as the disulfide linkage in high potential, respiratory Rieske proteins [138, 153]. DDBJ-EMBL-GenBank accession codes: bovine mitochondrial cytochrome bc_1-associated Rieske protein fragment, P13272; *S. solfataricus* ARF (hypothetical ORF c06009), CAA669492, AB047031; *P. furiosus* rubredoxin, P24297; *C. pasteurianum* rubredoxin, P00268. The metal-binding motifs are underlined (left), and the structure of the cluster ligand residues of a bovine mitochondrial Rieske protein domain fragment (PDB code, 1rie.pdb) [145] is shown, but with the *S. solfataricus* ARF numbering (right). Brown square symbols (left), wild-type or mutant proteins containing a [2Fe-2S] cluster; magenta open circles (left), wild-type or mutant proteins containing a Rd-like, mononuclear (Fe/Zn) center.

motif" in the genome-wide bioinformatics, could select and bind a specific Fe-S cluster in the Fe-S protein biogenesis is considered in this section.

Our group used an archaeal Rieske-type [2Fe-2S] ferredoxin (called ARF) from *Sulfolobus solfataricus* P1 [137–142] as a tractable model (Figure 3). Rieske-type [2Fe-2S] clusters are ubiquitous in a variety of organisms, playing crucial electron transfer functions in respiratory chains, photosynthetic chains, and multicomponent oxygenase systems for biodegradation of aromatic and alkene compounds [85, 143, 144]. In contrast to regular plant- and vertebrate-type [2Fe-2S] ferredoxins having complete cysteinyl ligations, the Rieske-type cluster has an asymmetric [2Fe-2S] core with the S_γ atom of each of the two cysteine residues coordinated to one iron site and the N_δ atom of each of the two histidine residues coordinated to the other iron site (e.g., PDB codes, 1rie, 1rfs, 1ndo, 1fqt, 1jm1, 1nyk and 2nuk.pdb [145–151]) (Figure 3, right). The structure of a bovine mitochondrial Rieske protein domain fragment suggests that its cluster-binding loops have a similar geometry to those found in the rubredoxin and zinc ribbon scaffolds [145]. We have addressed the influence of substitution of each of the two outermost histidine ligands (His44 and His64) by cysteine on the properties of the Rieske-type [2Fe-2S] cluster in *S. solfataricus* ARF (Figure 3). Replacement of one of the histidine ligands, His64, by cysteine allowed the assembly of a new low-potential [2Fe-2S] cluster with one-hisitidine plus three-cysteine ligands in the archaeal Rieske-type protein scaffold whereas replacement of the other ligand, His44, by cysteine generated a protein that failed in cluster insertion and/or assembly [138]. Replacement of the two histidine ligands to the [2Fe-2S] cluster of *S. solfataricus* ARF by cysteine residues (in the H44C/H64C double mutant) largely impaired the cluster assembly in the recombinant variant protein. In contrast, replacement of three residues (His-44, Lys-45, and His-64) in ARF by cysteines and isoleucine (H44I/K45C/H64C triple mutant), to mimic the mononuclear $Fe(Cys)_4$ site in the *P. furiosus* rubredoxin [152], has allowed a rational design of the thermostable rubredoxin-like, mononuclear $Fe(Cys)_4$ site in the recombinant ARF-triple mutant protein [153] (Figure 3, left).

These experiments demonstrate that the *in vivo* assembly of a [2Fe-2S] cluster in the Rieske protein scaffold is determined primarily by the nature and spacing of the ligands at the cluster binding loops which are often located near the protein surface in modular Fe-S proteins [138, 153] (Figure 3). The two innermost cysteinyl ligand residues (Cys42 and Cys61) of *S. solfataricus* ARF are also essential for the cluster assembly and/or stability [138], suggesting that the thiophilicity of iron ions with the thiol-containing loop region is also important for the Fe-S cluster binding and/or stability. It seems plausible that a (kinetic) "native-like" semiordered structure of the cluster binding site in a folding intermediate may behave as a substrate in the enzyme-assisted [2Fe-2S] cluster assembly/maturation steps, where (i) the geometric tolerance of the metal-binding loops, allowed by the spacing and types of ligands near the protein surface, and (ii) the thiophilicity of iron ions with the thiol-containing loops should play decisive roles [153]. This is in accord with the previous report by Meyer et al. [154], clearly showing the (unexpected) assembly of an oxidized [2Fe-2S] cluster into a recombinant, single-ligand-substituted (C42A) variant of *Clostridium pasteurianum* rubredoxin, whose polypeptide chain normally accommodates a mononuclear $Fe(Cys)_4$ site in the wild-type protein (see Figure 3, left).

Although not experimentally tested, generality of this "geometrical tolerance plus thiophilicity" concept seems to also apply to the biogenesis of a cubane [4Fe-4S] cluster, considering also the established interconversion of the Fe-S cluster types (two [2Fe-2S] ↔ one [4Fe-4S]) on the IscU scaffold protein [114, 116]. Here the minimal requirement

for the number of terminal cysteinyl ligands to a cubane [4Fe-4S] cluster is usually three in most simple and complex Fe-S proteins, and the fourth ligand at a (spatially) particular position can be an external ligand [2] (e.g., see Figure 1(c)). This may be the reason why a cubane [4Fe-4S] core is often employed for the substrate binding/activation in some Fe-S enzymes, such as aconitase and related hydratases, and the radical *S*-adenosylmethionine (SAM) superfamily [5, 6, 155, 156].

A likely biological and evolutionary benefit of having a polynuclear cluster site in a complex metalloenzyme would be that the cluster synthesis/assembly can be more strictly controlled by one or more specific synthesis-and-assembly apparatuses [5, 96–98], thereby facilitating a unique redox chemistry for specific cellular needs—simple binding of a mononuclear transient metal site in a primordial metalloprotein might have been more severely influenced by the *in vivo* availability of environmental metal ions to the last universal common ancestors (due to the simpler metal binding equilibrium). Additionally, a cavity of sufficiently large size to accommodate a polynuclear cluster might reduce a potential problem of binding the wrong metal ion that is correlated with the Irving-Williams series [157] of the stability trend for aqueous metal-sulfur complexes in the order, $Mn^{2+} < Fe^{2+} < Co^{2+} < Ni^{2+} < Cu^{2+} > Zn^{2+}$ (even when diminishing iron levels posed a serious challenge for early aerobic archaea). Prototypal polynuclear cluster formations, followed by early modular evolutionary events afforded "stepwise" development of new catalytic and electron transfer functions of primordial complex metalloenzymes. These enzymes consist of ensembles of redox protein modules of convergent/divergent evolutionary origins, using a limited number of basic protein scaffolds, and could meet versatile requirements of early metabolisms and environmental conditions [153]. Contemporary aerobic and thermoacidophilic archaea inherited the resultant intracellular Fe-S world from their anaerobic ancestors, and this world keeps running in an extraordinary environment by powering the enzyme-assisted Fe-S cluster biogenesis machinery.

5. Conclusion

The majority of thermophilic archaea are anaerobic organisms because molecular oxygen is often scarce in their habitats. Early biochemical evidence has established that one of the characteristic features in the central metabolic pathways of *both* anaerobic and aerobic archaea is the involvement of ferredoxins in electron transport. In the aerobic and thermoacidophilic archaea, zinc-containing ferredoxin [17] is abundant in the cytoplasm and functions as a key electron carrier; in addition, many other Fe-S enzymes are operative in the central metabolic and bioenergetic pathways [17, 35, 68]. These Fe-S proteins must be protected by keeping intracellular pH at an acceptable value (typically 5.6–6.5 in *Sulfolobus* and *Thermoplasma* [20, 21, 58–60]) in the face of a huge proton gradient ΔpH across the membrane. Thus, in addition to expected structural adaptations of a local Fe-S cluster binding site by natural selection, the Fe-S enzymes of aerobic and thermoacidophilic archaea obligately require the stringent intracellular pH homeostasis mechanism, as well as the reactive oxygen species-scavenging system. Some thermoacidophilic archaea such as *Thermoplasma* do this by reducing the proton influx by the generation of an inside positive membrane potential ΔΨ, which is generated by a difference in electrical potential formed between a greater influx of cations (such as potassium ions) and the outward flux of protons [19, 21, 59]. In *Sulfolobus*, the inside negative ΔΨ is rather low and the PMFis largely composed of a ΔpH of greater than 2 units [21, 58, 60, 68], where the cognate aerobic respiratory chain probably fulfills the role as an effective proton pump *in vivo* and preserves the cognate Fe-S world descendant from their anaerobic ancestors.

De novo formation of intracellular Fe-S clusters does not occur spontaneously but requires specific biosynthetic pathways: of three types of the Fe-S cluster biosynthesis systems (NIF, ISC, and SUF) identified in the bacterial and eukaryal systems [7, 98–101], the thermoacidophilic archaea apparently contain only the SUF system. More specifically, only the SufB, SufC, and SufD homologs are conserved in some archaea including *Sulfolobus*, which most likely function as a putative Fe-S scaffold complex [119, 120]. On the other hand, cysteine desulfurase (CdsA/IscS/SufS) homologs are rather poorly conserved in these archaea, and remain to be assigned in future study. A transient Fe-S cluster dislocated from the archaeal SUF scaffold protein is subsequently transferred (presumably using an ApbC/NBP35 homolog) and inserted into recipient apoproteins, either during or shortly after the apoprotein generation and before the folding into its native-like conformation. In many recipient Fe-S protein modules, the Fe-S cluster is assembled to loop regions and is often located near the protein surface. The *in vivo* assembly of a biological Fe-S cluster in a (recipient) protein scaffold is determined primarily by the nature and spacing of the ligands in the cluster binding loops. These loops probably define the geometric tolerance and thiophilicity of iron ions and thereby play a decisive role in a (kinetic) "native-like" semiordered folding intermediate. I hope that this short review will stimulate further research work, through which the answers to many open questions will be integrated into a comprehensive view on the biogenesis and maintenance of the archaeal Fe-S world.

Acknowledgments

The author would like to thank numerous colleagues and collaborators whose names appear in the references, and Professor Tairo Oshima and Professor Takeshi Nishino, whose support over many years is gratefully acknowledged. The investigation of this author's group was supported in part by a JSPS Grant BSAR-507, JSPS Grants-in-Aid (11169237, 15770088, 18608004, 21659111) to T. Iwasaki, and more recently approved by JSPS (to T. Iwasaki) and NSF (CHE-1026541 to S. A. Dikanov) under the JSPS-NSF International Collaborations in Chemistry (ICC).

References

[1] R. H. Holm, P. Kennepohl, and E. I. Solomon, "Structural and functional aspects of metal sites in biology," *Chemical Reviews*, vol. 96, no. 7, pp. 2239–2314, 1996.

[2] H. Beinert, R. H. Holm, and E. Münck, "Iron-sulfur clusters: nature's modular, multipurpose structures," *Science*, vol. 277, no. 5326, pp. 653–659, 1997.

[3] D. C. Rees and J. B. Howard, "The interface between the biological and inorganic worlds: iron-sulfur metalloclusters," *Science*, vol. 300, no. 5621, pp. 929–931, 2003.

[4] M. W. W. Adams, "Novel iron-sulfur centers in metalloenzymes and redox proteins from extremely thermophilic bacteria," *Advances in Inorganic Chemistry*, vol. 38, pp. 341–396, 1992.

[5] H. Beinert, "Iron-sulfur proteins: ancient structures, still full of surprises," *Journal of Biological Inorganic Chemistry*, vol. 5, no. 1, pp. 2–15, 2000.

[6] H. Beinert, M. C. Kennedy, and C. D. Stout, "Aconitase as iron-sulfur protein, enzyme, and iron-regulatory protein," *Chemical Reviews*, vol. 96, no. 7, pp. 2335–2373, 1996.

[7] D. C. Johnson, D. R. Dean, A. D. Smith, and M. K. Johnson, "Structure, function, and formation of biological iron-sulfur clusters," *Annual Review of Biochemistry*, vol. 74, pp. 247–281, 2005.

[8] R. Lill and U. Mühlenhoff, "Maturation of iron-sulfur proteins in eukaryotes: mechanisms, connected processes, and diseases," *Annual Review of Biochemistry*, vol. 77, pp. 669–700, 2008.

[9] T. A. Rouault and W.-H. Tong, "Iron-sulphur cluster biogenesis and mitochondrial iron homeostasis," *Nature Reviews Molecular Cell Biology*, vol. 6, no. 4, pp. 345–351, 2005.

[10] C. R. Woese, O. Kandler, and M. L. Wheelis, "Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 87, no. 12, pp. 4576–4579, 1990.

[11] G. J. Olsen, C. R. Woese, and R. Overbeek, "The winds of (evolutionary) change: breathing new life into microbiology," *Journal of Bacteriology*, vol. 176, no. 1, pp. 1–6, 1994.

[12] K. O. Stetter, "Microbial life in hyperthermal environments," *ASM News*, vol. 61, pp. 285–290, 1995.

[13] K. O. Stetter, "Hyperthermophiles in the history of life," *Philosophical Transactions of the Royal Society B*, vol. 361, no. 1474, pp. 1837–1842, 2006.

[14] L. Kerscher and D. Oesterhelt, "Ferredoxin is the coenzyme of α-ketoacid oxidoreductases in *Halobacterium halobium*," *FEBS Letters*, vol. 83, no. 2, pp. 197–201, 1977.

[15] L. Kerscher, S. Nowitzki, and D. Oesterhelt, "Thermoacidophilic archaebacteria contain bacterial-type ferredoxins acting as electron acceptors of 2-oxoacid: ferredoxin oxidoreductases," *European Journal of Biochemistry*, vol. 128, no. 1, pp. 223–230, 1982.

[16] T. Iwasaki, T. Wakagi, Y. Isogai, K. Tanaka, T. Iizuka, and T. Oshima, "Functional and evolutionary implications of a [3Fe-4S] cluster of the dicluster-type ferredoxin from the thermoacidophilic archaeon, *Sulfolobus* sp. strain 7," *Journal of Biological Chemistry*, vol. 269, no. 47, pp. 29444–29450, 1994.

[17] T. Iwasaki and T. Oshima, "Ferredoxin and related enzymes from *Sulfolobus*," *Methods in Enzymology*, vol. 334, pp. 3–22, 2001.

[18] H. Beinert, "Semi-micro methods for analysis of labile sulfide and of labile sulfide plus sulfane sulfur in unusually stable iron-sulfur proteins," *Analytical Biochemistry*, vol. 131, no. 2, pp. 373–378, 1983.

[19] M. Ciaramella, A. Napoli, and M. Rossi, "Another extreme genome: how to live at pH 0," *Trends in Microbiology*, vol. 13, no. 2, pp. 49–51, 2005.

[20] A. Angelov and W. Liebl, "Insights into extreme thermoacidophily based on genome analysis of *Picrophilus torridus* and other thermoacidophilic archaea," *Journal of Biotechnology*, vol. 126, no. 1, pp. 3–10, 2006.

[21] C. Baker-Austin and M. Dopson, "Life in acid: pH homeostasis in acidophiles," *Trends in Microbiology*, vol. 15, no. 4, pp. 165–171, 2007.

[22] K. S. Auernik, C. R. Cooper, and R. M. Kelly, "Life in hot acid: pathway analyses in extremely thermoacidophilic archaea," *Current Opinion in Biotechnology*, vol. 19, no. 5, pp. 445–453, 2008.

[23] M. W. W. Adams, "Enzymes and proteins from organisms that grow near and above 100°C," *Annual Review of Microbiology*, vol. 47, pp. 627–658, 1993.

[24] M. W. W. Adams, "Biochemical diversity among sulfur-dependent, hyperthermophilic microorganisms," *FEMS Microbiology Reviews*, vol. 15, no. 2-3, pp. 261–277, 1994.

[25] M. K. Johnson, R. E. Duderstadt, and E. C. Duin, "Biological and synthetic $[Fe_3S_4]$ clusters," *Advances in Inorganic Chemistry*, vol. 47, pp. 1–82, 1999.

[26] F. E. Jenney Jr. and M. W. W. Adams, "Hydrogenases of the model hyperthermophiles," *Annals of the New York Academy of Sciences*, vol. 1125, pp. 252–266, 2008.

[27] L. Kerscher and D. Oesterhelt, "Pyruvate: ferredoxin oxidoreductase—new findings on an ancient enzyme," *Trends in Biochemical Sciences*, vol. 7, no. 10, pp. 371–374, 1982.

[28] L. Kerscher and D. Oesterhelt, "The catalytic mechanism of 2-oxoacid: ferredoxin oxidoreductases from *Halobacterium halobium*. One electron transfer at two distinct steps of the catalytic cycle," *European Journal of Biochemistry*, vol. 116, no. 3, pp. 595–600, 1981.

[29] L. Kerscher and D. Oesterhelt, "Purification and properties of two 2-oxoacid: ferredoxin oxidoreductases from *Halobacterium halobium*," *European Journal of Biochemistry*, vol. 116, no. 3, pp. 587–594, 1981.

[30] Q. Zhang, T. Iwasaki, T. Wakagi, and T. Oshima, "2-oxoacid:ferredoxin oxidoreductase from the thermoacidophilic archaeon, *Sulfolobus* sp. strain 7," *Journal of Biochemistry*, vol. 120, no. 3, pp. 587–599, 1996.

[31] A. Kletzin and M. W. W. Adams, "Molecular and phylogenetic characterization of pyruvate and 2-ketoisovalerate ferredoxin oxidoreductases from *Pyrococcus furiosus* and pyruvate ferredoxin oxidoreductase from *Thermotoga maritima*," *Journal of Bacteriology*, vol. 178, no. 1, pp. 248–257, 1996.

[32] E. Chabrière, M.-H. Charon, A. Volbeda, L. Pieulle, E. C. Hatchikian, and J.-C. Fontecilla-Camps, "Crystal structures of the key anaerobic enzyme pyruvate ferredoxin oxidoreductase free and in complex with pyruvate," *Nature Structural Biology*, vol. 6, no. 2, pp. 182–190, 1999.

[33] E. Chabrière, X. Vernède, B. Guigliarelli, M.-H. Charon, E. C. Hatchikian, and J. C. Fontecilla-Camps, "Crystal structure of the free radical intermediate of pyruvate: ferredoxin oxidoreductase," *Science*, vol. 294, no. 5551, pp. 2559–2563, 2001.

[34] W. Plaga, F. Lottspeich, and D. Oesterhelt, "Improved purification, crystallization and primary structure of pyruvate: ferredoxin oxidoreductase from *Halobacterium halobium*," *European Journal of Biochemistry*, vol. 205, no. 1, pp. 391–397, 1992.

[35] T. Iwasaki, A. Kounosu, M. Aoshima et al., "Novel [2Fe-2S]-type redox center C in sdhC of archaeal respiratory complex II from *Sulfolobus tokodaii* strain 7," *Journal of Biological Chemistry*, vol. 277, no. 42, pp. 39642–39648, 2002.

[36] C. Enroth, B. T. Eger, K. Okamoto, T. Nishino, T. Nishino, and E. F. Pai, "Crystal structures of bovine milk xanthine dehydrogenase and xanthine oxidase: structure-based mechanism of conversion," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 97, no. 20, pp. 10723–10728, 2000.

[37] L. A. Sazanov and P. Hinchliffe, "Structure of the hydrophilic domain of respiratory complex I from *Thermus thermophilus*," *Science*, vol. 311, no. 5766, pp. 1430–1436, 2006.

[38] A. Hirata, B. J. Klein, and K. S. Murakami, "The X-ray crystal structure of RNA polymerase from Archaea," *Nature*, vol. 451, no. 7180, pp. 851–854, 2008.

[39] M. Teixeira, R. Batista, A. P. Campos et al., "A seven-iron ferredoxin from the thermoacidophilic archaeon *Desulfurolobus ambivalens*," *European Journal of Biochemistry*, vol. 227, no. 1-2, pp. 322–327, 1995.

[40] J. L. Breton, J. L. C. Duff, J. N. Butt et al., "Identification of the iron-sulfur clusters in a ferredoxin from the archaeon *Sulfolobus acidocaldarius*—evidence for a reduced [3Fe-4S] cluster with pH-dependent electronic properties," *European Journal of Biochemistry*, vol. 233, no. 3, pp. 937–946, 1995.

[41] T. Iwasaki, T. Suzuki, T. Kon et al., "Novel zinc-containing ferredoxin family in thermoacidophilic archaea," *Journal of Biological Chemistry*, vol. 272, no. 6, pp. 3453–3458, 1997.

[42] C. M. Gomes, A. Faria, J. C. Carita et al., "Di-cluster, seven-iron ferredoxins from hyperthermophilic *Sulfolobales*," *Journal of Biological Inorganic Chemistry*, vol. 3, no. 5, pp. 499–507, 1998.

[43] N. J. Cosper, C. M. V. Stålhandske, H. Iwasaki, T. Oshima, R. A. Scott, and T. Iwasaki, "Structural conservation of the isolated zinc site in archaeal zinc- containing ferredoxins as revealed by X-ray absorption spectroscopic analysis and its evolutionary implications," *Journal of Biological Chemistry*, vol. 274, no. 33, pp. 23160–23168, 1999.

[44] T. Iwasaki, E. Watanabe, D. Ohmori et al., "Spectroscopic investigation of selective cluster conversion of archaeal zinc-containing ferredoxin from *Sulfolobus* sp. strain 7," *Journal of Biological Chemistry*, vol. 275, no. 33, pp. 25391–25401, 2000.

[45] R. C. Conover, A. T. Kowal, W. Fu et al., "Spectroscopic characterization of the novel iron-sulfur cluster in *Pyrococcus furiosus* ferrodoxin," *Journal of Biological Chemistry*, vol. 265, no. 15, pp. 8533–8541, 1990.

[46] T. Imai, K. Taguchi, Y. Ogawara et al., "Characterization and cloning of an extremely thermostable, *Pyrococcus furiosus*-type 4Fe ferredoxin from *Thermococcus profundus*," *Journal of Biochemistry*, vol. 130, no. 5, pp. 649–655, 2001.

[47] T. Fujii, Y. Hata, T. Wakagi, N. Tanaka, and T. Oshima, "Novel zinc-binding centre in thermoacidophilic archaeal ferredoxins," *Nature Structural Biology*, vol. 3, no. 10, pp. 834–837, 1996.

[48] T. Fujii, Y. Hata, M. Oozeki et al., "The crystal structure of zinc-containing ferredoxin from the thermoacidophilic archaeon *Sulfolobus* sp. strain 7," *Biochemistry*, vol. 36, no. 6, pp. 1505–1513, 1997.

[49] C. Frazão, D. Aragão, R. Coelho et al., "Crystallographic analysis of the intact metal centres $[3Fe\text{-}4S]^{1+/0}$ and $[4Fe\text{-}4S]^{2+/1+}$ in a Zn^{2+}-containing ferredoxin," *FEBS Letters*, vol. 582, no. 5, pp. 763–767, 2008.

[50] K. Kojoh, H. Matsuzawa, and T. Wakagi, "Zinc and an N-terminal extra stretch of the ferredoxin from a thermoacidophilic archaeon stabilize the molecule at high temperature," *European Journal of Biochemistry*, vol. 264, no. 1, pp. 85–91, 1999.

[51] S. S. Leal and C. M. Gomes, "Studies of the molten globule state of ferredoxin: structural characterization and implications on protein folding and iron-sulfur center assembly," *Proteins*, vol. 68, no. 3, pp. 606–616, 2007.

[52] S. S. Leal and C. M. Gomes, "On the relative contribution of ionic interactions over iron-sulfur clusters to ferredoxin stability," *Biochimica et Biophysica Acta*, vol. 1784, no. 11, pp. 1596–1600, 2008.

[53] R. Rocha, S. S. Leal, V. H. Teixeira et al., "Natural domain design: enhanced thermal stability of a zinc-lacking ferredoxin isoform shows that a hydrophobic core efficiently replaces the structural metal site," *Biochemistry*, vol. 45, no. 34, pp. 10376–10384, 2006.

[54] S. Janssen, J. Trincão, M. Teixeira, G. Schäfer, and S. Anemüller, "Ferredoxins from the archaeon *Acidianus ambivalens*: overexpression and characterization of the non-zinc-containing ferredoxin FdB," *Biological Chemistry*, vol. 382, no. 10, pp. 1501–1507, 2001.

[55] L. Kerscher, D. Oesterhelt, R. Cammack, and D. O. Hall, "A new plant type ferredoxin from halobacteria," *European Journal of Biochemistry*, vol. 71, no. 1, pp. 101–108, 1976.

[56] F. Pfeifer, J. Griffig, and D. Oesterhelt, "The *fdx* gene encoding the [2Fe-2S] ferredoxin of Halobacterium salinarium (H. halobium)," *Molecular and General Genetics*, vol. 239, no. 1-2, pp. 66–71, 1993.

[57] F. Frolow, M. Harel, J. L. Sussman, M. Mevarech, and M. Shoham, "Insights into protein adaptation to a saturated salt environment from the crystal structure of a halophilic 2Fe-2S ferredoxin," *Nature Structural Biology*, vol. 3, no. 5, pp. 452–458, 1996.

[58] R. Moll and G. Schäfer, "Chemiosmotic H^+ cycling across the plasma membrane of the thermoacidophilic archaebacterium *Sulfolobus acidocaldarius*," *FEBS Letters*, vol. 232, no. 2, pp. 359–363, 1988.

[59] D. G. Searcy, "*Thermoplasma acidophilum*: intracellular pH and potassium concentration," *Biochimica et Biophysica Acta*, vol. 451, no. 1, pp. 278–286, 1976.

[60] M. Lübben and G. Schäfer, "Chemiosmotic energy conversion of the archaebacterial thermoacidophile *Sulfolobus acidocaldarius*: oxidative phosphorylation and the presence of an F_o-related N,N'-dicyclohexylcarbodiimide-binding proteolipid," *Journal of Bacteriology*, vol. 171, no. 11, pp. 6106–6116, 1989.

[61] O. Djaman, F. W. Outten, and J. A. Imlay, "Repair of oxidized iron-sulfur clusters in *Escherichia coli*," *Journal of Biological Chemistry*, vol. 279, no. 43, pp. 44590–44599, 2004.

[62] S. Knapp, S. Kardinahl, N. Hellgren, G. Tibbelin, G. Schäfer, and R. Ladenstein, "Refined crystal structure of a superoxide dismutase from the hyperthermophilic archaeon *Sulfolobus acidocaldarius* at 2.2 Å resolution," *Journal of Molecular Biology*, vol. 285, no. 2, pp. 689–702, 1999.

[63] T. Ursby, B. S. Adinolfi, S. Al-Karadaghi, E. De Vendittis, and V. Bocchini, "Iron superoxide dismutase from the archaeon *Sulfolobus solfataricus*: analysis of structure and

thermostability," *Journal of Molecular Biology*, vol. 286, no. 1, pp. 189–205, 1999.

[64] B. Wiedenheft, J. Mosolf, D. Willits et al., "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*, vol. 102, no. 30, pp. 10551–10556, 2005.

[65] G. H. Gauss, P. Benas, B. Wiedenheft, M. Young, T. Douglas, and C. M. Lawrence, "Structure of the DPS-like protein from *Sulfolobus solfataricus* reveals a bacterioferritin-like dimetal binding site within a DPS-like dodecameric assembly," *Biochemistry*, vol. 45, no. 36, pp. 10815–10827, 2006.

[66] W. S. Maaty, B. Wiedenheft, P. Tarlykov et al., "Something old, something new, something borrowed; how the thermoacidophilic archaeon *Sulfolobus solfataricus* responds to oxidative stress," *PLoS ONE*, vol. 4, no. 9, Article ID e6964, 2009.

[67] J. A. Hayden and M. P. Hendrich, "EPR spectroscopy and catalase activity of manganese-bound DNA-binding protein from nutrient starved cells," *Journal of Biological Inorganic Chemistry*, vol. 15, no. 5, pp. 729–736, 2010.

[68] G. Schäfer, M. Engelhard, and V. Müller, "Bioenergetics of the archaea," *Microbiology and Molecular Biology Reviews*, vol. 63, no. 3, pp. 570–620, 1999.

[69] K. Denda, J. Konishi, K. Hajiro, T. Oshima, T. Date, and M. Yoshida, "Structure of an ATPase operon of an acidothermophilic archaebacterium, *Sulfolobus acidocaldarius*," *Journal of Biological Chemistry*, vol. 265, no. 35, pp. 21509–21513, 1990.

[70] M. Lübben and K. Morand, "Novel prenylated hemes as cofactors of cytochrome oxidases. Archaea have modified hemes A and O," *Journal of Biological Chemistry*, vol. 269, no. 34, pp. 21473–21479, 1994.

[71] T. Iwasaki, K. Matsuura, and T. Oshima, "Resolution of the aerobic respiratory system of the thermoacidophilic archaeon, *Sulfolobus* sp. strain 7: I. The archaeal terminal oxidase supercomplex is a functional fusion of respiratory complexes III and IV with no *c*-type cytochromes," *Journal of Biological Chemistry*, vol. 270, no. 52, pp. 30881–30892, 1995.

[72] M. Lübben, "Cytochromes of archaeal electron transfer chains," *Biochimica et Biophysica Acta*, vol. 1229, no. 1, pp. 1–22, 1995.

[73] T. Suzuki, T. Iwasaki, T. Uzawa et al., "*Sulfolobus tokodaii* sp. nov. (f. *Sulfolobus* sp. strain 7), a new member of the genus *Sulfolobus* isolated from Beppu Hot Springs, Japan," *Extremophiles*, vol. 6, no. 1, pp. 39–44, 2002.

[74] Y. Kawarabayasi, Y. Hino, H. Horikawa et al., "Complete genome sequence of an aerobic thermoacidophilic crenarchaeon, *Sulfolobus tokodaii* strain," *DNA Research*, vol. 8, no. 4, pp. 123–140, 2001.

[75] Q. She, R. K. Singh, F. Confalonieri et al., "The complete genome of the crenarchaeon *Sulfolobus solfataricus* P2," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 98, no. 14, pp. 7835–7840, 2001.

[76] M. Lübben, B. Kolmerer, and M. Saraste, "An archaebacterial terminal oxidase combines core structures of two mitochondrial respiratory complexes," *EMBO Journal*, vol. 11, no. 3, pp. 805–812, 1992.

[77] M. Lübben, S. Arnaud, J. Castresana, A. Warne, S. P. J. Albracht, and M. Saraste, "A second terminal oxidase in *Sulfolobus acidocaldarius*," *European Journal of Biochemistry*, vol. 224, no. 1, pp. 151–159, 1994.

[78] J. Castresana, M. Lübben, and M. Saraste, "New archaebacterial genes coding for redox proteins: implications for the evolution of aerobic metabolism," *Journal of Molecular Biology*, vol. 250, no. 2, pp. 202–210, 1995.

[79] L. Chen, K. Brügger, M. Skovgaard et al., "The genome of *Sulfolobus acidocaldarius*, a model organism of the Crenarchaeota," *Journal of Bacteriology*, vol. 187, no. 14, pp. 4992–4999, 2005.

[80] J. Hemp and R. B. Gennis, "Diversity of the heme-copper superfamily in archaea: insights from genomics and structural modeling," *Results and Problems in Cell Differentiation*, vol. 45, pp. 1–31, 2008.

[81] J. P. Hosler, S. Ferguson-Miller, and D. A. Mills, "Energy transduction: proton transfer through the respiratory complexes," *Annual Review of Biochemistry*, vol. 75, pp. 165–187, 2006.

[82] K. Shimokata, Y. Katayama, H. Murayama et al., "The proton pumping pathway of bovine heart cytochrome *c* oxidase," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 104, no. 10, pp. 4200–4205, 2007.

[83] L. Qin, J. Liu, D. A. Mills, D. A. Proshlyakov, C. Hiser, and S. Ferguson-Miller, "Redox-dependent conformational changes in cytochrome *c* oxidase suggest a gating mechanism for proton uptake," *Biochemistry*, vol. 48, no. 23, pp. 5121–5130, 2009.

[84] G. Brändén, R. B. Gennis, and P. Brzezinski, "Transmembrane proton translocation by cytochrome *c* oxidase," *Biochimica et Biophysica Acta*, vol. 1757, no. 8, pp. 1052–1063, 2006.

[85] E. A. Berry, M. Guergova-Kuras, L.-S. Huang, and A. R. Crofts, "Structure and function of cytochrome *bc* complexes," *Annual Review of Biochemistry*, vol. 69, pp. 1005–1075, 2000.

[86] A. R. Crofts, "The cytochrome bc_1 complex: function in the context of structure," *Annual Review of Physiology*, vol. 66, pp. 689–733, 2004.

[87] W. A. Cramer, H. Zhang, J. Yan, G. Kurisu, and J. L. Smith, "Transmembrane traffic in the cytochrome b_6f complex," *Annual Review of Biochemistry*, vol. 75, pp. 769–790, 2006.

[88] U. Kappler, L. I. Sly, and A. G. McEwan, "Respiratory gene clusters of *Metallosphaera sedula*—differential expression and transcriptional organization," *Microbiology*, vol. 151, no. 1, pp. 35–43, 2005.

[89] S. Bathe and P. R. Norris, "Ferrons iron- and sulfur-induced genes in *Sulfolobus metallicus*," *Applied and Environmental Microbiology*, vol. 73, no. 8, pp. 2491–2497, 2007.

[90] K. S. Auernik and R. M. Kelly, "Identification of components of electron transport chains in the extremely thermoacidophilic crenarchaeon *Metallosphaera sedula* through iron and sulfur compound oxidation transcriptomes," *Applied and Environmental Microbiology*, vol. 74, no. 24, pp. 7723–7732, 2008.

[91] A. Brooun, J. Bell, T. Freitas, R. W. Larsen, and M. Alam, "An archaeal aerotaxis transducer combines subunit I core structures of eukaryotic cytochrome *c* oxidase and eubacterial methyl-accepting chemotaxis proteins," *Journal of Bacteriology*, vol. 180, no. 7, pp. 1642–1646, 1998.

[92] R. I. Samoilova, D. Kolling, T. Uzawa, T. Iwasaki, A. R. Crofts, and S. A. Dikanov, "The interaction of the Rieske iron-sulfur protein with occupants of the Q_o-site of the bc_1 complex, probed by electron spin echo envelope modulation," *Journal of Biological Chemistry*, vol. 277, no. 7, pp. 4605–4608, 2002.

[93] M. L. Reno, N. L. Held, C. J. Fields, P. V. Burke, and R. J. Whitaker, "Biogeography of the *Sulfolobus islandicus* pan-genome," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 106, no. 21, pp. 8605–8610, 2009.

[94] K. Deigweiher, T. L. Drell IV, A. Prutsch, A. J. Scheidig, and M. Lübben, "Expression, isolation, and crystallization of the catalytic domain of CopB, a putative copper transporting ATPase from the thermoacidophilic archaeon *Sulfolobus solfataricus*," *Journal of Bioenergetics and Biomembranes*, vol. 36, no. 1, pp. 151–159, 2004.

[95] B. De Hertogh, A.-C. Lantin, P. V. Baret, and A. Goffeau, "The archaeal P-type ATPases," *Journal of Bioenergetics and Biomembranes*, vol. 36, no. 1, pp. 135–142, 2004.

[96] R. Lill and G. Kispal, "Maturation of cellular Fe-S proteins: an essential function of mitochondria," *Trends in Biochemical Sciences*, vol. 25, no. 8, pp. 352–356, 2000.

[97] J. Frazzon, J. R. Fick, and D. R. Dean, "Biosynthesis of iron-sulphur clusters is a complex and highly conserved process," *Biochemical Society Transactions*, vol. 30, no. 4, pp. 680–685, 2002.

[98] Y. Takahashi and U. Tokumoto, "A third bacterial system for the assembly of iron-sulfur clusters with homologs in Archaea and plastids," *Journal of Biological Chemistry*, vol. 277, no. 32, pp. 28380–28383, 2002.

[99] S. Bandyopadhyay, K. Chandramouli, and M. K. Johnson, "Iron-sulfur cluster biosynthesis," *Biochemical Society Transactions*, vol. 36, no. 6, pp. 1112–1119, 2008.

[100] M. Fontecave and S. Ollagnier-de-Choudens, "Iron-sulfur cluster biosynthesis in bacteria: mechanisms of cluster assembly and transfer," *Archives of Biochemistry and Biophysics*, vol. 474, no. 2, pp. 226–237, 2008.

[101] C. Ayala-Castro, A. Saini, and F. W. Outten, "Fe-S cluster assembly pathways in bacteria," *Microbiology and Molecular Biology Reviews*, vol. 72, no. 1, pp. 110–125, 2008.

[102] Y. Takahashi and M. Nakamura, "Functional assignment of the *ORF2-iscS-iscU-iscA-hscB-hscA-fdx-ORF3* gene cluster involved in the assembly of Fe-S clusters in *Escherichia coli*," *Journal of Biochemistry*, vol. 126, no. 5, pp. 917–926, 1999.

[103] U. Tokumoto and Y. Takahashi, "Genetic analysis of the *isc* operon in *Escherichia coli* involved in the biogenesis of cellular iron-sulfur proteins," *Journal of Biochemistry*, vol. 130, no. 1, pp. 63–71, 2001.

[104] U. Tokumoto, S. Nomura, Y. Minami et al., "Network of protein-protein interactions among iron-sulfur cluster assembly proteins in *Escherichia coli*," *Journal of Biochemistry*, vol. 131, no. 5, pp. 713–719, 2002.

[105] S. I. Patzer and K. Hantke, "SufS is a NifS-like protein, and sufd is necessary for stability of the [2Fe-2S] FhuF protein in *Escherichia coli*," *Journal of Bacteriology*, vol. 181, no. 10, pp. 3307–3309, 1999.

[106] N. Rouhier, J. Couturier, M. K. Johnson, and J.-P. Jacquot, "Glutaredoxins: roles in iron homeostasis," *Trends in Biochemical Sciences*, vol. 35, no. 1, pp. 43–52, 2010.

[107] D. J. A. Netz, A. J. Pierik, M. Stümpfig, U. Mühlenhoff, and R. Lill, "The Cfd1-Nbp35 complex acts as a scaffold for iron-sulfur protein assembly in the yeast cytosol," *Nature Chemical Biology*, vol. 3, no. 5, pp. 278–286, 2007.

[108] L. Zheng and D. R. Dean, "Catalytic formation of a nitrogenase iron-sulfur cluster," *Journal of Biological Chemistry*, vol. 269, no. 29, pp. 18723–18726, 1994.

[109] L. Zheng, V. L. Cash, D. H. Flint, and D. R. Dean, "Assembly of iron-sulfur clusters. Identification of an *iscSUA-hscBA-fdx* gene cluster from *Azotobacter vinelandii*," *Journal of Biological Chemistry*, vol. 273, no. 21, pp. 13264–13272, 1998.

[110] D. H. Flint, "*Escherichia coli* contains a protein that is homologous in function and N-terminal sequence to the protein encoded by the *nifS* gene of *Azotobacter vinelandii* and that can participate in the synthesis of the Fe-S cluster of dihydroxy-acid dehydratase," *Journal of Biological Chemistry*, vol. 271, no. 27, pp. 16068–16074, 1996.

[111] S.-I. Kato, H. Mihara, T. Kurihara et al., "Cys-328 of IscS and Cys-63 of IscU are the sites of disulfide bridge formation in a covalently bound IscS/IscU complex: implications for the mechanism of iron-sulfur cluster assembly," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 99, no. 9, pp. 5948–5952, 2002.

[112] L. Loiseau, S. Ollagnier-de-Choudens, L. Nachin, M. Fontecave, and F. Barras, "Biogenesis of Fe-S cluster by the bacterial suf system. SufS and SufE form a new type of cysteine desulfurase," *Journal of Biological Chemistry*, vol. 278, no. 40, pp. 38352–38359, 2003.

[113] F. W. Outten, M. J. Wood, F. M. Muñoz, and G. Storz, "The SufE protein and the SufBCD complex enhance SufS cysteine desulfurase activity as part of a sulfur transfer pathway for Fe-S cluster assembly in *Escherichia coli*," *Journal of Biological Chemistry*, vol. 278, no. 46, pp. 45713–45719, 2003.

[114] J. N. Agar, C. Krebs, J. Frazzon, B. H. Huynh, D. R. Dean, and M. K. Johnson, "IscU as a scaffold for iron-sulfur cluster biosynthesis: sequential assembly of [2Fe-2S] and [4Fe-4S] clusters in IscU," *Biochemistry*, vol. 39, no. 27, pp. 7856–7862, 2000.

[115] K. Chandramouli and M. K. Johnson, "HscA and HscB stimulate [2Fe-2S] cluster transfer from IscU to apoferredoxin in an ATP-dependent reaction," *Biochemistry*, vol. 45, no. 37, pp. 11087–11095, 2006.

[116] K. Chandramouli, M.-C. Unciuleac, S. Naik, D. R. Dean, H. H. Boi, and M. K. Johnson, "Formation and properties of [4Fe-4S] clusters on the IscU scaffold protein," *Biochemistry*, vol. 46, no. 23, pp. 6804–6811, 2007.

[117] M.-C. Unciuleac, K. Chandramouli, S. Naik et al., "*In vitro* activation of apo-aconitase using a [4Fe-4S] cluster-loaded form of the IscU [Fe-S] cluster scaffolding protein," *Biochemistry*, vol. 46, no. 23, pp. 6812–6821, 2007.

[118] E. C. Raulfs, I. P. O'Carroll, P. C. Dos Santos, M.-C. Unciuleac, and D. R. Dean, "*In vivo* iron-sulfur cluster formation," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 105, no. 25, pp. 8591–8596, 2008.

[119] H. K. Chahal, Y. Dai, A. Saini, C. Ayala-Castro, and F. W. Outten, "The SufBCD Fe-S scaffold complex interacts with sufa for Fe-s cluster transfer," *Biochemistry*, vol. 48, no. 44, pp. 10644–10653, 2009.

[120] S. Wollers, G. Layer, R. Garcia-Serres et al., "Iron-sulfur (Fe-S) cluster assembly: the SufBCD complex is a new type of Fe-S scaffold with a flavin redox cofactor," *Journal of Biological Chemistry*, vol. 285, no. 30, pp. 23331–23341, 2010.

[121] C. J. Schwartz, J. L. Giel, T. Patschkowski et al., "IscR, an Fe-S cluster-containing transcription factor, represses expression of *Escherichia coli* genes encoding Fe-S cluster assembly proteins," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 98, no. 26, pp. 14895–14900, 2001.

[122] F. W. Outten, O. Djaman, and G. Storz, "A *suf* operon requirement for Fe-S cluster assembly during iron starvation in *Escherichia coli*," *Molecular Microbiology*, vol. 52, no. 3, pp. 861–872, 2004.

[123] U. Tokumoto, S. Kitamura, K. Fukuyama, and Y. Takahashi, "Interchangeability and distinct properties of bacterial Fe-S cluster assembly systems: functional replacement of the isc and suf operons in *Escherichia coli* with the *nifSU*-like operon from *Helicobacter pylori*," *Journal of Biochemistry*, vol. 136, no. 2, pp. 199–209, 2004.

[124] M. Zheng, X. Wang, L. J. Templeton, D. R. Smulski, R. A. LaRossa, and G. Storz, "DNA microarray-mediated transcriptional profiling of the *Escherichia coli* response to hydrogen peroxide," *Journal of Bacteriology*, vol. 183, no. 15, pp. 4562–4570, 2001.

[125] K.-C. Lee, W.-S. Yeo, and J.-H. Roe, "Oxidant-responsive induction of the suf operon, encoding a Fe-S assembly system, through Fur and IscR in *Escherichia coli*," *Journal of Bacteriology*, vol. 190, no. 24, pp. 8244–8247, 2008.

[126] D. Vinella, C. Brochier-Armanet, L. Loiseau, E. Talla, and F. Barras, "Iron-sulfur (Fe/S) protein biogenesis: phylogenomic and genetic studies of A-type carriers," *PLoS Genetics*, vol. 5, no. 5, Article ID e1000497, 2009.

[127] L. Nachin, L. Loiseau, D. Expert, and F. Barras, "SufC: an unorthodox cytoplasmic ABC/ATPase required for [Fe-S] biogenesis under oxidative stress," *EMBO Journal*, vol. 22, no. 3, pp. 427–437, 2003.

[128] A. L. Menon, F. L. Poole II, A. Cvetkovic et al., "Novel multiprotein complexes identified in the hyperthermophilic archaeon *Pyrococcus furiosus* by non-denaturing fractionation of the native proteome," *Molecular & Cellular Proteomics*, vol. 8, no. 4, pp. 735–751, 2009.

[129] V. Gupta, M. Sendra, S. G. Naik et al., "Native *Escherichia coli* SufA coexpressed with SufBCDSE purifies as a [2Fe-2S] protein and acts as an Fe-S transporter to Fe-S target enzymes," *Journal of the American Chemical Society*, vol. 131, no. 17, pp. 6149–6153, 2009.

[130] J. M. Boyd, A. J. Pierik, D. J. A. Netz, R. Lill, and D. M. Downs, "Bacterial ApbC can bind and effectively transfer iron-sulfur clusters," *Biochemistry*, vol. 47, no. 31, pp. 8195–8202, 2008.

[131] J. M. Boyd, J. L. Sondelski, and D. M. Downs, "Bacterial apbC protein has two biochemical activities that are required for *in vivo* function," *Journal of Biological Chemistry*, vol. 284, no. 1, pp. 110–118, 2009.

[132] H. Kohbushi, Y. Nakai, S. Kikuchi, T. Yabe, H. Hori, and M. Nakai, "*Arabidopsis* cytosolic Nbp35 homodimer can assemble both [2Fe-2S] and [4Fe-4S] clusters in two distinct domains," *Biochemical and Biophysical Research Communications*, vol. 378, no. 4, pp. 810–815, 2009.

[133] J. M. Boyd, R. M. Drevland, D. M. Downs, and D. E. Graham, "Archaeal ApbC/Nbp35 homologs function as iron-sulfur cluster carrier proteins," *Journal of Bacteriology*, vol. 191, no. 5, pp. 1490–1497, 2009.

[134] E. Skovran and D. M. Downs, "Lack of the ApbC or ApbE protein results in a defect in Fe-S cluster metabolism in *Salmonella enterica* serovar *typhimurium*," *Journal of Bacteriology*, vol. 185, no. 1, pp. 98–106, 2003.

[135] B. Zafrilla, R.M. Martínez-Espinosa, J. Esclapez, F. Pérez-Pomares, and M. J. Bonete, "SufS protein from *Haloferax volcanii* involved in Fe-S cluster assembly in haloarchaea," *Biochimica et Biophysica Acta*, vol. 1804, no. 7, pp. 1476–1482, 2010.

[136] Y. Kawarabayasi, Y. Hino, H. Horikawa et al., "Complete genome sequence of an aerobic hyper-thermophilic crenarchaeon, *Aeropyrum pernix* K1," *DNA Research*, vol. 6, no. 2, pp. 83–101, 1999.

[137] N. J. Cosper, D. M. Eby, A. Kounosu et al., "Redox-dependent structural changes in archaeal and bacterial Rieske-type [2Fe-2S] clusters," *Protein Science*, vol. 11, no. 12, pp. 2969–2973, 2002.

[138] A. Kounosu, Z. Li, N. J. Cosper et al., "Engineering a three-cysteine, one-histidine ligand environment into a new hyperthermophilic archaeal rieske-type [2Fe-2S] ferredoxin from *Sulfolobus solfataricus*," *Journal of Biological Chemistry*, vol. 279, no. 13, pp. 12519–12528, 2004.

[139] T. Iwasaki, A. Kounosu, D. R. J. Kolling et al., "Characterization of the pH-dependent resonance Raman transitions of archaeal and bacterial Rieske [2Fe-2S] proteins," *Journal of the American Chemical Society*, vol. 126, no. 15, pp. 4788–4789, 2004.

[140] S. A. Dikanov, A. A. Shubin, A. Kounosu, T. Iwasaki, and R. I. Samoilova, "A comparative, two-dimensional ^{14}N ESEEM characterization of reduced [2Fe-2S] clusters in hyperthermophilic archaeal high- and low-potential Rieske-type proteins," *Journal of Biological Inorganic Chemistry*, vol. 9, no. 6, pp. 753–767, 2004.

[141] T. Iwasaki, R. I. Samoilova, A. Kounosu, and S. A. Dikanov, "Two-dimensional pulsed electron spin resonance characterization of ^{15}N-labeled archaeal Rieske-type ferredoxin," *FEBS Letters*, vol. 583, no. 21, pp. 3467–3472, 2009.

[142] A. Kounosu, K. Hasegawa, T. Iwasaki, and T. Kumasaka, "Crystallization and preliminary X-ray diffraction studies of hyperthermophilic archaeal Rieske-type ferredoxin (ARF) from *Sulfolobus solfataricus* P1," *Acta Crystallographica F*, vol. 66, no. 7, pp. 842–845, 2010.

[143] J. R. Mason and R. Cammack, "The electron-transport proteins of hydroxylating bacterial dioxygenases," *Annual Review of Microbiology*, vol. 46, pp. 277–305, 1992.

[144] T. A. Link, "The structures of Rieske and Rieske-type proteins," *Advances in Inorganic Chemistry*, vol. 47, pp. 83–157, 1999.

[145] S. Iwata, M. Saynovits, T. A. Link, and H. Michel, "Structure of a water soluble fragment of the 'Rieske' iron-sulfur protein of the bovine heart mitochondrial cytochrome bc_1 complex determined by MAD phasing at 1.5 Å resolution," *Structure*, vol. 4, no. 5, pp. 567–579, 1996.

[146] C. J. Carrell, H. Zhang, W. A. Cramer, and J. L. Smith, "Biological identity and diversity in photosynthesis and respiration: structure of the lumen-side domain of the chloroplast Rieske protein," *Structure*, vol. 5, no. 12, pp. 1613–1625, 1997.

[147] B. Kauppi, K. Lee, E. Carredano et al., "Structure of an aromatic-ring-hydroxylating dioxygenasenaphthalene 1,2-dioxygenase," *Structure*, vol. 6, no. 5, pp. 571–586, 1998.

[148] C. L. Colbert, M. M.-J. Couture, L. D. Eltis, and J. T. Bolin, "A cluster exposed: structure of the rieske ferredoxin from biphenyl dioxygenase and the redox properties of Rieske Fe-S proteins," *Structure*, vol. 8, no. 12, pp. 1267–1278, 2000.

[149] H. Bönisch, C. L. Schmidt, G. Schäfer, and R. Ladenstein, "The structure of the soluble domain of an archaeal Rieske iron-sulfur protein at 1.1 Å resolution," *Journal of Molecular Biology*, vol. 319, no. 3, pp. 791–805, 2002.

[150] L. M. Hunsicker-Wang, A. Heine, Y. Chen et al., "High-resolution structure of the soluble, respiratory-type Rieske protein from *Thermus thermophilus*: analysis and comparison," *Biochemistry*, vol. 42, no. 24, pp. 7303–7317, 2003.

[151] D. J. Kolling, J. S. Brunzelle, S. Lhee, A. R. Crofts, and S. K. Nair, "Atomic resolution structures of Rieske iron-sulfur protein: role of hydrogen bonds in tuning the redox potential of iron-sulfur clusters," *Structure*, vol. 15, no. 1, pp. 29–38, 2007.

[152] M. W. Day, B. T. Hsu, L. Joshua-Tor et al., "X-ray crystal structures of the oxidized and reduced forms of the rubredoxin from the marine hyperthermophilic archaebacterium *Pyrococcus furiosus*," *Protein Science*, vol. 1, no. 11, pp. 1494–1507, 1992.

[153] T. Iwasaki, A. Kounosu, Y. Tao et al., "Rational design of a mononuclear metal site into the archaeal Rieske-type protein scaffold," *Journal of Biological Chemistry*, vol. 280, no. 10, pp. 9129–9134, 2005.

[154] J. Meyer, J. Gagnon, J. Gaillard et al., "Assembly of a $[2Fe\text{-}2S]^{2+}$ cluster in a molecular variant of *Clostridium pasteurianum* rubredoxin," *Biochemistry*, vol. 36, no. 43, pp. 13374–13380, 1997.

[155] P. A. Frey, A. D. Hegeman, and F. J. Ruzicka, "The radical SAM superfamily," *Critical Reviews in Biochemistry and Molecular Biology*, vol. 43, no. 1, pp. 63–88, 2008.

[156] K. S. Duschene, S. E. Veneziano, S. C. Silver, and J. B. Broderick, "Control of radical chemistry in the AdoMet radical enzymes," *Current Opinion in Chemical Biology*, vol. 13, no. 1, pp. 74–83, 2009.

[157] H. Irving and R. J. P. Williams, "The stability of transition-metal complexes," *Journal of the Chemical Society*, pp. 3192–3210, 1953.

Hindawi Publishing Corporation
Archaea
Volume 2010, Article ID 710303, 10 pages
doi:10.1155/2010/710303

Research Article

Archaeal Ubiquitin-Like Proteins: Functional Versatility and Putative Ancestral Involvement in tRNA Modification Revealed by Comparative Genomic Analysis

Kira S. Makarova and Eugene V. Koonin

National Center for Biotechnology Information, NLM, National Institutes of Health, Bethesda, MD 20894, USA

Correspondence should be addressed to Eugene V. Koonin, koonin@ncbi.nlm.nih.gov

Received 18 May 2010; Accepted 20 July 2010

Academic Editor: Julie Maupin-Furlow

The recent discovery of protein modification by SAMPs, ubiquitin-like (Ubl) proteins from the archaeon *Haloferax volcanii*, prompted a comprehensive comparative-genomic analysis of archaeal Ubl protein genes and the genes for enzymes thought to be functionally associated with Ubl proteins. This analysis showed that most archaea encode members of two major groups of Ubl proteins with the β-grasp fold, the ThiS and MoaD families, and indicated that the ThiS family genes are rarely linked to genes for thiamine or Mo/W cofactor metabolism enzymes but instead are most often associated with genes for enzymes of tRNA modification. Therefore it is hypothesized that the ancestral function of the archaeal Ubl proteins is sulfur insertion into modified nucleotides in tRNAs, an activity analogous to that of the URM1 protein in eukaryotes. Together with additional, previously described genomic associations, these findings indicate that systems for protein quality control operating at different levels, including tRNA modification that controls translation fidelity, protein ubiquitination that regulates protein degradation, and, possibly, mRNA degradation by the exosome, are functionally and evolutionarily linked.

1. Introduction

Ubiquitination (ubiquitylation) of proteins is an ancestral, pivotal process in eukaryotes that governs protein trafficking and turnover, signaling, heterochromatin remodeling, and other processes [1–3]. All eukaryotes possess an elaborate system that includes a variety of small proteins of the ubiquitin (Ub) family, E1 Ub-activating, E2 Ub-conjugating, and E3 Ub-ligase enzymes, as well as a broad diversity of deubiquitinating enzymes (DUBs) [1, 2, 4]. Ubiquitin conjugation through the formation of isopeptide bonds by the e-amino groups of two conserved lysines of the Ub molecule (K48 and K63) determines the fate of most proteins in eukaryotic cells, in terms of both topogenesis and degradation. The functioning of Ub-centered signaling systems is regulated through the activities of numerous, specific Ub-binding domains and proteins.

Ubiquitin is one of the most highly conserved eukaryotic proteins, and the evolution of the Ub system is fairly well studied [1, 5–8]. In particular, it has been shown that Ub homologs in bacteria and most likely in archaea are involved in thiamine and molybdenum (Mo)/tungsten (W) cofactor biosynthesis along with functionally linked homologs of E1 enzymes; in addition, E2 family proteins and homologs of metal-dependent DUBs of the Jab1/MPN family have been detected in several bacteria in association with Ub-like (Ubl) and E1-like proteins, leading to the hypothesis that these proteins could give rise to the Ub-system of eukaryotes; in contrast, E3 enzymes appear to be specific to eukaryotes [1, 7]. Indeed, there are some steps of thiamine and Mo/W cofactor biosynthesis that are biochemically equivalent to Ub conjugation. These steps include incorporation of sulfur into the respective molecules mediated by the Ubl sulfur-carrier proteins of the ThiS or MoaD family. These Ubl proteins are activated by adenylating E1-like enzymes of the ThiF and MoeB families, and in the next step, sulfur is incorporated by sulfur transferases of the IscS or rhodanese family, that transfer sulfur to its target via an intermediate persulfide (-S-S-H) formed by the active site cysteine [1, 7, 9–13].

The eukaryote Ub proteins and the prokaryote ThiS/MoaD family proteins possess the same β-grasp fold [14, 15] and a conserved carboxyl-terminal glycine which is crucial for the activation by E1-like enzymes [9, 10, 12, 13]. Recently, a protein modification system, known as pupylation, that is functionally equivalent but not homologous to the Ub system has been discovered in *Mycobacterium tuberculosis* [16, 17]. The two key components of this system are the small protein Pup and the enzyme PafA that is essential for Pup conjugation to the ε-NH_2 groups of lysines on several target proteins [16, 17]. The pupylated proteins are targeted for degradation by the mycobacterial proteasome [18]. Until recently, there were no indications that in archaea Ubl proteins perform functions other than cofactor biosynthesis, especially given that no archaeal E2-like proteins have been detected [7, 8]. Furthermore, there were some doubts that ThiS-like proteins in archaea are actually involved in thiamine biosynthesis because, unlike the bacterial case, the respective genes do not belong in the same gene neighborhoods with other thiamine biosynthesis genes, and an alternative pathway for thiamine biosynthesis has been proposed to function in archaea and eukaryotes [7, 19, 20].

In a striking recent development, the involvement of two Ubl proteins called SAMPs (small archaeal modifier proteins) in protein conjugation has been demonstrated in the halobacterium *Haloferax volcanii* [21]. Because SAMPylated proteins seem to accumulate in proteasome-deficient mutants and the targets of SAMPylation include ubiquitous metabolic and house-keeping systems of archaea, Humbard et al. hypothesized that the eukaryotic Ub system evolved from the SAMPylation machinery or a related archaeal system [21]. These groundbreaking results prompted us to perform an in-depth comparative genomic and sequence analysis of archaeal Ubl proteins and associated gene products; this analysis led to a number of functional predictions and a shift of the perspective on the likely ancestral functions of Ub-like proteins.

2. Materials and Methods

The recent update of the arCOG database [22] that includes 70 complete archaeal genomes (ftp://ftp.ncbi.nih.gov/pub/wolf/COGs/arCOG/) was used for the analysis of phyletic patterns of the relevant genes. The same database was also used for sequence retrieval. The NCBI Refseq database [23] was used for retrieval of information on genomic context. Protein sequence database searches were performed using PSI-BLAST [24] with an inclusion threshold *E*-value of 0.01 and no composition-based statistical correction. Additional sequence database searches were performed using the HHPred program which includes secondary structure prediction as part of the search [25]. The PSI-BLAST and HHPred searches allow prediction of protein fold through similarity to proteins of known structure.

Multiple alignments of protein sequences were constructed using the Promals3D program [26], followed by a minimal manual correction on the basis of local alignments obtained using PSI-BLAST [24]. Protein secondary structure was predicted using the PSIPRED program that constructs multiple alignments of the query proteins with their homologs (whenever available) and employs these alignments for prediction [27]. Maximum likelihood (ML) phylogenetic trees were constructed by using MOLPHY program [28] with the JTT substitution matrix to perform local rearrangement of an original Fitch tree [29]. The MOLPHY program was also used to compute RELL bootstrap values from 10,000 replicates.

3. Results and Discussion

3.1. Ubl Proteins in Archaea and Their Classification. For the purpose of this paper, we define Ubl proteins broadly and in functional terms, rather than in terms of homology, that is, as small proteins that function as sulfur carriers in coenzyme biosynthesis and other metabolic reactions or that modify other proteins through conjugation that includes isopeptide bond formation. So defined, the Ubl proteins include the Ub homologs that adopt the β-grasp fold, the Pup-like proteins, and the additional proteins that are inferred to function via a similar mechanism on the basis of gene fusions, genomic neighborhoods and distinct sequence motifs (see below).

In order to identify potential Ubl proteins in archaea as completely as possible, we employed two approaches. First, we performed PSI-BLAST searches against the archaeal subset of the NR database using as queries representatives of all previously identified Ubl protein families [1, 7, 8]. All proteins identified by these searches were linked to the updated arCOG database (see [22] and Section 2). The list of arCOGs that encompass potential Ubl proteins is given in Supplementary Table S1 available at doi:10.1155/2010/710303. This search allowed us to detect a few missing members of Ubl protein families, including a ThiS-like protein (NEQ520) in *Nanoarchaeum equitans*, an organism that was not previously noticed to encode Ubl proteins. The second approach was based on the identification of C-terminal motifs in multiple alignments of arCOGs. It has been shown that Ubl proteins (both β-grasp proteins and Pup-related proteins) possess a functionally essential double glycine (GG) motif at the C-terminus [1, 7, 8, 21]. Additionally, we noticed that one of the β-grasp related arCOGs from Halobacteria (arCOG00539) contains a double cysteine (CC) C-terminal motif. So we reconstructed consensus sequences for multiple alignments of all arCOGs and searched families that consisted of small proteins (<200 aa) with a conserved GG or CC C-terminal motif. Altogether we identified 8 arCOGs that met these criteria: 6 of which belong to the β-grasp fold, the 7th one (arCOG06308) possesses a TATA-binding protein-(TBP-) like fold (these proteins contain a C-terminal GG motif and are unique to Halobacteria), and the 8th one is an uncharacterized family (arCOG08988) with a "CC" C-terminal motif that is also specific to Halobacteria. The proteins in the latter family are predicted to possess a pattern of secondary structure elements (helix-helix-β-strand) that is clearly distinct from the β-grasp fold or the TBP-like fold but resembles the Pup domain [7, 8]. The phyletic patterns of all these arCOGs show that, among Archaea, Ubl proteins (primarily, the β-grasp domain proteins) are missing only from the genomes of several methanogens,

namely, *Methanococcus jannaschii*, *Methanopyrus kandleri* and *Methanococcus aeolicus*.

We analyzed all arCOGs that include β-grasp fold Ubl proteins by constructing a multiple alignment (Supplementary Figure S1) and a phylogenetic tree (Figure 1: The maximum likelihood tree was reconstructed using MOLPHY program [28] from 76 informative positions in the multiple alignment. The RELL bootstrap values are indicated for selected major branches: the branches supported at $\geq$50% are marked by black circles. The sequences are denoted by their GI numbers, abbreviated species name, and arCOG number to which this sequence has been assigned in arCOG database. Color codes for sequences are given as follows: blue—euryarchaea; orange—crenarchaea; brown—thaumarchaea; pink—korarchaea; black—*Nanoarchaeum equitans*. Major haloarchaeal branches are shaded. Proteins analyzed in the recent study of SAMPylation [21] are denoted by *Haloferax volcanii* protein identifiers and colored red. For the MoaD subtree, the expected associations with one or more MoCo biosynthesis genes are shown by green circles. Other gene neighbors are indicated on the right side of the tree (red) by indication of gene name, by full protein name, or by arCOG. Genes associated with Ubl are the following: E1-Ubl activating enzyme, ThiF/HesA family; AOR, tungsten cofactor containing enzyme aldehyde ferredoxin oxidoreductase; SseA, Rhodanese-related sulfurtransferase; GloB, glyoxalase; SfsA, sugar fermentation stimulation protein; OcmC, peroxiredoxin.). In this case, a highly reliable tree topology could not be obtained owing to the small size of the Ubl proteins resulting in a small number of informative positions. This caveat notwithstanding, the tree consisted of the two major previously established branches that correspond, respectively, to the ThiS and MoaD families [7]; moreover, the topology is reasonably compatible with the archaeal taxonomy and with the classification of the Ubl protein derived from the arCOGs (Figure 1). Therefore, this tree provides a useful framework for classification and potential functional inferences. The MoaD branch includes almost twice as many proteins as the ThiS branch. Several lineage-specific duplications are traceable in the MoaD branch including Crenarchaea- and Halobacteria-specific duplications. Several cases of likely horizontal gene transfer are also noticeable, for example, several euryarchaeal branches within the crenarchaeal part of the MoaD branch and, conversely, some crenarchaea embedded within the euryarchaeal part of the ThiS branch. The proteins in arCOG00540 that is specific to Sulfolobales, which so far have not been annotated as Ubl proteins, and those in arCOG00537 that is specific to Thermoproteales appear to cluster within the ThiS branch, pointing to additional duplications in crenarchaea. The tree also reveals a probable error in arCOG assignments for Thaumarchaea because two Thaumarchaeal proteins (GI: 161528937 and GI: 118195088) belong to arCOG00535 rather than arCOG00536. Given the diversity within both branches in the Ubl protein tree, it seems most likely that the last archaeal common ancestor (LACA) encoded at least two Ubl proteins with the β-grasp fold that represented the ThiS and MoaD families.

3.2. Gene Context and Domain Fusion Analysis for Ubl Proteins. Gene context and domain fusion analysis are central tools of inference under the "guilt by association" approach that is broadly used for prediction of functional connections for uncharacterized genes [30–33]. Most domain fusions can be automatically retrieved from arCOGs because the algorithm of arCOG construction includes splitting proteins into domains unless a fusion is conserved to the extent that it dominates the corresponding arCOG [22]. To analyze neighborhoods we retrieved three upstream and three downstream genes for each Ubl gene from a representative set of archaeal genomes (Supplementary Table S2) and identified the most common gene associations (Figure 1, Table 1, and Supplementary Table S3). Generally, we observed the same trends that have been pointed out previously [7, 20]. Most of the genes from the MoaD subfamily in archaea are associated with MoCo biosynthesis enzymes and the gene for aldehyde ferredoxin oxidoreductase (AOR) which utilizes the tungsten cofactor (a derivative of the molybdopterin cofactor). Like in bacteria, many MoaD-family domains are fused to the MoaE enzyme which is responsible for sulfur transfer to activated MoaD-like protein. We also confirmed the absence of contextual association of ThiS genes with any of the genes for thiamine cofactor biosynthesis.

In addition, we identified several strong connections that have not been noticed previously, partly, because recently sequenced genomes help us to ascertain the evolutionary conservation of these associations. Mostly, these new associations are links between ThiS family genes and genes for proteins involved in translation. The most notable case is the association with PP-loop family ATPases that catalyze various tRNA modifications. In particular, the connection with the MesJ protein (arCOG0042) recurs in several archaeal lineages (Figure 1). The MesJ protein is nearly ubiquitous in prokaryotes and, in bacteria, is responsible for lysidine formation [35].

Recently, a tRNA modification pathway in yeast and in the nematode *Caenorhabditis elegans* that includes the Ubl protein URM1, two PP-loop ATPases (Nsc6p and Ncs2p), and two additional enzymes whose orthologs in bacteria are involved in thiamine biosynthesis (E1-like protein and rhodanese) has been characterized [36–38]. It has been shown that URM1 acts as a sulfur carrier protein for thiolation of uridine in the wobble position of some tRNAs; this modification results in an increased translational fidelity, in particular, preventing frame shift errors [37, 39]. Strikingly, three proteins that are homologous to URM1 pathway components (HVO_0558, arCOG01676; HVO_0025, arCOG02019; HVO_0580, arCOG00042) are SAMPylated with both SAMP1 and SAMP2 in *H. volcanii* [21]. The HVO_0580 protein, which is the ortholog of Nsc6p and a member of arCOG00042, is SAMPylated only with SAMP2 (HVO_0202), a ThiS family protein. Our observations complement these results and suggest that, even in those archaea where there is no genomic association between Ubl and PP-loop ATPases of arCOG00042 genes (which is the case in Halobacteria), these proteins function in concert.

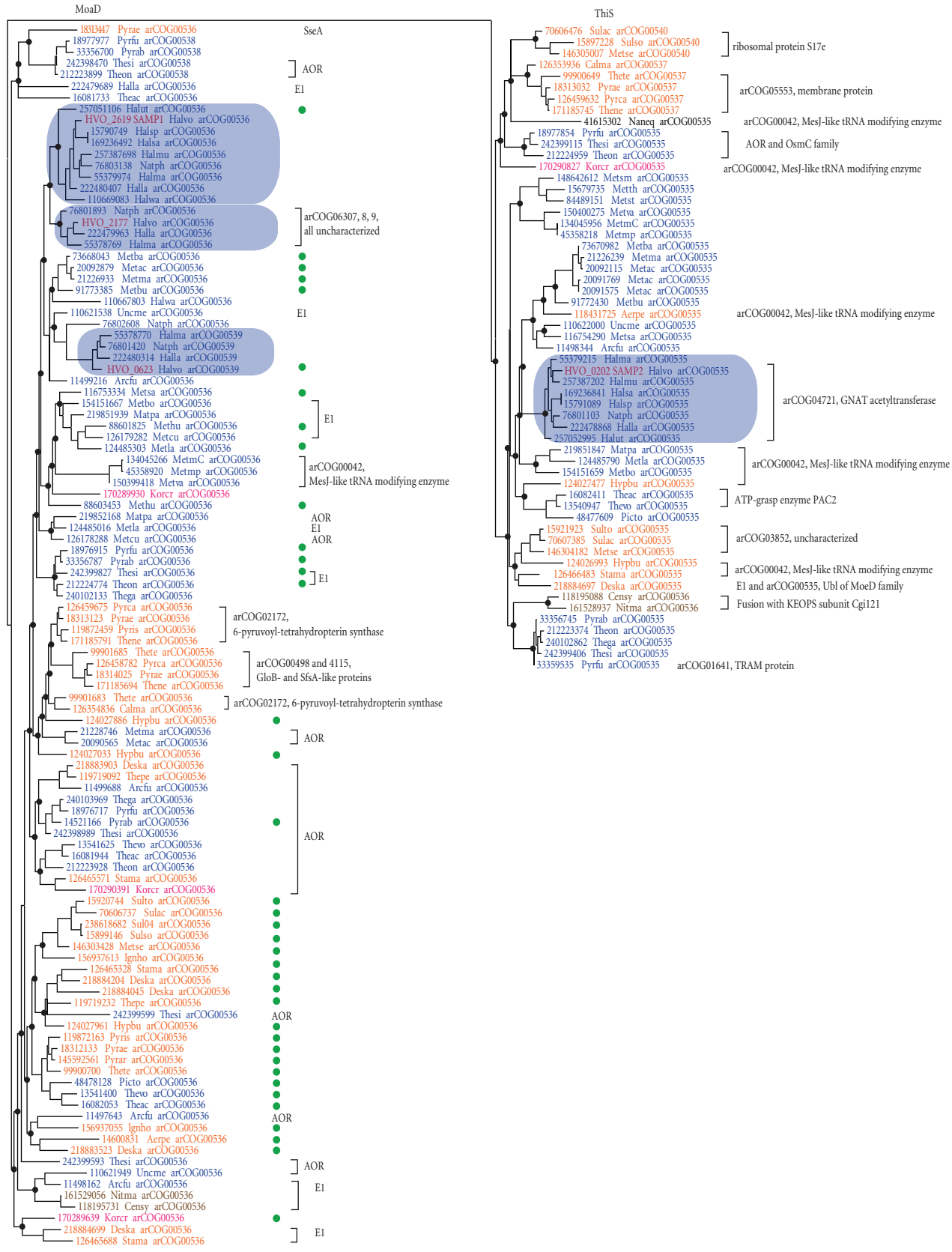


Figure 1: The phylogenetic tree and gene associations for the archaeal Ubl proteins of the β-grasp fold.

Table 1: Components of Ubl-related pathways in archaea predicted in this paper[a].

Reference arCOG	Representative of the reference arCOG	arCOGs identified by genomic context analysis	Lineage	Comment
Genomic associations of genes for UBL (β-grasp) proteins				
arCOG00540	70606476	arCOG01885 arCOG07188 arCOG00286	Sulfolobales	Predicted operon contains genes for 30 S ribosomal protein S17e, small uncharacterized protein, and a distinct membrane-associated HerA-like ATPase of the SSO0283 family. Thermosome beta subunit is a divergently encoded gene located within a conserved region which includes a variety of informational genes.
arCOG00537	18313032	arCOG05553	Thermoproteales	Membrane protein
arCOG00535	15791089	arCOG04721	Halobacteria	GNAT N-acetyltransferase
arCOG00535	15679735	arCOG00042	Many diverse archaea	tRNA(Ile)-lysidine synthase MesJ
arCOG00535	70607385	arCOG03852	Sulfolobales	Uncharacterized protein
Genomic associations of arCOG06308 (TBP-like fold protein with [GG] C-terminal motif)				
arCOG06308	76801892	arCOG06307 arCOG06309 arCOG00536	Halobacteria	Two small uncharacterized proteins and MoaD-like Ubl protein
Genomic associations of arCOG08988 (uncharacterized protein with [CC] C-terminal motif)				
arCOG08988	257373014	arCOG04404	Halobacteria	ComK-like protein, in bacteria, is involved in regulation of competence; in archaea, its role is unclear [34].
Genomic associations of genes for E1-like enzymes (ThiF/MoeB)				
arCOG01677	15679158	arCOG04863 arCOG04865 arCOG04864 arCOG04454	Methanococcales	First two genes are distant CinA C-terminal domain homologs; next is an NAD-binding domain containing protein and an NIF3 homolog.
Genomic associations of genes for the Jab protease				
arCOG01139	257387955	arCOG01222	Many diverse archaea	Cytidylyltransferase family protein

[a]The table lists genes (arCOGs) that are consistently found within genomic neighborhoods of genes that encode components of Ubl-related pathways. On the basis of theses associations and, in some cases, their domain content as well, the protein products of these neighboring genes are predicted to be functionally related to Ubl systems as well.

In Thermococcales, several Ubl genes are associated with genes encoding peroxiredoxins of the OcmC family (Figure 1), and indeed, a highly similar homolog of these proteins accumulates in proteasome mutants and is SAMPylated in *H. volcanii* [21, 40].

Several representatives of Sulfolobales encode a distinct family of Ubl proteins (arCOG00540) that are most similar to the eukaryotic URM1 family (Supplementary Figure S2) and therefore can be predicted to be involved in a URM1-like pathway. These *Sulfolobus* proteins are encoded in a distinct neighborhood which also includes genes for the ribosomal protein S17, an uncharacterized small protein of arCOG07188, a distinct membrane-associated HerA-like ATPase of the SSO0283 family [41], and a gene for an HSP60 family chaperonin, a thermosome subunit [42], which is transcribed in the opposite direction compared to the rest of the above genes (Table 1). Considering the data on SAMPylation of proteins encoded by genes adjacent to Ubl genes, it seems likely that the URM1 homologs in Sulfolobales regulate translation, proteolysis, and/or cell division through SAMPylation of, respectively, S17, HSP60, or HerA proteins, in addition to or instead of functioning in tRNA modification.

Another notable observation is the fusion of a Ubl domain with the KEOPS complex subunit Cgi121. This fusion is conserved in all available genomes of Thaumarchaea (formerly known as mesophilic Crenarchaea [43]. The KEOPS (kinase, endopeptidase, and other peptides of small size) complex consists of 5 subunits (the names are those of the respective yeast genes that have been studied in most detail): Mn2+-dependent serine/threonine protein kinase Bud32p, ATPase of the ASKHA family (Kae1p), and three additional subunits: Pcc1p, Gon7p, and Cgi121p whose functions remain unclear. KEOPS complex has been shown

to be involved in telomere maintenance and transcription in yeast [44–47]. The orthologs of the Kae1 and Bud32p subunits are present in all Archaea, the Pcc1p ortholog is missing only in a few archaeal genomes, and the Cgi121p ortholog is absent in Sulfolobales/Desulfurococcales and Nanoarchaeon. Taken together, comparative-genomic findings suggest that the counterpart of the KEOPS complex performs an essential function in archaea. The structure of this complex has been solved but the details of its functioning are still scarce although there are indications that it is critical for the maintenance of genome integrity in archaea [45–47]. The gene for the Pcc1 subunit shows a strong genomic association with genes that encode subunits of the archaeal exosome, the RNA degradation machine [48, 49]. Furthermore, the exosome genes themselves are associated with genes for proteasome subunits suggesting that RNA and protein degradation in archaea are tightly coordinated [48]. Very recently, it has been shown that in bacteria homologs of the KEOPS complex subunits are required for a distinct, widespread tRNA modification, the formation of N6-threonylcarbamoyladenosine (t6A) [50]. These findings suggest the possibility of regulation of the KEOPS complex by SAMPylation or coordinated functioning of the KEOPS complex, along with the Ubl-based system, proteasome, and exosome, in RNA and protein turnover control in archaea. Interestingly, the gene for the Cgi121-Ubl fusion protein is apparently cotranscribed with a gene for the ribosomal protein S17 in *Nitrosopumilus maritimus* and some other unfinished genomes of marine Thaumarchaeota, resembling the gene neighborhood in Sulfolobales described above.

The emerging trend of the association of Ubl proteins with genes involved in key information processing function in archaea suggests that several less frequent associations seen in a variety of different genomes also merit attention. For example, in two Thermoplasma genomes, the genes for ThiS family proteins are associated with the gene for the proteasome assembly chaperone PAC2 (Figure 1). In Pyrococci, ThiS family genes are associated with RNA-binding TRAM domain (Figure 1). Proteins containing TRAM domains are common in archaea; in particular, it is notable that a TRAM domain is fused to the essential enzyme 2-methylthioadenine synthetase that is involved in the thiolation of both tRNA and ribosomal proteins in bacteria [51–53]. In this case, again, the Ubl protein might possess a dual function: it could be involved in thiolation of tRNA (and/or ribosomal proteins) as a sulfur carrier or could regulate this process by SAMPylation or both. Finally, the only Ubl protein in Nanoarchaeon is located in the neighborhood of several informational genes including the proteasome alpha subunit and tRNA modification enzymes (Supplementary Table S2).

Surprisingly, it appears that either the functional specificity of Ubl proteins from different subfamilies can be easily switched or functional flexibility is an intrinsic feature of these proteins. For instance, the two functionally characterized SAMP proteins of *H. volcanii* belong to the two distinct branches of archaeal Ubl proteins, ThiS and MoaD (Figure 1). This hypothesis seems to be further supported by gene context and the dendrogram analysis, in particular, the association of Ubl proteins of the MoeB family with tRNA-modifying PP-loop ATPases and association of the ThiS family genes with the AOR enzyme (Figure 1).

3.3. Gene Context and Domain Fusions of E1-Like Enzymes. All known pathways involving Ubl proteins require E1 enzymes which activate these proteins via adenylation of the carboxy-terminal glycine residue of the Ub/Ubl polypeptide [54]. E1 enzymes possess a core Rossmann-fold ATP-binding domain [5]. Four distinct families of E1-like enzymes have been identified in archaea, namely, MoeB/ThiF/MOSC3 like, MJ0639-like, PaaA-like, and GodD-like enzymes [5] which in arCOGs are assigned to arCOG1676, arCOG1677, arCOG4786, and arCOG02882-2883,5002, respectively. However, PaaA and GodD-like enzymes are probably not involved in pathways that rely on Ubl proteins [5] and therefore are not considered here. Representatives of arCOG1676 are present in most archaea with the exception of the same methanogens that lack Ubl proteins (see above). However, all these methanogens encode a representative of the closely related arCOG1677 (Supplementary Table S1). The reconstructed phylogeny of arCOG1676 shows that the major euryarchaeal branch is well separated from the major crenarchaeal branch (Supplementary Figure S3). Some euryarchaea seem to have acquired from different bacterial sources additional E1-like enzymes; in Thermoplasma, these enzymes apparently have replaced the ancestral form.

Most of the archaeal E1-like enzymes possess the same domain architecture (E1 core and a TBP-like C-terminal domain) as most of the bacterial homologs. There are also several other telling fusions shared with bacteria: Ubl-E1-TBP in Thaumarchaeota and Jab-E1 in methanogen RC1 (Jab is a predicted protease and/or DUB—see below). In addition, a unique architecture, with a small C-terminal small domain containing two conserved cysteines, is seen in *Sulfolobus* genomes. Analysis of gene neighborhoods for arCOG01676 did not reveal any new strong functional links. We detected many associations with Ubl-like genes and fewer links with enzymes of MoCo biosynthesis, thiamine biosynthesis enzyme ThiI, and cysteine synthase, all of which have been described before (see [5] and Supplementary Tables S2 and S3). However, it should be emphasized that the essential function of ThiI-like enzymes in prokaryotes is 4-thiouridine (S4U) modification of tRNAs [55], so it seems plausible that in archaea, which apparently synthesize thiamine via a distinct pathway [7, 19, 20], tRNA modification is the only function of ThiI. Furthermore, recently it has been shown that E1 enzymes and Ubl-proteins are also involved in thiolation of tRNA in Thermus thermophilus [56]. Thus, the same function can be proposed for at least some of the E1-MoaD associations seen in archaea.

Interestingly, several representatives of the second E1-like family (arCOG01677) in methanogens are located in a conserved neighborhood which includes a gene for PP-loop superfamily enzyme, a predicted subunit of tRNA(5-methylaminomethyl-2-thiouridylate) methyltransferase (arCOG00037) [57]. However, the strongest potential functional association of arCOG1677 family genes remains enigmatic. In most methanogens, these genes are associated

with genes for arCOG04865, which is homologous to the C-terminal domain of CinA, and arCOG04454, a NIF3 homolog (Table 1). In bacteria, CinA is a competence-induced gene often located in the same operon with RecA [58]. The NIF3 gene encodes a conserved metal-binding regulatory protein whose exact function remains unknown [59]. Given that arCOG01677 genes are never associated with genes for Ubl proteins, it seems unlikely that this group of E1-like enzymes is functionally linked to Ubl-dependent pathways.

3.4. Gene Context and Domain Fusions of Jab Proteases and Rhodanese-Like Enzymes. Metal-dependent proteases of the Jab family that in eukaryotes function as the primary proteasome-associated DUBs [4, 60, 61] and rhodanese-related enzymes that are involved in sulfur transfer reactions together with Ubl proteins [62] show similar but not identical distributions in archaea (Supplementary Table S1). These proteins are missing in many crenarchaea and methanogens. In archaea, the homologs of Jab proteases are rarely associated with Ubl genes or other genes involved in Ubl-related pathways. However, Jab genes are often associated with a gene for a cytidylyltransferase (Table 1 and Supplementary Tables S2 and S3), an association that could be of particular interest given that E2 and E3 enzymes required for Ub conjugation in eukaryotes have not been detected in archaea [6, 7]. A nucleotidyltransferase potentially could transfer an adenylated (activated) Ubl to a target protein, that is, perform the function of Ub ligase without sulfur-containing intermediates. The Jab protease is likely to function as a DUB similarly to its homologs in eukaryotes. Thus, it is tempting to propose the cytidylyltransferase-Jab tandem of enzymes as a candidate for an archaeal Ubl-conjugation/deubiqiutination system.

Sulfur transferases of the rhodanese family catalyze the incorporation of sulfur into activated Ubl proteins via an intermediate persulfide. Rhodanese domains are often fused to ThiI like enzymes that also contain an N-terminal RNA-binding THUMP domain (Supplementary Tables S2 and S3). Many bacteria posses the same domain architecture and, as pointed out above, these enzymes are probably involved in tRNA modification. Only a few other associations of rhodanese-like proteins could be related to Ubl pathways (with AOR genes, for example), but most of other proteins of the rhodanese family are involved in either sulfur metabolism or redox pathways, which are likely Ubl independent.

4. Discussion

Comparative-genomic analysis indicates that most archaea encode members of two major groups of Ubl proteins with the β-grasp fold, the ThiS and MoaD families. The ThiS family genes are rarely found together with genes for thiamine and Mo/W cofactor metabolism enzymes but instead are often associated with various highly conserved and probably essential genes with functions related to translation, especially, tRNA modification. Thus, most if not all ThiS family proteins are predicted to function as sulfur carrier proteins for reactions similar to those recently characterized for the URM1 pathway in yeast [37]. In contrast, genomic associations suggest that the primary function of the MoaD family proteins is indeed the Mo/W cofactor biosynthesis. The absence of Ubl proteins and E1-like Ubl-activating enzymes of the arCOG1676 in such autotrophic archaea as *M. jannaschii* and *M. kandleri* and the absence of association of Ubl genes with thiamine biosynthesis genes (other than ThiI family enzymes which are probably involved in tRNA modification) is compatible with the existence of an alternative thiamine biosynthesis pathway in archaea.

Surprisingly, despite their apparent functional preferences, ThiS and MoaD family members appear to be interchangeable in pathways that employ Ubl proteins either as sulfur carriers or for protein modification. This possibility is born out both through analysis of gene associations for both subfamilies as described here and by the experimental data on the two SAMP proteins of *Haloferax volcanii* one of which belongs to the ThiS family and the other one to the MoaD family [21].

The most prominent associations revealed by comparative genomics for the archaeal Ubl proteins are with enzymes of tRNA modification. This finding leads to the hypothesis that the majority of the β-grasp Ubl proteins in archaea, at least those of the ThiS family, are involved in sulfur insertion steps of the biosynthesis of modified nucleotides. Given the ubiquity of a variety of tRNA modifications across cellular life [63], this is likely to be the ancestral function of the Ubl proteins that subsequently were recruited for other chemically similar reactions, such as MoCo and thiamine biosynthesis, as well as protein modification. This hypothesis is compatible with the role of the eukaryotic Urm1 protein in specific tRNA modification and with fusion of the Ubl domain to the KEOPS complex subunit Cgi121, given the requirement of KEOPS for the t6A modification. Experimental study of the involvement of Ubl proteins in tRNA modification appears to be an extremely promising research direction.

From a more general perspective, tRNA modification is undoubtedly a major mechanism of the quality control of translation [64, 65]. Considering also the association of another KEOPS subunit (Pcc1) with the exosome and the proteasome, it is tempting to view the Ubl proteins as general devices for protein quality control, both at the most fundamental level of translation fidelity and at the secondary levels of regulated protein and RNA degradation. In eukaryotes, the latter mechanisms assumed hugely diversified roles which required the evolution of the enormously complex Ub-centered signaling systems.

The comparative-genomic analysis of the genes for Ubl proteins and the enzymes that appear functionally linked to them suggests that archaea might possess still uncharacterized Ubl-related functional systems. In particular, the association of the Jab protease with a cytidylyltransferase-like enzyme appears to be a candidate for a Ubl conjugation/deubiquitination system. In addition, archaea are likely to possess functional analogs of Ubl proteins that are structurally and hence evolutionarily unrelated to the β-grasp fold. This group includes small proteins of the TBP-like fold

that bend at a GG doublet and are often fused to E1 family enzymes, in a strong indication of their Ubl-type activity, along with putative homologs of the bacterial Pup protein.

In conclusion, the comparative-genomic analysis triggered by the seminal discovery of the SAMPylation reactions in *H. volcanii* reveals unexpected potential complexity of archaeal Ubl-centered systems and offers several directions for further experimentation, the most important of which arguably is the validation of the hypothesis on the involvement of Ubl proteins in tRNA modification. In addition, this analysis opens up an unexpected and potentially fundamental area of inquiry into the evolution of cells, namely, the ancestral connection between systems of protein quality control that operate at different levels.

Species Abbreviations

Aerpe: *Aeropyrum pernix* K1
Arcfu: *Archaeoglobus fulgidus*
Calma: *Caldivirga maquilingensis* IC-167
Korar: *Candidatus Korarchaeum cryptofilum* OPF8
Metbo: *Candidatus Methanoregula boonei* 6A8
Censy: *Cenarchaeum symbiosum*
Deska: *Desulfurococcus kamchatkensis* 1221n
Ferac: *Ferroplasma acidarmanus* fer1
Halma: *Haloarcula marismortui* ATCC 43049
Halsa: *Halobacterium salinarum* R1
Halsp: *Halobacterium sp.*
Halmu: *Halomicrobium mukohataei* DSM 12286
Halwa: *Haloquadratum walsbyi*
Halut: *Halorhabdus utahensis* DSM 12940
Halla: *Halorubrum lacusprofundi* ATCC 49239
Haltu: *Haloterrigena turkmenica* DSM 5511
Hypbu: *Hyperthermus butylicus* DSM 5456
Ignho: *Ignicoccus hospitalis* KIN4/I
Metse: *Metallosphaera sedula* DSM 5348
Metsm: *Methanobrevibacter smithii* ATCC 35061
Metin: *Methanocaldococcus infernus* ME
Metvu: *Methanocaldococcus vulcanius* M7
Metbu: *Methanococcoides burtonii* DSM 6242
Metmp: *Methanococcus maripaludis* S2
Metva: *Methanococcus vannielii* SB
Metla: *Methanocorpusculum labreanum* Z
Metcu: *Methanoculleus marisnigri* JR1
Metsa: *Methanosaeta thermophila* PT
Metac: *Methanosarcina acetivorans*
Metba: *Methanosarcina barkeri str. Fusaro*
Metma: *Methanosarcina mazei*
Metst: *Methanosphaera stadtmanae*
Matpa: *Methanosphaerula palustris* E1-9c
Methu: *Methanospirillum hungatei* JF-1
Metth: *Methanothermobacter thermautotrophicus*
Naneq: *Nanoarchaeum equitans*
Natph: *Natronomonas pharaonis*
Nitma: *Nitrosopumilus maritimus* SCM1
Picto: *Picrophilus torridus* DSM 9790
Pyrae: *Pyrobaculum aerophilum*
Pyrar: *Pyrobaculum arsenaticum* DSM 13514
Pyrca: *Pyrobaculum calidifontis* JCM 11548
Pyris: *Pyrobaculum islandicum* DSM 4184
Pyrab: *Pyrococcus abyssi*
Pyrfu: *Pyrococcus furiosus*
Pyrho: *Pyrococcus horikoshii*
Stama: *Staphylothermus marinus* F1
Sulac: *Sulfolobus acidocaldarius* DSM 639
Sulso: *Sulfolobus solfataricus* P2
Sulto: *Sulfolobus tokodaii* str. 7
Thega: *Thermococcus gammatolerans* EJ3
Theko: *Thermococcus kodakarensis* KOD1
Theon: *Thermococcus onnurineus* NA1
Thesi: *Thermococcus sibiricus* MM 739
Thepe: *Thermofilum pendens* Hrk 5
Theac: *Thermoplasma acidophilum*
Thevo: *Thermoplasma volcanium*
Thene: *Thermoproteus neutrophilus* V24Sta
Thete: *Thermoproteus tenax*
Uncme: Uncultured methanogenic archaeon.

Acknowledgments

The authors thank Valerie de Crecy-Lagard and Julie Maupin-Furlow for helpful discussions. The authors' research is supported by the intramural funds of the Department of Health and Human Services of the USA (National Library of Medicine).

References

[1] M. Hochstrasser, "Origin and function of ubiquitin-like proteins," *Nature*, vol. 458, no. 7237, pp. 422–429, 2009.
[2] O. Kerscher, R. Felberbaum, and M. Hochstrasser, "Modification of proteins by ubiquitin and ubiquitin-like proteins," *Annual Review of Cell and Developmental Biology*, vol. 22, pp. 159–180, 2006.
[3] A. M. Weissman, "Themes and variations on ubiquitylation," *Nature Reviews Molecular Cell Biology*, vol. 2, no. 3, pp. 169–178, 2001.
[4] L. M. Iyer, E. V. Koonin, and L. Aravind, "Novel predicted peptidases with a potential role in the ubiquitin signaling pathway," *Cell Cycle*, vol. 3, no. 11, pp. 1440–1450, 2004.
[5] A. M. Burroughs, L. M. Iyer, and L. Aravind, "Natural history of the E1-like superfamily: implication for adenylation, sulfur transfer, and ubiquitin conjugation," *Proteins: Structure, Function and Bioformatics*, vol. 75, no. 4, pp. 895–910, 2009.
[6] A. M. Burroughs, M. Jaffee, L. M. Iyer, and L. Aravind, "Anatomy of the E2 ligase fold: implications for enzymology and evolution of ubiquitin/Ub-like protein conjugation," *Journal of Structural Biology*, vol. 162, no. 2, pp. 205–218, 2008.
[7] L. M. Iyer, A. M. Burroughs, and L. Aravind, "The prokaryotic antecedents of the ubiquitin-signaling system and the early evolution of ubiquitin-like β-grasp domains," *Genome Biology*, vol. 7, no. 7, article no. R60, 2006.
[8] L. M. Iyer, A. M. Burroughs, and L. Aravind, "Unraveling the biochemistry and provenance of pupylation: a prokaryotic analog of ubiquitination," *Biology Direct*, vol. 3, article no. 45, 2008.
[9] K. Furukawa, N. Mizushima, T. Noda, and Y. Ohsumi, "A protein conjugation system in yeast with homology to biosynthetic enzyme reaction of prokaryotes," *Journal of Biological Chemistry*, vol. 275, no. 11, pp. 7462–7465, 2000.

[10] A. S. Goehring, D. M. Rivers, and G. F. Sprague Jr., "Attachment of the ubiquitin-related protein Urm1p to the antioxidant protein Ahp1p," *Eukaryotic Cell*, vol. 2, no. 5, pp. 930–936, 2003.

[11] D. Kessler, "Enzymatic activation of sulfur for incorporation into biomolecules in prokaryotes," *FEMS Microbiology Reviews*, vol. 30, no. 6, pp. 825–840, 2006.

[12] C. Lehmann, T. P. Begley, and S. E. Ealick, "Structure of the Escherichia coli ThiS-ThiF complex, a key component of the sulfur transfer system in thiamin biosynthesis," *Biochemistry*, vol. 45, no. 1, pp. 11–19, 2006.

[13] S. Leimkühler, M. M. Wuebbens, and K. V. Rajagopalan, "Characterization of Escherichia coli MoeB and its involvement in the activation of molybdopterin synthase for the biosynthesis of the molybdenum cofactor," *Journal of Biological Chemistry*, vol. 276, no. 37, pp. 34695–34701, 2001.

[14] M. J. Rudolph, M. M. Wuebbens, K. V. Rajagopalan, and H. Schindelin, "Crystal structure of molybdopterin synthase and its evolutionary relationship to ubiquitin activation," *Nature Structural Biology*, vol. 8, no. 1, pp. 42–46, 2001.

[15] C. Wang, J. Xi, T. P. Begley, and L. K. Nicholson, "Solution structure of ThiS and implications for the evolutionary roots of ubiquitin," *Nature Structural Biology*, vol. 8, no. 1, pp. 47–51, 2001.

[16] R. A. Festa, M. J. Pearce, and K. H. Darwin, "Characterization of the proteasome accessory factor (paf) operon in Mycobacterium tuberculosis," *Journal of Bacteriology*, vol. 189, no. 8, pp. 3044–3050, 2007.

[17] M. J. Pearce, J. Mintseris, J. Ferreyra, S. P. Gygi, and K. H. Darwin, "Ubiquitin-like protein involved in the proteasome pathway of Mycobacterium tuberculosis," *Science*, vol. 322, no. 5904, pp. 1104–1107, 2008.

[18] M. J. Pearce, P. Arora, R. A. Festa, S. M. Butler-Wu, R. S. Gokhale, and K. H. Darwin, "Identification of substrates of the Mycobacterium tuberculosis proteasome," *EMBO Journal*, vol. 25, no. 22, pp. 5423–5432, 2006.

[19] A. Chatterjee, C. T. Jurgenson, F. C. Schroeder, S. E. Ealick, and T. P. Begley, "Biosynthesis of thiamin thiazole in eukaryotes: conversion of NAD to an advanced intermediate," *Journal of the American Chemical Society*, vol. 129, no. 10, pp. 2914–2922, 2007.

[20] D. A. Rodionov, A. G. Vitreschak, A. A. Mironov, and M. S. Gelfand, "Comparative genomics of thiamin biosynthesis in procaryotes. New genes and regulatory mechanisms," *Journal of Biological Chemistry*, vol. 277, no. 50, pp. 48949–48959, 2002.

[21] M. A. Humbard, H. V. Miranda, J.-M. Lim et al., "Ubiquitin-like small archaeal modifier proteins (SAMPs) in Haloferax volcanii," *Nature*, vol. 463, no. 7277, pp. 54–60, 2010.

[22] K. S. Makarova, A. V. Sorokin, P. S. Novichkov, Y. I. Wolf, and E. V. Koonin, "Clusters of orthologous genes for 41 archaeal genomes and implications for evolutionary genomics of archaea," *Biology Direct*, vol. 2, article no. 33, 2007.

[23] K. D. Pruitt, T. Tatusova, and D. R. Maglott, "NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins," *Nucleic Acids Research*, vol. 35, no. 1, pp. D61–D65, 2007.

[24] S. F. Altschul, T. L. Madden, A. A. Schäffer et al., "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs," *Nucleic Acids Research*, vol. 25, no. 17, pp. 3389–3402, 1997.

[25] A. Hildebrand, M. Remmert, A. Biegert, and J. Söding, "Fast and accurate automatic structure prediction with HHpred," *Proteins: Structure, Function and Bioformatics*, vol. 77, no. S9, pp. 128–132, 2009.

[26] J. Pei, B.-H. Kim, and N. V. Grishin, "PROMALS3D: a tool for multiple protein sequence and structure alignments," *Nucleic Acids Research*, vol. 36, no. 7, pp. 2295–2300, 2008.

[27] L. J. McGuffin, K. Bryson, and D. T. Jones, "The PSIPRED protein structure prediction server," *Bioinformatics*, vol. 16, no. 4, pp. 404–405, 2000.

[28] J. Adachi and M. Hasegawa, *MOLPHY: Programs for Molecular Phylogenetics*, Computer Science Monographs 27, Institute of Statistical Mathematics, Tokyo, Japan, 1992.

[29] J. Felsenstein, "Inferring phylogenies from protein sequences by parsimony, distance, and likelihood methods," *Methods in Enzymology*, vol. 266, pp. 418–427, 1996.

[30] L. Aravind, "Guilt by association: contextual information in genome analysis," *Genome Research*, vol. 10, no. 8, pp. 1074–1077, 2000.

[31] M. Y. Galperin and E. V. Koonin, "Who's your neighbor? New computational approaches for functional genomics," *Nature Biotechnology*, vol. 18, no. 6, pp. 609–613, 2000.

[32] E. V. Koonin and M. Y. Galperin, *Sequence-Evolution-Function: Computational Approaches in Comparative Genomics*, Kluwer Academic Publishers, London, UK, 2003.

[33] L. J. Jensen, M. Kuhn, M. Stark et al., "STRING 8—a global view on proteins and their functional interactions in 630 organisms," *Nucleic Acids Research*, vol. 37, no. 1, pp. D412–D416, 2009.

[34] H. Maamar, A. Raj, and D. Dubnau, "Noise in gene expression determines cell fate in Bacillus subtilis," *Science*, vol. 317, no. 5837, pp. 526–529, 2007.

[35] A. Soma, Y. Ikeuchi, S. Kanemasa et al., "An RNA-modifying enzyme that governs both the codon and amino acid specificities of isoleucine tRNA," *Molecular Cell*, vol. 12, no. 3, pp. 689–698, 2003.

[36] S. Kim, W. Johnson, C. Chen, A. K. Sewell, A. S. Bystrom, and M. Han, "Allele specific suppressors of lin-1(R175Opal) identify functions of MOC-3 and DPH-3 in tRNA modification complexes in caenorhabditis elegans," *Genetics*, vol. 185, no. 4, pp. 1235–1247, 2010.

[37] S. Leidel, P. G. A. Pedrioli, T. Bucher et al., "Ubiquitin-related modifier Urm1 acts as a sulphur carrier in thiolation of eukaryotic transfer RNA," *Nature*, vol. 458, no. 7235, pp. 228–232, 2009.

[38] A. Noma, Y. Sakaguchi, and T. Suzuki, "Mechanistic characterization of the sulfur-relay system for eukaryotic 2-thiouridine biogenesis at tRNA wobble positions," *Nucleic Acids Research*, vol. 37, no. 4, pp. 1335–1352, 2009.

[39] S. S. Ashraf, E. Sochacka, R. Cain, R. Guenther, A. Malkiewicz, and P. F. Agris, "Single atom modification (O → S) of tRNA confers ribosome binding," *RNA*, vol. 5, no. 2, pp. 188–194, 1999.

[40] P. A. Kirkland, C. J. Reuter, and J. A. Maupin-Furlow, "Effect of proteasome inhibitor clasto-lactacystin-β-lactone on the proteome of the haloarchaeon Haloferax volcanii," *Microbiology*, vol. 153, no. 7, pp. 2271–2280, 2007.

[41] L. M. Iyer, K. S. Makarova, E. V. Koonin, and L. Aravind, "Comparative genomics of the FtsK-HerA superfamily of pumping ATPases: implications for the origins of chromosome segregation, cell division and viral capsid packaging," *Nucleic Acids Research*, vol. 32, no. 17, pp. 5260–5279, 2004.

[42] M. Klumpp and W. Baumeister, "The thermosome: archetype of group II chaperonins," *FEBS Letters*, vol. 430, no. 1-2, pp. 73–77, 1998.

[43] C. Brochier-Armanet, B. Boussau, S. Gribaldo, and P. Forterre, "Mesophilic crenarchaeota: proposal for a third archaeal phylum, the Thaumarchaeota," *Nature Reviews Microbiology*, vol. 6, no. 3, pp. 245–252, 2008.

[44] A. Bianchi and D. Shore, "The KEOPS complex: a rosetta stone for telomere regulation?" *Cell*, vol. 124, no. 6, pp. 1125–1128, 2006.

[45] A. Hecker, M. Graille, E. Madec et al., "The universal Kae1 protein and the associated Bud32 kinase (PRPK), a mysterious protein couple probably essential for genome maintenance in Archaea and Eukarya," *Biochemical Society Transactions*, vol. 37, no. 1, pp. 29–35, 2009.

[46] A. Hecker, R. Lopreiato, M. Graille et al., "Structure of the archaeal Kae1/Bud32 fusion protein MJ1130: a model for the eukaryotic EKC/KEOPS subcomplex," *EMBO Journal*, vol. 27, no. 17, pp. 2340–2351, 2008.

[47] D. Y. L. Mao, D. Neculai, M. Downey et al., "Atomic structure of the KEOPS complex: an ancient protein kinase-containing molecular machine," *Molecular Cell*, vol. 32, no. 2, pp. 259–275, 2008.

[48] E. V. Koonin, Y. I. Wolf, and L. Aravind, "Prediction of the archeal exosome and its connections with the proteasome and the translation and transcription machineries by a comparative-genomic approach," *Genome Research*, vol. 11, no. 2, pp. 240–252, 2001.

[49] A. Lebreton, R. Tomecki, A. Dziembowski, and B. Séraphin, "Endonucleolytic RNA cleavage by a eukaryotic exosome," *Nature*, vol. 456, no. 7224, pp. 993–996, 2008.

[50] B. El Yacoubi, H. McGuirk, I. Hatin, D. Iwata-Reuyl, A. G. Murzin, and V. De Crecy-Lagard, "Function of the YrdC/YgjD conserved protein network: the t6A lead," in *Proceedings of the 23rd tRNA Workshop: From the Origin of Life to Biomedicine*, T. Weil and M. Santos, Eds., p. 7, 2010.

[51] V. Anantharaman, E. V. Koonin, and L. Aravind, "TRAM, a predicted RNA-binding domain, common to tRNA uracil methylation and adenine thiolation enzymes," *FEMS Microbiology Letters*, vol. 197, no. 2, pp. 215–221, 2001.

[52] B. P. Anton, L. Saleh, J. S. Benner, E. A. Raleigh, S. Kasif, and R. J. Roberts, "RimO, a MiaB-like enzyme, methylthiolates the universally conserved Asp88 residue of ribosomal protein S12 in Escherichia coli," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 105, no. 6, pp. 1826–1831, 2008.

[53] B. Esberg, H.-C. E. Leung, H.-C. T. Tsui, G. R. Bjork, and M. E. Winkler, "Identification of the miaB gene, involved in methylthiolation of isopentenylated A37 derivatives in the tRNA of Salmonella typhimurium and Escherichia coli," *Journal of Bacteriology*, vol. 181, no. 23, pp. 7256–7265, 1999.

[54] B. A. Schulman and J. W. Harper, "Ubiquitin-like protein activation by E1 enzymes: the apex for downstream signalling pathways," *Nature Reviews Molecular Cell Biology*, vol. 10, no. 5, pp. 319–331, 2009.

[55] E. G. Mueller, C. J. Buck, P. M. Palenchar, L. E. Barnhart, and J. L. Paulson, "Identification of a gene involved in the generation of 4-thiouridine in tRNA," *Nucleic Acids Research*, vol. 26, no. 11, pp. 2606–2610, 1998.

[56] N. Shigi, Y. Sakaguchi, S.-I. Asai, T. Suzuki, and K. Watanabe, "Common thiolation mechanism in the biosynthesis of tRNA thiouridine and sulphur-containing cofactors," *EMBO Journal*, vol. 27, no. 24, pp. 3267–3278, 2008.

[57] T. G. Hagervall, C. G. Edmonds, J. A. McCloskey, and G. R. Björk, "Transfer RNA(5-methylaminomethyl-2-thiouridine)-methyltransferase from Escherichia coli K-12 has two enzymatic activities," *Journal of Biological Chemistry*, vol. 262, no. 18, pp. 8488–8495, 1987.

[58] B. Martin, P. García, M.-P. Castanié, and J.-P. Claverys, "The recA gene of Streptococcus pneumoniae is part of a competence-induced operon and controls lysogenic induction," *Molecular Microbiology*, vol. 15, no. 2, pp. 367–379, 1995.

[59] M. H. Godsey, G. Minasov, L. Shuvalova et al., "The 2.2 Å resolution crystal structure of Bacillus cereus Nif3-family protein YqfO reveals a conserved dimetal-binding motif and a regulatory domain," *Protein Science*, vol. 16, no. 7, pp. 1285–1293, 2007.

[60] G. A. Cope, G. S. B. Suh, L. Aravind et al., "Role of predicted metalloprotease motif of Jab1/Csn5 in cleavage of Nedd8 from Cul1," *Science*, vol. 298, no. 5593, pp. 608–611, 2002.

[61] R. Verma, L. Aravind, R. Oania et al., "Role of Rpn11 metalloprotease in deubiquitination and degradation by the 26S proteasome," *Science*, vol. 298, no. 5593, pp. 611–615, 2002.

[62] R. Cipollone, P. Ascenzi, and P. Visca, "Common themes and variations in the rhodanese superfamily," *IUBMB Life*, vol. 59, no. 2, pp. 51–59, 2007.

[63] H. Grosjean, V. de Crécy-Lagard, and C. Marck, "Deciphering synonymous codons in the three domains of life: co-evolution with specific tRNA modification enzymes," *FEBS Letters*, vol. 584, no. 2, pp. 252–264, 2010.

[64] P. F. Agris, "Decoding the genome: a modified view," *Nucleic Acids Research*, vol. 32, no. 1, pp. 223–238, 2004.

[65] P. F. Agris, F. A. P. Vendeix, and W. D. Graham, "tRNA's wobble decoding of the genome: 40 years of modification," *Journal of Molecular Biology*, vol. 366, no. 1, pp. 1–13, 2007.

Hindawi Publishing Corporation
Archaea
Volume 2010, Article ID 481725, 10 pages
doi:10.1155/2010/481725

Research Article

Phosphorylation and Methylation of Proteasomal Proteins of the Haloarcheon *Haloferax volcanii*

Matthew A. Humbard, Christopher J. Reuter, Kheir Zuobi-Hasona, Guangyin Zhou, and Julie A. Maupin-Furlow

Department of Microbiology and Cell Science, University of Florida, Gainesville, FL 32611, USA

Correspondence should be addressed to Julie A. Maupin-Furlow, jmaupin@ufl.edu

Received 14 April 2010; Accepted 8 May 2010

Academic Editor: Jerry Eichler

Proteasomes are composed of 20S core particles (CPs) of α- and β-type subunits that associate with regulatory particle AAA ATPases such as the proteasome-activating nucleotidase (PAN) complexes of archaea. In this study, the roles and additional sites of post-translational modification of proteasomes were investigated using the archaeon *Haloferax volcanii* as a model. Indicative of phosphorylation, phosphatase-sensitive isoforms of α1 and α2 were detected by 2-DE immunoblot. To map these and other potential sites of post-translational modification, proteasomes were purified and analyzed by tandem mass spectrometry (MS/MS). Using this approach, several phosphosites were mapped including α1 Thr147, α2 Thr13/Ser14 and PAN-A Ser340. Multiple methylation sites were also mapped to α1, thus, revealing a new type of proteasomal modification. Probing the biological role of α1 and PAN-A phosphorylation by site-directed mutagenesis revealed dominant negative phenotypes for cell viability and/or pigmentation for α1 variants including Thr147Ala, Thr158Ala and Ser58Ala. An *H. volcanii* Rio1p Ser/Thr kinase homolog was purified and shown to catalyze autophosphorylation and phosphotransfer to α1. The α1 variants in Thr and Ser residues that displayed dominant negative phenotypes were significantly reduced in their ability to accept phosphoryl groups from Rio1p, thus, providing an important link between cell physiology and proteasomal phosphorylation.

1. Introduction

Proteasomes are multicatalytic proteases found in all three domains of life and are essential for growth in many organisms, including haloarchaea and eukaryotes [1–5]. Accessory proteins, such as 19S cap and proteasome activating nucleotidase (PAN) complexes, and protein-modification pathways such as ubiquitination are important in regulating protein degradation by proteasomes. Proteasomal proteins are also subject to co- and posttranslational modification (PTM) including N-terminal acetylation [6–9], phosphorylation [10–27], S-glutathionation [28], N-myristoylation [29], O-linked glycosylation [30], and processing of N-terminal propeptides including exposure of the Thr active site residues of 20S core particles (CPs) [31].

Despite decades of study, the mechanisms and biological significance of many of the posttranslational modifications of proteasomes are poorly understood. In particular, our knowledge of the phosphorylation of proteasomes remains limited. Polo-kinase, casein kinase II and calcium/calmodulin-dependent protein kinase II are known to mediate the phosphorylation of 26S proteasomes in vitro [11, 13, 14, 16–19, 27, 32, 33]. There is also evidence that proteasomal phosphorylation is regulated by growth and/or signaling molecules such as γ-interferon [11, 12, 17, 22, 24, 26].

Our previous work demonstrated that archaeal proteasomes undergo a variety of PTMs and alterations in subunit composition. In *H. volcanii*, these PTMs include N-terminal acetylation of the CP α1 and α2, removal of the CP β-propeptide and phosphorylation of the CP β Ser129 [34, 35]. The subunit composition of *H. volcanii* proteasomal complexes also appears to change as cells enter stationary phase. *H. volcanii* synthesizes two different proteasome-activating nucleotidase proteins (PAN-A and PAN-B) and three different CP subunits (α1, α2, and β) [36, 37].

Proteasomal CPs and PAN complexes of different subunit composition have been isolated from *H. volcanii* ([36], this study and unpublished results). While the transcript levels encoding these proteasomal proteins all increase as cells enter stationary phase, only the protein levels of $\alpha 2$ and PAN-B increase several-fold during this transition [37]. Thus, the population of different proteasomal subtypes appears to be modulated by growth.

In this report, we elaborate on the PTM of *H. volcanii* proteasomal proteins including the mapping of additional sites of phosphorylation as well as a new type of modification for proteasomes, methyl-esterification. We also report a correlation between the phosphorylation state of the CP $\alpha 1$ protein and growth phase, provide evidence that Rio1p Ser/Thr kinase phosphorylates $\alpha 1$ and suggest through site-directed mutagenesis that the state of proteasomal phosphorylation influences cell physiology including viability and pigmentation.

2. Materials and Methods

2.1. Materials. Biochemicals were purchased from Sigma-Aldrich (St. Louis, MO). Other organic and inorganic analytical-grade chemicals were from Fisher Scientific (Atlanta, GA) and Bio-Rad (Hercules, CA). Radioactive chemicals were purchased from Perkin Elmer (Waltham, MA), and Strep chromatography resin was purchased from Qiagen (Valencia, CA). Desalted oligonucleotide primers were from Integrated DNA Technologies (Coralville, IA). Site-directed mutagenesis materials were from Invitrogen (Carlsbad, CA). Polyvinyl difluoride (PVDF) membranes for Western blots were from Amersham Biosciences (Piscataway, NJ). Alpha-casein was from ICN Biochemicals (Cleveland, OH).

2.2. Strains, Media, and Plasmids. Strains and plasmids used in this study are summarized in Supplementary Table 1 available online at doi:10.1155/2010/481725. *H. volcanii* strains H26 and GZ130 (Δ*psmA*) and *E. coli* strain BL21(DE3) were used as hosts for protein production and purification. *E. coli* strains were grown in Luria-Bertani medium at 37°C and 200 rpm, and *H. volcanii* strains were grown in ATCC 974 medium at 42°C and 200 rpm. Media were supplemented with 100 μg/mL of ampicillin, 50 μg/mL of kanamycin, or 0.1 μg/mL of novobiocin as needed. For salt stress tests, the concentration of NaCl was reduced from the typical 2.25 M NaCl of the "normal" ATCC 974 medium to 1, 1.125, 1.25, 1.375, and 1.5 M NaCl. For the salt stress test inoculum, cells were grown in "normal salt" ATCC 974 medium by streaking cells from −80°C glycerol stocks onto fresh plates and growing these cells twice to log-phase in 2 mL medium. Each subculture was inoculated to a final OD_{600} of 0.01 to 0.02. Experiments were performed in biological triplicate and the mean ± S.D. was calculated.

2.3. Cloning and Site-Directed Mutagenesis. Site-directed mutagenesis (SDM) was performed using the Quikchange SDM kit as per manufacturer's instructions (Stratagene, La Jolla, CA) with the primers used for mutagenesis listed in Supplementary Table 2. The fidelity of all PCR-amplified products was confirmed by DNA sequencing (Genomics Core, Interdisiplinary Center for Biotechnology Research, University of Florida, Gainesville, FL) using the dideoxy termination method with Applied Biosystems (Foster City, CA) Model 3130 Genetic analyzer.

2.4. Protein Purification. Proteasomal proteins $\alpha 1$, $\alpha 2$, and β with C-terminal hexahistidine-(His_6) tags were purified from recombinant *E. coli* BL21 (DE3) strains after induction isopropyl β-D-1-thiogalactopyranoside (IPTG) at 0.4 mM for 2 hours as previously described [36]. In addition, proteasomal proteins ($\alpha 1$, $\alpha 2$, β, PAN-A, and PAN-B) with C-terminal His_6-tags, a Rio1p kinase homolog (HVO_0135) with a C-terminal -WSHPQFEK (-StrepII) tag and CPs composed of $\alpha 1$-His_6 and β-StrepII were purified from recombinant *H. volcanii*. For protein purification, the various *H. volcanii* strains expressing these derivatives were grown to stationary phase (OD_{600} of 2.0–2.5), unless otherwise indicated. Phosphatase inhibitor cocktail I and II were included in all lysis buffers and diluted according to supplier (Sigma). For His_6-tagged proteins, cells were lysed in 20 mM Tris pH 7.2 with 2 M NaCl (buffer A) supplemented with 5 mM imidazole. Cell lysate was loaded onto a nickel nitrilotriacetic acid (Ni-NTA) column (1.6 × 2.5 cm, Pharmacia) with a step gradient at 60 mM imidazole, and proteins were eluted from the column in buffer A with 500 mM imidazole. For StrepII-tagged proteins, cells were lysed and applied to a StrepTactin column (0.5 × 5 cm, Qiagen) in buffer A. Proteins were eluted from the column in buffer A with 2.5 mM desthiobiotin. For cells expressing both StrepII- and His_6-tagged proteins, "two-step" purification was performed which incorporated sequential Ni-NTA and StrepTactin chromatography [38]. Native molecular mass of Rio1p was determined by applying protein fractions eluted from a StrepTactin column to a calibrated Superose 200 HR 10/30 column as recommended by supplier (Pharmacia-GE Healthcare, Piscataway, NJ). Molecular mass standards for calibration included cytochrome c (12.4 kDa), carbonic anhydrase (29 kDa), bovine serum albumin (66 kDa), alcohol dehydrogenase (150 kDa), β-amylase (200 kDa), and apoferritin (443 kDa) (Sigma). For molecular mass standards, the gel filtration column was equilibrated in 20 mM Tris buffer at pH 7.2 supplemented with 150 mM NaCl. Salt concentration was increased to 2 M NaCl for Rio1 kinase.

2.5. Two-Dimensional Gel Electrophoresis and Western Blot Analysis. TRIzol-extracted and benzonase-treated protein (up to 250 μg) was applied to an 11 cm immobilized pH gradient (IPG) strip with a pH range of 3.9–5.1 (Bio-Rad). Proteins were separated according to their pI using manufacturer's protocol and previously published methods [39]. Proteins were further separated using either 10% or 12% SDS-PAGE Criterion gels (Bio-Rad) at 16°C at 200 V for 50 or 70 minutes, respectively. Proteins were transferred from gels to PVDF membranes at 100 V and 4°C for 90–100 minutes in 10 mM 2-(N-morpholino) ethanesulfonic acid

(MES) buffer at pH 6.0 supplemented with 10% (vol/vol) methanol. Polyclonal antibodies raised in rabbit against PAN-A and α1 and phosphoamino monoclonal antibodies (clone 22a, Becton Dickinson Biosciences, San Jose, CA) were used for Western blot analysis as previously described [37].

2.6. Phosphatase Treatment. Protein (up to 80 μg) was incubated (30°C, 20 minutes) in the presence and absence of lambda phosphatase (400 U) in 100 μL 50 mM Tris buffer at pH 7.5 supplemented with 0.1 mM EDTA, 5 mM dithiothreitol and 2 mM $MnCl_2$ (lambda phosphatase buffer).

2.7. Peptidase Assay. Chymotrypsin-like peptide hydrolyzing activity was assayed by release of 7-amino-4-methylcoumarin from N-Suc-Leu-Leu-Val-Tyr-7-amido-4-methylcoumarin (Suc-LLVY-AMC; Sigma) at 60°C for 10–30 minutes by an increase in fluorescence (excitation = 348 nm, emission = 440 nm) as previously described [40]. The assay mixture (0.3 mL) contained 1–3 μg protein and 20 μM fluorogenic substrate in buffer A with 0.4% (v/v) dimethyl sulfoxide (DMSO).

2.8. Kinase Assay. A Rio1p kinase homolog (HVO_0135) was expressed in *H. volcanii* with a C-terminal StrepII tag (-WSHPQFEK) that included a GT linker (encoded by a KpnI site) and purified to electrophoretic homogeneity by StrepTactin chromatography as described above. The purified kinase was dialyzed into 20 mM Tris buffer (pH 7.2) with 2 M NaCl or 2 M KCl at 4°C using a D-tube dialysis tube (EMD Chemicals, Gibbstown, NJ). Autophosphorylation was monitored by incubating purified Rio1p (1 μg) with $MnCl_2$ or $MgCl_2$ (0–250 mM) for one hour at 37°C in the presence of 200–400 μCi of [γ-^{32}P] adenosine 5′-triphosphate (ATP) (Perkin Elmer) in a reaction volume of 20 μL. For *in vitro* phosphorylation, 10 μg of α1 variants were included in the assay mixture. Kinase reactions were inactived by addition of an equal volume of reducing SDS-PAGE loading buffer and separated by 12.5% SDS-PAGE gels. Gels were stained with Coomassie brilliant blue and dried to Whatman paper. Blue sensitive X-ray film (Research Products International Corp., Mt. Prospect, IL) was exposed to the dried gel for 1–10 days at −80°C.

2.9. Mass Spectrometry. Proteasomal fractions (3 to 6 μg protein) from nickel affinity chromatography and/or two-step affinity chromatography were separated by reducing 12% SDS-PAGE. Gels were stained with Bio-Safe Coomassie (Bio-Rad). Protein bands corresponding to the molecular masses of the CP subunits were excised from the gel and destained with 100 mM NH_4HCO_3 in 50% (vol/vol) acetonitrile (4°C, overnight). Protein samples were reduced, alkylated in-gel, and digested with trypsin (Promega, Madison, WI). Capillary reverse-phase high-performance liquid chromatography (HPLC) separation of the protein digests was performed using a PepMap C18 column (75-μm inside diameter, 15-cm length; LC Packings, San Francisco, CA) with an Ultimate capillary HPLC system (LC Packings). A gradient (90 or 120 minutes) from 5 to 50% acetonitrile in 0.1% acetic acid was used at a flow rate of 200 nL per min. Tandem mass spectrometry (MS/MS) analysis was performed using a hybrid quadrupole time-of-flight (QTOF) instrument (QSTAR) equipped with a nanoelectrospray source and operated with Analyst QS 1.1 data acquisition software (Applied Biosystems).

2.10. Database Searching. Tandem mass spectrometry (MS/MS) data generated by information-dependent acquisition with QSTAR were searched against the deduced proteome of the *H. volcanii* DS2 genome sequence [41] by using the Mascot (Matrix Science, Boston, MA) database search engine. Probability-based Mascot scores were determined by a comparison of search results against estimated random match population and are reported as $\sim 10 \times \log_{10}(p)$, where p is the absolute probability. Individual Mascot ion scores greater than 32 were considered to indicate identity or extensive homology ($P < .05$). Scores below this default significant value were considered for protein identification in addition to validation by manual interpretation of MS/MS spectra. Carbamidomethylation of cysteine was used as a fixed modification based on sample preparation. Variable modifications that were implemented in the database search included deamidation of asparagine and glutamine, N-terminal acetylation, oxidation (single and double) of methionine, and phosphorylation of serine, threonine, and tyrosine.

3. Results

3.1. Isoforms of α1 Modulated during Growth-Phase Transitions. Posttranslational modification of α1, one of the predominant proteasomal proteins in *H. volcanii* cells, was examined during various stages of growth from log- to late-stationary phase by 2-DE immunoblot of cell lysate. Two α1-specific isoforms of pI 3.0 and 4.4 were detected, both of which migrated at 37.5 kDa consistent with the mobility of the α1 protein in reducing SDS-PAGE gels [34]. The α1-isoform of pI 4.4 was present at all stages of growth while the second more acidic α1-isoform of pI 3.0 was present in earlier phases of growth and disappeared in late-stationary phase (Figure 1(a)).

Affinity purification of CPs from early stationary-phase cells expressing a hexahistidine tagged (-His_6) derivative of the catalytic β subunit revealed that both isoforms of α1 (pI 3.0 and 4.4) associate as CPs active in hydrolysis of the peptide reporter Suc-LLVY-AMC (Figure 1(b)). The α1 isoform of pI 4.4 was also detected in non-CP fractions. Treatment of these purified 20S CPs with phosphatase resulted in a basic shift in pI of both α1 and α2 proteins signifying both α-type subunits were phosphorylated (Figure 1(c)).

Together these results suggest that the α-type proteins of archaeal proteasomal CPs can be modified by phosphorylation. Our previous studies demonstrate that the β subunits of *H. volcanii* CPs are also phosphorylated at Ser129 [35]. Thus, phosphorylation appears to be a major modification of the proteasome system of this haloarchaeon. Based on

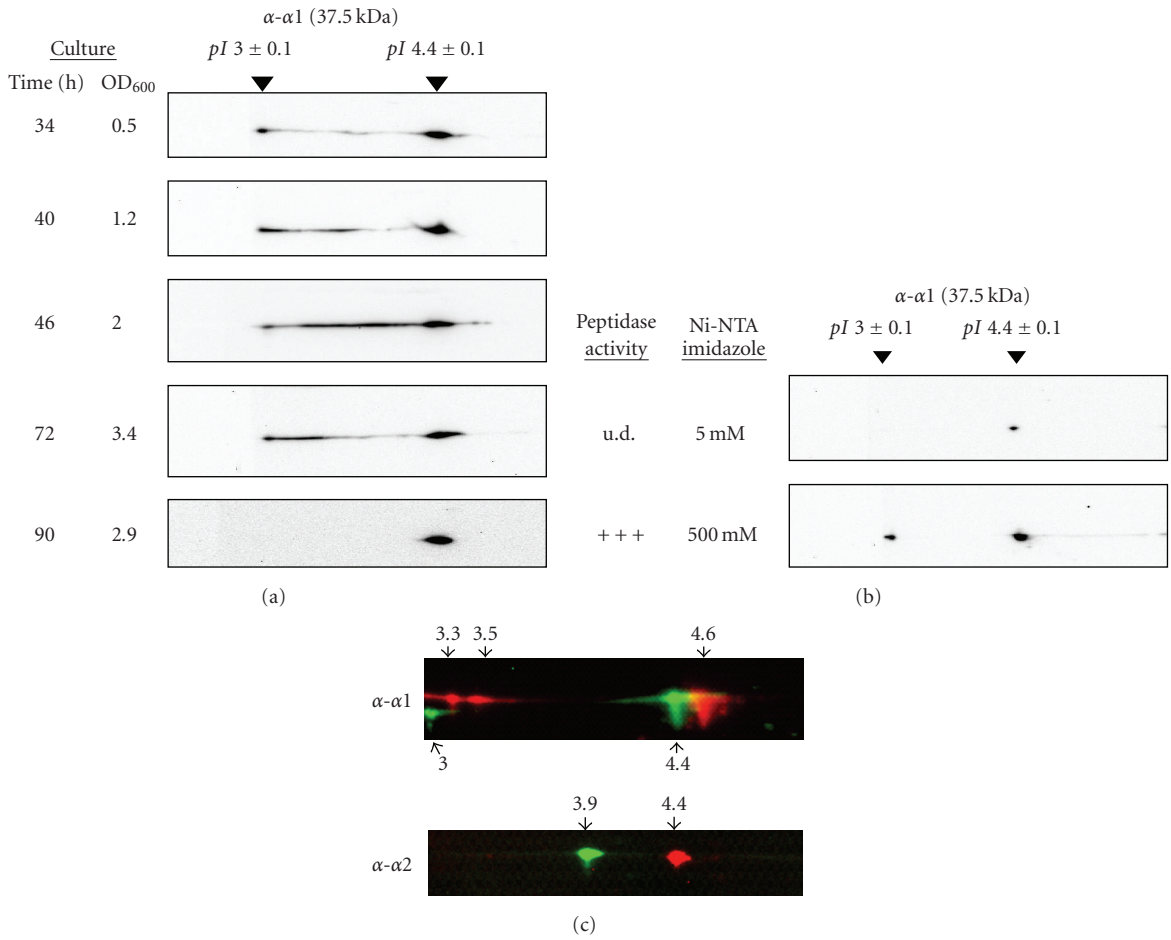


FIGURE 1: Phosphatase-sensitive isoforms of α-type proteins of proteasomal CPs as a function of growth in *H. volcanii* cells. (a) An α1 isoform of pI 4.4 was present at all stages of growth with a more acidic isoform of pI 3.0 present in log- and early-stationary phase, but absent in late-stationary phase. Cell lyase was prepared from various stages of growth as indicated (1 OD_{600} unit ~1×10^9 CFU·mL^{-1}), separated by 2-DE and analyzed by immunoblot using anti-α1 antibody (α-α1). (b) The two α1 isoforms of pI 3.0 and 4.4 are associated in proteasomal CPs. Proteasomal CPs were purified by Ni-NTA chromatography from early stationary-phase cells expressing β-His_6. Proteins were separated into two fractions: (i) flowed through the Ni-NTA column at 5 mM imidazole and (ii) bound and were eluted from the column at 500 mM imidazole. Protein fractions were separated by 2-DE and analyzed by immunoblot using anti-α1 antibody (α-α1). Fractions were also assayed for peptidase activity using Suc-LLVY-AMC. u.d., undetectable. (c) Both α1 and α2 isoforms are sensitive to phosphatase treatment. Proteasomal CPs (purified as above) were treated with (red) and without (green) phosphatase, separated by 2-DE, and probed by immunoblot with anti-α1 and anti-α2 antibodies as indicated (α-α1 and α-α2, resp.).

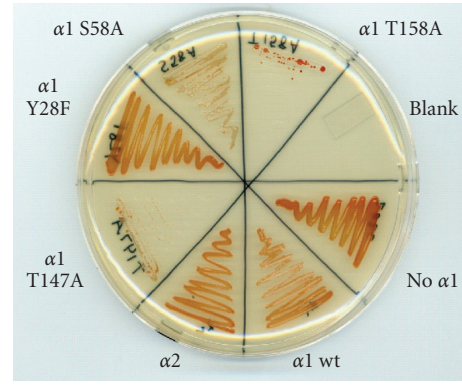
the isoelectric focusing (IEF) migration and phosphatase sensitivity of both α1 isoforms of pI 3.0 and 4.4, we propose that α1 harbors at least two phosphosites. The least phosphorylated form of α1 (pI 4.4) was detected at all stages of growth and occurred in both free and CP-associated forms. In contrast, the more extensively phosphorylated form of α1 (pI 3.0) was only detected in association with CPs and was not detected in late-stationary phase. In addition to these modified forms of α1, CP subtypes appear to incorporate a phosphorylated form of α2. Thus, *H. volcanii* not only modulates the types of subunits present in the CP and PAN subtypes during the transition from log- to stationary-phase growth [37], but appears also to posttranslationally modify the CP proteasomal proteins during this transition.

3.2. α1, α2, and PAN-A Are Phosphorylated at Ser/Thr Residues. To further map the phosphosites of CPs and PAN complexes, the various subunits (α1, α2, β, PAN-A, and PAN-B) were enriched from *H. volcanii* by nickel chromatography using His_6 epitope-tags and analyzed by MS after tryptic digest. Initially, CPs were purified from log-phase *H. volcanii* H26-pJAM204 cells expressing α1-His_6 and separated by SDS-PAGE prior to RP-HPLC MS/MS-analysis. Using this approach, a high-probability phosphopeptide was reproducibly detected in the α1-specific protein band that mapped to amino acid residues 137 to 149 of α1 (ALLIGGVENGpTPR, where pT represents phospho-Thr147) (Supplementary Figure 1). MS/MS-fragmentation of this peptide included a characteristic neutral loss (−98) indicative of CID removal of the phosphoryl group and

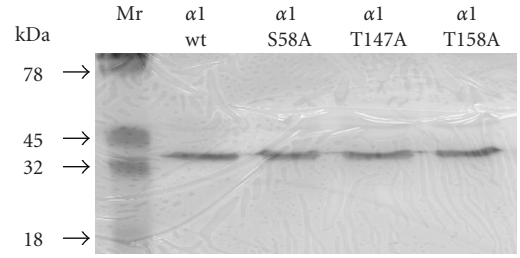
confirmed that Thr147 was a site of α1 phosphorylation. The Thr147 residue was conserved in eukaryotic α-type CP proteins and was predicted to be phosphorylated by NetPhos (with a score of 0.851). An α2-specific phosphopeptide was also detected by precursor ion scanning of α2-specific tryptic peptides purified from *H. volcanii* H26-pJAM205 cells expressing α2-His_6. The α2-specific phosphopeptide was composed of amino acid residues 12 to 21 (GTSLF-SPDGR) and again displayed a neutral loss characteristic of phosphorylated peptides (Supplementary Figure 2). MS/MS, however, could not distinguish between Thr13 and Ser14 as the phosphosite due to the close vicinity of these residues within the peptide. Of these two residues, Ser14 has a higher probability of serving as a phosphosite compared to Thr13 based on modeling to known eukaryotic phosphoproteins with NetPhos scores of 0.957 and 0.243, respectively. In addition to the α1- and α2-specific phosphopeptides, a PAN-A-specific phosphopeptide was reproducibly detected by ESI-QTOF analysis of tryptic digestions of PAN complexes purified from *H. volcanii* expressing either PAN-A-His_6 or PAN-B-His_6 (DS70-pJAM650 or pJAM1012, resp.) (Supplementary Figure 3). This phosphopeptide mapped to amino acid residues 337 to 361 of PAN-A (MNVpSDDVD-FVELAEMADNASGADIK, where pS represents phospho-Ser340). MS/MS fragmentation confirmed that Ser340 was the phosphorylated amino acid including a neutral loss peak (-98 for phosphorylation) and several internal fragmentation ions. It is important to note that the PAN-A-Ser340-specific phosphopeptide was detected in strains expressing either PAN-A-His_6 or PAN-B-His_6. This indicates that phosphorylated PAN-A associates with PAN-B in *H. volcanii* cells. Interestingly, the PAN-A Ser340 and amino acids surrounding this residue are highly conserved among eukaryotic 26S proteasome Rpt-like proteins including those of humans; however, the prediction of this as a phoshosite by NetPhos was low with a score of 0.464. Similar to the Ser129 phosphosite that was mapped to the CP β subunit [35], analogous nonphosphorylated peptides also mapped to α1 Thr147, α2 Thr13/Ser14, and PAN-A Ser340 residues. These results suggest that the phosphosites of PAN and CP complexes are not fully occupied by phosphoryl groups and, thus, the phosphorylation of proteasomes may be regulated in the cell.

3.3. Variant α1 Proteins Affect Pigmentation and Cell Viability. Single and double knock-out of *psmA* (α1) and *panA* (PAN-A) genes diminishes the ability of *H. volcanii* cells to adapt to low salt and other stressful conditions [5]. These phenotypes are readily complemented by providing a copy of the corresponding wild type gene in *trans* and, thus, provide an ideal assay for examining how phosphorylation may influence in the function of α1 and PAN-A in the cell.

Initially, *panA* mutant cells were tested for their ability to be complemented for growth on low-salt media with *panB*, *panA*, or $panA_{t1018g}$ (PAN-A Ser340Ala) genes. For complementation, the genes were expressed in *trans* from a multicopy pHV2-based plasmid using a strong rRNA P2 promoter. While overexpression of *panB* from the plasmid did not rescue the sensitivity of the *panA* mutant to low



(a)



(b)

Figure 2: Analysis of an *H. volcanii psmA* (α1) knockout expressing wild type and variant α-type proteasomal genes *in trans*. (a) Growth and pigmentation of an *H. volcanii psmA* mutant (GZ130) were influenced by expression of α1 variant proteins on ATCC novobiocin plates. Synthesis of α1 Thr147Ala and α1 Thr158Ala inhibited growth of the *psmA* mutant, while α1 Ser58Ala and α1 Thr147Ala altered pigmentation. (b) No significant difference in α1 protein levels were detected in the *H. volcanii psmA* mutant synthesizing the wild type and variant α1 proteins from genes provided *in trans*. The α-type proteins encoded by the genes expressed *in trans* are indicated with the following abbreviations: wt, wild type; S58A, Ser58Ala; Y28F, Tyr28Phe; T158A, Thr158Ala; T147A, Thr147Ala.

salt conditions, expression of either *panA* or *panA* Ser340Ala restored growth to wild-type levels. This indicates that phosphorylation of Ser340 is not important for response to hypo-osmotic stress (data not shown) and suggests that *panB* cannot substitute for *panA* under these growth conditions.

Site-directed mutations in the *psmA* gene encoding single amino acid substitutions of Ser58Ala, Thr147Ala, Thr158Ala, and Tyr28Phe in α1 were tested (along with overexpression of *psmC* encoding α2) for their ability to complement the hypo-osmotic stress response phenotype of a *psmA* mutant. Similar to *panA*, the modified *psmA* and *psmC* genes were provided in *trans* using a pHV2-based plasmid and strong rRNA P2 promoter. The α1-residues selected for site-directed modification were based on the following: α1 Thr147 was detected in this study as a phosphosite by MS/MS (see above for details), α1 Tyr28 was homologous to the Tyr27 phosphosite identified in the α2-type subunit of eukaryotic CPs [22], and Ser58 and Thr158 are predicted to be phosphorylated based on on NetPhos 2.0 scores of 0.996 and 0.981, respectively, (significantly above the significance threshold of

0.500) [42]. Overexpression of the *psmC* (α2) gene did not complement the *psmA* mutant while the *psmA* and *psmA-a83t* (α1 Tyr28Phe) genes complemented the reduced growth phenotype of the *psmA* mutant on low salt media to wild type levels. In contrast, the other substitutions to the *psmA* gene provided *in trans* severely inhibited overall growth of the *psmA* mutant strain. While each of the constructs was successfully transformed into the *psmA* knockout strain, growth of the transformed cells was drastically reduced after being transferred to liquid media or streaked onto a fresh plate (Figure 2(a)). Some of the substitutions in α1 including Ser58Ala and α1 Thr147Ala appeared to affect carotenoid biosynthesis as cells expressing these α1 variants were whiter than cells expressing wild-type α1 or no α1 at all (Figure 2(a)). Whole cell chymotrypsin-like (CL) peptidase activity of the different α1 substitution mutants expressed *in trans* was compared to wild type after initial transfer to liquid media. Only modest differences in specific activity were detected between the strains with the most pronounced difference in cells expressing α1 Thr158Ala, which had a ~2-fold increase in CL activity compared to wild-type (data not shown). It is unclear at this time if these modest differences in CL activity can account for the phenotypes observed. Anti-α1 immunoblot of cell lysate revealed all of the α1 variant proteins were produced at levels similar to wild type when the *psmA* genes were expressed *in trans* in the *psmA* knockout (Figure 2(b)). Both the hypo-osmotic stress and pigmentation phenotypes were recessive based on the observation that these α1 variants did not alter growth or pigmentation when expressed *in trans* in wild-type *H. volcanii* cells (data not shown).

3.4. Haloferax volcanii Encodes a Number of Putative Protein Kinases. Based on genome sequence, *H. volcanii* encodes for a number of putative serine/threonine kinases including two different Rio-type protein kinases, type-1 (HVO_0135) and type-2-like (HVO_0569) (Figure 3(a)) as well as homologs of the serine protein kinase PrkA, originally described in *Bacillus subtilis* [43]. Rio-type kinases are a relatively new atypical protein kinase family with four unique types, Rio1, Rio2, Rio3, and RioB. Rio-type kinases are widespread throughout eukaryotes and archaea. Archaea typically have two different Rio-type kinases, Rio1 and Rio2 [44]. Rio1 kinases are characterized by a consensus sequence STGKEA in the N-terminal domain of the protein and a second region of homology in the C-terminal domain, IDxxQ (where x represents any amino acid residue). While the Rio2 kinases can vary somewhat in domain structure, these kinases are characterized by a helix-turn-helix motif in the N-terminus followed by the amino acid sequence GxGKES and a C-terminal IDFPQ sequence [45]. The Rio1p from *H. volcanii* contains the signature domain sequence but the Rio2p does not (Figure 3(a)).

3.5. H. volcanii Rio Type-I Kinase Catalyzes Autophosphorylation. To investigate the activity of the Rio1p of *H. volcanii*, the C-terminal residue of this kinase was fused to a StrepII tag and synthesized in recombinant *H. volcanii* (H26-pJAM2558). Rio1p was purified to electrophoretic homogeneity by StrepTactin chromatography (Figure 3(b)) and determined to be a 50-kDa monomer based on gel filtration chromatography compared to migration by reducing SDS-PAGE and deduced polypeptide sequence. Rio1p catalyzed its autophosphorylation in the presence of either Mg^{2+} or Mn^{2+} as the divalent cation (Figure 3(c)). The ideal concentration of cation was 10 to 50 mM Mg^{2+} in 20 mM Tris-buffer pH 7.2 supplemented with 2 M NaCl with 4 μM enzyme. The levels of Rio1p autokinase activity were comparable when assayed in the presence of either 2 M KCl or NaCl (data not shown). All subsequent kinase

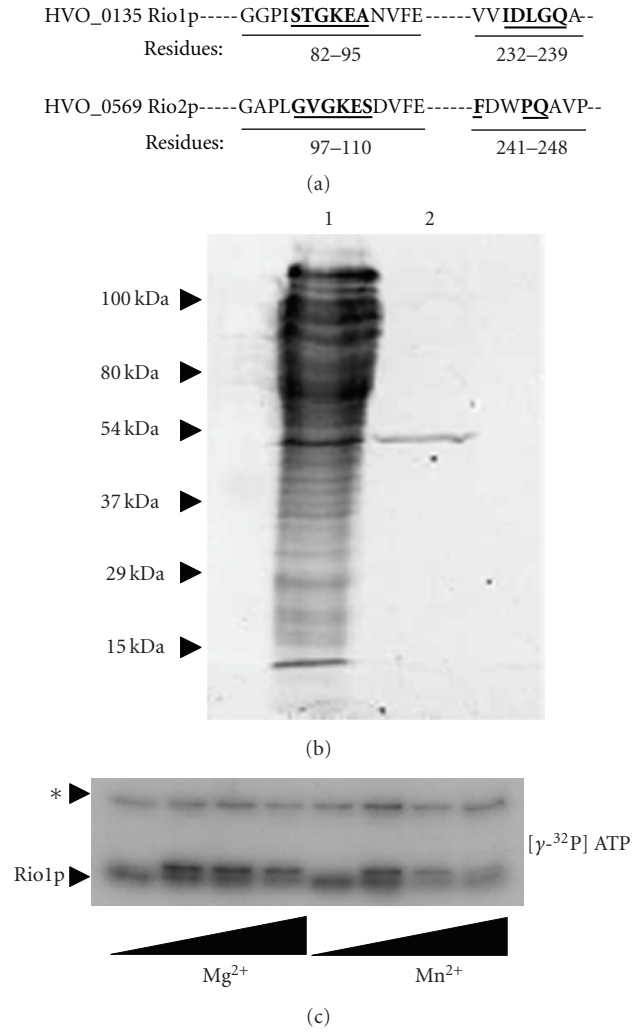
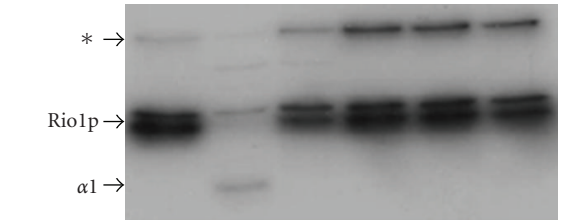
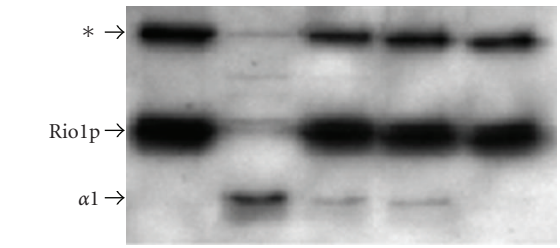


Figure 3: Rio-type protein kinases of *H. volcanii*. (a) Rio-type protein kinase homologs of *H. volcanii*, type-1 (HVO_0135) and type-2-like (HVO_0569), with conserved amino acid residues in bold and underlined. While HVO_0569 does not have a complete C-terminal domain, it does contain three of the five conserved residues of Rio type-2 kinases. (b) Purification of Rio1p (HVO_0135) from *H. volcanii*. Reducing SDS-PAGE gel of cell lysate (lane 1) and Rio1p purified by StrepTactin chromatography (lane 2) from in *H. volcanii*-pJAM2558. (c) Rio1p catalyzes its autophosphorylation. Autophosphotransfer was detected *in vitro* by autoradiography using [γ-^{32}P] ATP as a substrate with optimal activity detected at 10 to 50 mM $MgCl_2$. *, identity of band remains to be determined.



[γ-^{32}P] ATP	+	+	+	+	+	+
Rio1p	+	+	+	+	+	+
α1	+	+	–	–	–	–
α1 (boiled)	–	–	+	–	–	–
CP	–	–	–	+	–	–
CP (boiled)	–	–	–	–	+	–
α-casein	–	–	–	–	–	+

(a)



[γ-^{32}P] ATP	+	+	+	+	+
Rio1p	+	+	+	+	+
α1 wt	–	+	–	–	–
α1 Ser58Ala	–	–	+	–	–
α1 Thr147Ala	–	–	–	+	–
α1 Thr158Ala	–	–	–	–	+

(b)

Figure 4: Rio1p phosphorylates α1 at Ser and Thr residues. (a) Phosphotransfer of Rio1p to α1 is catalyzed using α1-His$_6$ from recombinant *E. coli* as a substrate. Phosphotransfer was assayed using the protein substrates indicated as follows: α1-His$_6$ (boiled and unboiled) from recombinant *E. coli*, 20S CPs of α1-His$_6$ and β-StrepII subunits (boiled and unboiled) from *H. volcanii* and α-casein from bovine milk (boiled), as indicated. (b) Phosphotransfer of Rio1p to α1 is reduced by site-directed α1 protein variants Ser58Ala, Thr147Ala and Thr158Ala. The α1 wild type and protein variants were purified from recombinant *E. coli* and were not boiled prior to assay. Phosphotransfer was detected by autoradiography using Rio1p-StrepII purified by StrepTactin chromatography and γ-^{32}P P-ATP as a substrate.

reactions were carried out in 20 mM Tris buffer (pH 7.2) with 2 M NaCl and 50 mM $MgCl_2$. Multiple bands were visible on autoradiographs after the Rio1p-mediated autophosphorylation reaction including a doublet at the apparent molecular mass of the Rio1p monomer in addition to a band running at approximately twice the mass. MS analysis of Rio1p (from solution) did not identify any contaminating proteins in the purified fraction (Supplementary Table 3). The multiple phospho-bands observed after assay may represent a transient dimer that forms between two Rio1p proteins in the presence of ATP. It is also possible that there is a contaminating protein that is phosphorylated in the reaction mixture and is not detected by MS or SYPRO staining.

3.6. Rio1p Phosphorylates α1 Proteins In Vitro. Purified Rio1p was assayed for its ability to phosphorylate α1-His$_6$ purified from recombinant *E. coli*, 20S CPs purified from *H. volcanii* and α-casein from bovine milk. While the α1-proteins purified from *E. coli* ($\alpha1_{EC}$) form a mixture of hexameric rings and monomers, these recombinant proteins are not anticipated to be phosphorylated in recombinant *E. coli* based on their migration as a single spot by 2-DE (data not shown) as well as the distant phylogenetic relationship between *E. coli* and *H. volcanii*. Rio1p phosphorylated the unboiled $\alpha1_{EC}$ preparation but was unable to phosphorylate the CPs or boiled $\alpha1_{EC}$ or α-casein (Figure 4(a)). When equal amounts of $\alpha1_{EC}$ and site-directed α1 variants all purified from *E. coli* were used in the kinase assay, varying levels of α1-phosphorylation were observed. Wild-type $\alpha1_{EC}$ was the most heavily phosphorylated under these conditions. Both $\alpha1_{EC}$ Ser58Ala and $\alpha1_{EC}$ Thr147Ala were phosphorylated but to a lesser extent than wild-type $\alpha1_{EC}$ (2.5-fold lower based on densitometry of autoradiograph). The variant $\alpha1_{EC}$ Thr158Ala was not phosphorylated under these conditions (Figure 4(b)). Based on these results, coupled with our earlier findings by 2-DE, α1 is most likely phosphorylated at 2 or 3 different sites. It is unclear at this time why $\alpha1_{EC}$ Thr158Ala was not phoshporylated by Rio1p *in vitro*. There may be drastic structural changes in this α1 protein that prevent meaningful interaction with the kinase.

3.7. α1 Is Methylesterified on Acidic Residues. In addition to phosphorylation, the α1 protein was also methylated. Five unique methylation sites were reproducibly detected with high probability by MS/MS for α1 (Asp20, Glu27, Glu62, Glu112, and Glu161) (Supplementary Figure 4). Substantiating this finding, several of the MS/MS-fragmentations also showed neutral losses of methyl groups (−14). Unmodified α1 peptides that mapped to the same residues as those that were methylated were also detected. The analogous peptide in its unmodified state was almost always present in the same preparation, suggesting that methylated and unmethylated α1 proteins exist simultaneously in the cell. O-linked methylesterification is a reversible posttranslational modification, and therefore α1 methylation may be regulated. However, there was no clear correlation to growth phase or environmental conditions examined that increased or decreased the frequency of the methylated state of α1 protein (data not shown).

The *psmA* gene, encoding for the CP α1 protein, is cotranscribed with a putative S-adenylosylmethionine-dependent methyltransferase (HVO_1093) [46]. Whether this enzyme mediates α1 methylation remains to be determined; its *in vitro* α1-metylating activity could not be demonstrated (data not shown). Future work is needed to better understand how and why this modification takes place.

4. Discussion

Our previous work established that the three CP subunits of *H. volcanii* proteasomes (α1, α2, and β) are modified post- and/or co-translational [34, 35]. The β subunit is phosphorylated and cleaved to expose the active site Thr, and

the α1 and α2 subunits are N^{α}-acetylated on the initiator methionine residue. However, additional modifications were suspected based on the detection of long isoform chains of CP proteins including at least four α1 isoforms that were separated by 2-DE [35].

In this study, the covalent modification of archaeal proteasomal proteins was further examined using *H. volcanii* as a model system. Additional sites of posttranslational modification were identified including the phosphorylation of α1, α2, and PAN-A as well as the methyl esterification of α1. The phosphorylation of α1 was temporally regulated with the number of phosphatase-sensitive α1 isoforms reduced during late-stationary phase. The most acidic, presumably most phosphorylated isoform of α1, was only detected in association with β as 20S CPs and was not detected in late-stationary phase.

In addition to elaborating on the prevalence and temporal nature of archaeal proteasomal PTMs, a member of the Rio-kinase family, Rio1p, was demonstrated to catalyze the phosphorylation of α1 *in vitro*. Modification of Ser58, Thr147, or Thr158 residues of α1 to alanine significantly reduced this Rio1p-dependent phosphotransfer activity. Although the finding that Rio1p phosphorylates α1 *in vitro* does not imply that this interaction is physiological, preliminary data on this substrate/kinase pair may prove valuable in understanding the role phosphorylation plays in regulating 20S CPs or other proteins in archaea. Substrate/kinase pairs are rarely known in this domain of life [47].

In vivo expression of phosphosite mutant forms of α1 revealed a biological connection between the phosphorylation of proteasomes and the pigmentation and viability of archaeal cells. While the details of this link remain to be determined at the molecular level, cell division is known to be highly regulated by the ubiquitin (Ub)-proteasome system in eukaryotic cells [48]. In addition, 3-hydroxy-3-methylglutaryl coenzyme A (HMG-CoA) reductase, a key enzyme in the mevalonate pathway for production of isoprenoids (including carotenoids) is heavily regulated by proteasomes [49], and the Ub ligase CrgA is required for proper regulation of carotenogenesis in fungi [50].

In addition to phosphorylation, the α1 subunit of 20S CPs was found to be heavily methylated in *H. volcanii*. Although methylation has not been previously reported as a PTM of proteasomes in any organism, there are examples of reversible methylation of proteins in bacteria and archaea. Like bacteria, the haloarchaea have several taxis proteins that are O-methylated resulting in altered protein-protein interactions that enable cells to respond to external stimuli [51–53]. Haloarchaea may regulate the activity of other systems, such as proteasomes, using a similar strategy. Alternatively, the methyl esterification of α1 may be used to reversibly modulate the overall charge of proteasomal CPs. Most haloarchaeal proteins, including α1, are highly acidic (low pI) with the acidic residues typically on the surface of the protein forming a hydration shell. This shell aids in "salting-in" the protein and maintaining the activity of the protein at the unusually high concentrations of salt found in the cytosol of these microorganisms. Although we did not detect quantitative differences in the methylation state of α1 under different growth conditions, α1 is required for *H. volcanii* to properly respond to low salt stress [5]. Whether the methyl esterification of α1 modifies protein-protein interactions in regulatory mechanisms such as PAN : CP interactions, buffers α1 in the cytosol at different concentrations of salt and/or mediates other physiological processes remains to be determined.

Acknowledgments

The authors thank Scott McClung and Dr. S. Chen at the University of Florida Interdisciplinary Center for Biological Research for assistance with mass spectrometry experiments and Dr. S. Shanker for DNA sequencing. This work was funded in part by NIH R01 GM057498 and DOE DE-FG02-05ER15650.

References

[1] D. H. Lee, K. Tanaka, T. Tamura, C. H. Chung, and A. Ichihara, "PRS3 encoding an essential subunit of yeast proteasomes homologous to mammalian proteasome subunit C5," *Biochemical and Biophysical Research Communications*, vol. 182, no. 2, pp. 452–460, 1992.

[2] D. H. Lee, T. Tamura, C. H. Chung, K. Tanaka, and A. Ichihara, "Molecular cloning of the yeast proteasome PRS2 gene identical to the suppressor gene scl1+," *Biochemistry International*, vol. 23, no. 4, pp. 689–696, 1991.

[3] E. Georgatsou, T. Georgakopoulos, and G. Thireos, "Molecular cloning of an essential yeast gene encoding a proteasomal subunit," *FEBS Letters*, vol. 299, no. 1, pp. 39–43, 1992.

[4] T. Fujiwara, K. Tanaka, E. Orino et al., "Proteasomes are essential for yeast proliferation. cDNA cloning and gene disruption of two major subunits," *Journal of Biological Chemistry*, vol. 265, no. 27, pp. 16604–16613, 1990.

[5] G. Zhou, D. Kowalczyk, M. A. Humbard, S. Rohatgi, and J. A. Maupin-Furlow, "Proteasomal components required for cell growth and stress responses in the haloarchaeon *Haloferax volcanii*," *Journal of Bacteriology*, vol. 190, no. 24, pp. 8096–8105, 2008.

[6] C. S. Arendt and M. Hochstrasser, "Eukaryotic 20S proteasome catalytic subunit propeptides prevent active site inactivation by N-terminal acetylation and promote particle assembly," *EMBO Journal*, vol. 18, no. 13, pp. 3575–3585, 1999.

[7] Y. Kimura, Y. Saeki, H. Yokosawa, B. Polevoda, F. Sherman, and H. Hirano, "N-terminal modifications of the 19S regulatory particle subunits of the yeast proteasome," *Archives of Biochemistry and Biophysics*, vol. 409, no. 2, pp. 341–348, 2003.

[8] Y. Kimura, M. Takaoka, S. Tanaka et al., "Nα-acetylation and proteolytic activity of the yeast 20 S proteasome," *Journal of Biological Chemistry*, vol. 275, no. 7, pp. 4635–4639, 2000.

[9] F. Tokunaga, R. Aruga, S. Iwanaga et al., "The NH_2-terminal residues of rat liver proteasome (multicatalytic proteinase complex) subunits, C2, C3 and C8, are Nα-acetylated," *FEBS Letters*, vol. 263, no. 2, pp. 373–375, 1990.

[10] F. Bardag-Gorce, R. Venkatesh, J. Li, B. A. French, and S. W. French, "Hyperphosphorylation of rat liver proteasome subunits: the effects of ethanol and okadaic acid are compared," *Life Sciences*, vol. 75, no. 5, pp. 585–597, 2004.

[11] S. Bose, F. L. L. Stratford, K. I. Broadfoot, G. G. F. Mason, and A. J. Rivett, "Phosphorylation of 20S proteasome alpha

subunit C8 (α7) stabilizes the 26S proteasome and plays a role in the regulation of proteasome complexes by γ-interferon," *Biochemical Journal*, vol. 378, no. 1, pp. 177–184, 2004.

[12] S. Bose, P. Brooks, G. G. F. Mason, and A. J. Rivett, "γ-interferon decreases the level of 26 S proteasomes and changes the pattern of phosphorylation," *Biochemical Journal*, vol. 353, no. 2, pp. 291–297, 2001.

[13] S. Bose, G. G. F. Mason, and A. J. Rivett, "Phosphorylation of proteasomes in mammalian cells," *Molecular Biology Reports*, vol. 26, no. 1-2, pp. 11–14, 1999.

[14] J. G. Castaño, E. Mahillo, P. Arizti, and J. Arribas, "Phosphorylation of C8 and C9 subunits of the multicatalytic proteinase by casein kinase II and identification of the C8 phosphorylation sites by direct mutagenesis," *Biochemistry*, vol. 35, no. 12, pp. 3782–3789, 1996.

[15] S. Claverol, O. Burlet-Schiltz, E. Girbal-Neuhauser, J. E. Gairin, and B. Monsarrat, "Mapping and structural dissection of human 20 S proteasome using proteomic approaches," *Molecular and Cellular Proteomics*, vol. 1, no. 8, pp. 567–578, 2002.

[16] P. Fernández Murray, P. S. Pardo, A. M. Zelada, and S. Passeron, "*In vivo* and *in vitro* phosphorylation of *Candida albicans* 20S proteasome," *Archives of Biochemistry and Biophysics*, vol. 404, no. 1, pp. 116–125, 2002.

[17] R. Horiguchi, M. Yoshikuni, M. Tokumoto, Y. Nagahama, and T. Tokumoto, "Identification of a protein kinase which phosphorylates a subunit of the 26S proteasome and changes in its activity during meiotic cell cycle in goldfish oocytes," *Cellular Signalling*, vol. 17, no. 2, pp. 205–215, 2005.

[18] Y. Iwafune, H. Kawasaki, and H. Hirano, "Identification of three phosphorylation sites in the α7 subunit of the yeast 20S proteasome in vivo using mass spectrometry," *Archives of Biochemistry and Biophysics*, vol. 431, no. 1, pp. 9–15, 2004.

[19] G. G. F. Mason, K. B. Hendil, and A. J. Rivett, "Phosphorylation of proteasomes in mammalian cells Identification of two phosphorylated subunits and the effect of phosphorylation on activity," *European Journal of Biochemistry*, vol. 238, no. 2, pp. 453–462, 1996.

[20] G. G. F. Mason, R. Z. Murray, D. Pappin, and A. J. Rivett, "Phosphorylation of ATPase subunits of the 26S proteasome," *FEBS Letters*, vol. 430, no. 3, pp. 269–274, 1998.

[21] P. S. Pardo, P. F. Murray, K. Walz, L. Franco, and S. Passeron, "*In vivo* and *in vitro* phosphorylation of the α7/PRS1 subunit of *Saccharomyces cerevisiae* 20 S proteasome: *in vitro* phosphorylation by protein kinase CK2 is absolutely dependent on polylysine," *Archives of Biochemistry and Biophysics*, vol. 349, no. 2, pp. 397–401, 1998.

[22] A. J. Rivett, S. Bose, P. Brooks, and K. I. Broadfoot, "Regulation of proteasome complexes by γ-interferon and phosphorylation," *Biochimie*, vol. 83, no. 3-4, pp. 363–366, 2001.

[23] K. Satoh, H. Sasajima, K.-I. Nyoumura, H. Yokosawa, and H. Sawada, "Assembly of the 26S proteasome is regulated by phosphorylation of the p45/Rpt6 ATPase subunit," *Biochemistry*, vol. 40, no. 2, pp. 314–319, 2001.

[24] M. Tokumoto, R. Horiguchi, Y. Nagahama, K. Ishikawa, and T. Tokumoto, "Two proteins, a goldfish 20S proteasome subunit and the protein interacting with 26S proteasome, change in the meiotic cell cycle," *European Journal of Biochemistry*, vol. 267, no. 1, pp. 97–103, 2000.

[25] M. Umeda, Y. Manabe, and H. Uchimiya, "Phosphorylation of the C2 subunit of the proteasome in rice (*Oryza sativa L.*)," *FEBS Letters*, vol. 403, no. 3, pp. 313–317, 1997.

[26] Y. Wakata, M. Tokumoto, R. Horiguchi, K. Ishikawa, Y. Nagahama, and T. Tokumoto, "Identification of α-type subunits of the *Xenopus* 20S proteasome and analysis of their changes during the meiotic cell cycle," *FEBS Letters*, vol. 5, article 18, 2004.

[27] F. Zhang, Y. Hu, P. Huang, C. A. Toleman, A. J. Paterson, and J. E. Kudlow, "Proteasome function is regulated by cyclic AMP-dependent protein kinase through phosphorylation of Rpt6," *Journal of Biological Chemistry*, vol. 282, no. 31, pp. 22460–22471, 2007.

[28] M. Demasi, G. M. Silva, and L. E. S. Netto, "20 S proteasome from *Saccharomyces cerevisiae* is responsive to redox modifications and is S-glutathionylated," *Journal of Biological Chemistry*, vol. 278, no. 1, pp. 679–685, 2003.

[29] S. C. Lee and B. D. Shaw, "A novel interaction between N-myristoylation and the 26S proteasome during cell morphogenesis," *Molecular Microbiology*, vol. 63, no. 4, pp. 1039–1053, 2007.

[30] N. E. Zachara and G. W. Hart, "O-GlcNAc modification: a nutritional sensor that modulates proteasome function," *Trends in Cell Biology*, vol. 14, no. 5, pp. 218–221, 2004.

[31] E. Seemuller, A. Lupas, D. Stock, J. Lowe, R. Huber, and W. Baumeister, "Proteasome from *Thermoplasma acidophilum*: a threonine protease," *Science*, vol. 268, no. 5210, pp. 579–582, 1995.

[32] Y. Feng, D. L. Longo, and D. K. Ferris, "Polo-like kinase interacts with proteasomes and regulates their activity," *Cell Growth and Differentiation*, vol. 12, no. 1, pp. 29–37, 2001.

[33] S. N. Djakovic, L. A. Schwarz, B. Barylko, G. N. DeMartino, and G. N. Patrick, "Regulation of the proteasome by neuronal activity and calcium/calmodulin-dependent protein kinase II," *Journal of Biological Chemistry*, vol. 284, no. 39, pp. 26655–26665, 2009.

[34] H. L. Wilson, H. C. Aldrich, and J. Maupin-Furlow, "Halophilic 20S proteasomes of the archaeon *Haloferax volcanii*: purification, characterization, and gene sequence analysis," *Journal of Bacteriology*, vol. 181, no. 18, pp. 5814–5824, 1999.

[35] M. A. Humbard, S. M. Stevens Jr., and J. A. Maupin-Furlow, "Post-translational modication of the 20S proteasomal proteins of the archaeon *Haloferax volcanii*," *Journal of Bacteriology*, vol. 188, no. 21, pp. 7521–7530, 2006.

[36] S. J. Kaczowka and J. A. Maupin-Furlow, "Subunit topology of two 20S proteasomes from *Haloferax volcanii*," *Journal of Bacteriology*, vol. 185, no. 1, pp. 165–174, 2003.

[37] C. J. Reuter, S. J. Kaczowka, and J. A. Maupin-Furlow, "Differential regulation of the PanA and PanB proteasome-activating nucleotidase and 20S proteasomal proteins of the haloarchaeon *Haloferax volcanii*," *Journal of Bacteriology*, vol. 186, no. 22, pp. 7763–7772, 2004.

[38] M. A. Humbard, G. Zhou, and J. A. Maupin-Furlow, "The N-terminal penultimate residue of 20S proteasome α1 influences its Nα acetylation and protein levels as well as growth rate and stress responses of *Haloferax volcanii*," *Journal of Bacteriology*, vol. 191, no. 12, pp. 3794–3803, 2009.

[39] P. A. Kirkland, J. Busby, S. Stevens Jr., and J. A. Maupin-Furlow, "Trizol-based method for sample preparation and isoelectric focusing of halophilic proteins," *Analytical Biochemistry*, vol. 351, no. 2, pp. 254–259, 2006.

[40] A. J. Barrett, "Fluorimetric assays for cathepsin B and cathepsin H with methylcoumarylamide substrates," *Biochemical Journal*, vol. 187, no. 3, pp. 909–912, 1980.

[41] A. L. Hartman, C. Norais, J. H. Badger, et al., "The complete genome sequence of *Haloferax volcanii* DS2, a model archaeon," *PLoS One*, vol. 5, article e9605, 2010.

[42] N. Blom, S. Gammeltoft, and S. Brunak, "Sequence and structure-based prediction of eukaryotic protein phosphorylation sites," *Journal of Molecular Biology*, vol. 294, no. 5, pp. 1351–1362, 1999.

[43] C. Fischer, C. Geourjon, C. Bourson, and J. Deutscher, "Cloning and characterization of the *Bacillus subtilis prkA* gene encoding a novel serine protein kinase," *Gene*, vol. 168, no. 1, pp. 55–60, 1996.

[44] N. LaRonde-LeBlanc, T. Guszczynski, T. Copeland, and A. Wlodawer, "Autophosphorylation of *Archaeoglobus fulgidus* Rio2 and crystal structures of its nucleotide-metal ion complexes," *FEBS Journal*, vol. 272, no. 11, pp. 2800–2810, 2005.

[45] N. LaRonde-LeBlanc and A. Wlodawer, "A family portrait of the RIO kinases," *Journal of Biological Chemistry*, vol. 280, no. 45, pp. 37297–37300, 2005.

[46] M. A. Gil, K. E. Sherwood, and J. A. Maupin-Furlow, "Transcriptional linkage of *Haloferax volcanii* proteasomal genes with non-proteasomal gene neighbours including RNase P, MOSC domain and SAM-methyltransferase homologues," *Microbiology*, vol. 153, no. 9, pp. 3009–3022, 2007.

[47] B. Wang, S. Yang, L. Zhang, and Z.-G. He, "Archaeal eukaryote-like serine/threonine protein kinase interacts with and phosphorylates a forkhead-associated-domain-containing protein," *Journal of Bacteriology*, vol. 192, no. 7, pp. 1956–1964, 2010.

[48] C. López-Otín and T. Hunter, "The regulatory crosstalk between kinases and proteases in cancer," *Nature Reviews Cancer*, vol. 10, no. 4, pp. 278–292, 2010.

[49] R. Y. Hampton and R. M. Garza, "Protein quality control as a strategy for cellular regulation: lessons from ubiqu it inmediated regulation of the sterol pathway," *Chemical Reviews*, vol. 109, no. 4, pp. 1561–1574, 2009.

[50] F. Silva, E. Navarro, A. Peñaranda, L. Murcia-Flores, S. Torres-Martínez, and V. Garre, "A RING-finger protein regulates carotenogenesis via proteolysis-independent ubiquitylation of a white collar-1-like activator," *Molecular Microbiology*, vol. 70, no. 4, pp. 1026–1036, 2008.

[51] A. Brooun, W. Zhang, and M. Alam, "Primary structure and functional analysis of the soluble transducer protein HtrXI in the archaeon *Halobacterium salinarium*," *Journal of Bacteriology*, vol. 179, no. 9, pp. 2963–2968, 1997.

[52] E. Hildebrand and A. Schimz, "The lifetime of photosensory signals in *Halobacterium halobium* and its dependence on protein methylation," *Biochimica et Biophysica Acta*, vol. 1052, no. 1, pp. 96–105, 1990.

[53] S. Hou, A. Brooun, H. S. Yu, T. Freitas, and M. Alam, "Sensory rhodopsin II transducer HtrII is also responsible for serine chemotaxis in the archaeon *Halobacterium salinarum*," *Journal of Bacteriology*, vol. 180, no. 6, pp. 1600–1602, 1998.

Hindawi Publishing Corporation
Archaea
Volume 2010, Article ID 485051, 11 pages
doi:10.1155/2010/485051

Review Article

Archaea Signal Recognition Particle Shows the Way

Christian Zwieb[1] and Shakhawat Bhuiyan[2]

[1] *Department of Molecular Biology, University of Texas Health Science Center at Tyler, 11937 US Highway 271, Tyler, TX 75708-3154, USA*
[2] *Department of Biology, Division of Business and Sciences, Jarvis Christian College, P.O. Box 1470, Hawkins, TX 75765, USA*

Correspondence should be addressed to Christian Zwieb, zwieb@uthct.edu

Received 15 April 2010; Accepted 14 May 2010

Academic Editor: Jerry Eichler

Archaea SRP is composed of an SRP RNA molecule and two bound proteins named SRP19 and SRP54. Regulated by the binding and hydrolysis of guanosine triphosphates, the RNA-bound SRP54 protein transiently associates not only with the hydrophobic signal sequence as it emerges from the ribosomal exit tunnel, but also interacts with the membrane-associated SRP receptor (FtsY). Comparative analyses of the archaea genomes and their SRP component sequences, combined with structural and biochemical data, support a prominent role of the SRP RNA in the assembly and function of the archaea SRP. The 5e motif, which in eukaryotes binds a 72 kilodalton protein, is preserved in most archaea SRP RNAs despite the lack of an archaea SRP72 homolog. The primary function of the 5e region may be to serve as a hinge, strategically positioned between the small and large SRP domain, allowing the elongated SRP to bind simultaneously to distant ribosomal sites. SRP19, required in eukaryotes for initiating SRP assembly, appears to play a subordinate role in the archaea SRP or may be defunct. The N-terminal A region and a novel C-terminal R region of the archaea SRP receptor (FtsY) are strikingly diverse or absent even among the members of a taxonomic subgroup.

1. Introduction

Protein sorting fundamentally maintains the identity and function of every cell with participation of the signal recognition particle (SRP). SRP components have been found in nearly all organisms [1]. Except in chloroplasts, SRP is a ribonucleoprotein [2]. The SRP RNA is typically composed of about 300 nucleotide residues and forms a complex with an extraordinarily conserved protein named SRP54 in archaea and eukarya or Ffh (fifty-four homolog) in the bacteria. A 19 kDa protein, SRP19, is present in archaea and eukarya, but absent in the bacteria. Polypeptides which are homologous to the eukaryal SRP9/14 and SRP68/72 heterodimers have not been found in the archaea genome sequences giving rise to an archaea SRP which is dominated by RNA [3, 4].

SRP interacts with secretory signal or membrane-anchor sequences upon their emergence from the ribosomal exit tunnel. In vitro and in vivo experiments carried out in eukaryotic protein sorting systems have shown that the SRP delays or blocks the translation of the to-be-targeted polypeptides. Translation resumes when the SRP-bound ribosome nascent chain complex (RNC) binds to the membrane-associated FtsY (filamentous temperature sensitive Y) or, in eukaryotes, the alpha subunit of the SRP receptor (SRα). The interaction between SRP54 and the SR increases the affinity of the proteins for guanosine triphosphate, promotes the release of the signal from the SRP, and interjects the signal sequence of the nascent polypeptide into the protein-conducting channel (PCC) of the cell membrane. Translation and protein translocation or membrane insertion take place during ongoing translation (cotranslational translocation), and, upon hydrolysis of two GTP molecules, the SRP returns to its free cytosolic state ready to initiate another protein targeting cycle (Figure 1) [5–9].

Even though archaea membranes differ significantly from the cell membranes of eukaryotes and bacteria with regard to the use of glycerol-ether lipids and isoprenoid side chains [10, 11], no obvious adaptations for survival under extreme

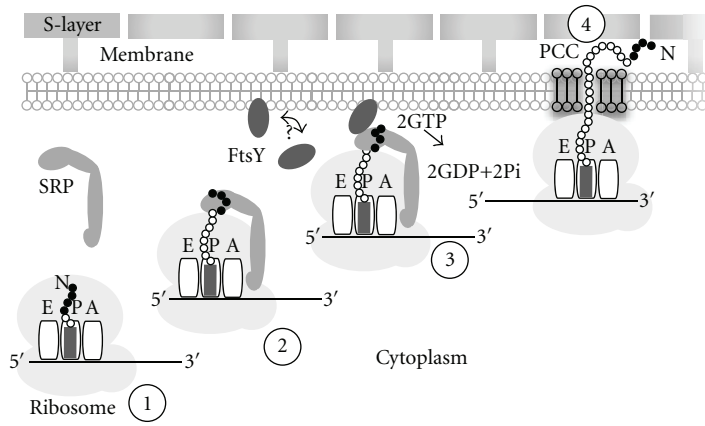


Figure 1: Hypothetical steps in the SRP-mediated targeting of archaea proteins. Step 1: A ribosome (gray, with A, P and E tRNA binding sites) in the cytoplasm translates a mRNA molecule (black, 5′ and 3′ ends are indicated) which encodes a N-terminal signal or membrane-anchor sequence (black dots). Step 2: As the signal emerges from the large ribosomal subunit, it is recognized by the elongated SRP and further translation may be halted. Step 3: The SRP-bound ribosome nascent chain complex (RNC) binds to free or membrane-associated FtsY (arrow). Step 4: After GTP hydrolysis, SRP has been released, translation resumes, and the targeted protein is threaded through the protein-conducting channel (PCC). The surface (S) layer, present in most archaea, is anchored to a glycerol-ether lipids-containing cell membrane.

conditions are apparent in the SRP components. Like bacteria, archaea contain only one SRP receptor polypeptide, FtsY, a homolog of the eukaryal SRα subunit. The signal sequences of the archaea and bacteria are interchangeable [12, 13], and archaea signal peptidases which remove the signal sequence after protein translocation have been identified [14]. Archaea lack homologs of the bacterial and eukaryal translocation ATPases SecA and Kar2p/BiP. They possess, however, Sec61 (the PCC) and a Tat translocase system [15]. These SRP independent means of protein delivery have been discussed recently [16, 17] and will not be reviewed here.

Solving the structures of numerous archaea SRP components and their complexes at high resolution (Table 1) has been crucial for understanding the intricacy of protein targeting in all organisms. Within this structural framework, the increasing number of newly identified archaea genome sequences provides an opportunity to review and discover not only archaea-specific SRP features, but also draw phylogenic distinctions which may pave the way for a better understanding of the function and evolution of every SRP.

2. Archaeal SRP RNA

Unlike the bacterial and eukaryal SRP RNAs, their archaea counterparts vary little in shape and size (approximately 300 nucleotide residues). This may be due to relatively slow evolutionary rates as has been observed previously when the relative conservation of archaea protein sequences was investigated [31]. Archaea SRP RNA secondary structures possess extensively base paired regions which form a prominent central helix flanked by a small (or Alu) and a large (or S) domain (Figure 2(a)). Thus, they resemble the secondary structures of the mammalian SRP RNAs. Helices have been assigned numbers from one to eight, and helical section are designated with letters a to f [32, 33]. The SRP RNAs of most archaea as well as certain bacteria (e.g., Bacilli and Clostridia) pair their terminal regions to form a helix 1. Helix 7 has been found only in eukaryal SRP RNAs where it is most prominent in some fungi and protozoans [1].

Using the previously described sequence identification procedures and covariation rules (see Methods) we aligned 81 archaea SRP RNA sequences and arranged them according to NCBI's taxonomy [35]. The shared alignment pairing mask allows to deduce phylogenetically supported SRP RNA secondary structures for each of the aligned sequences. With a few exceptions, a sequence corresponds to a known species.

The apical loops of SRP RNA helices 3 and 4 form a tertiary interaction which is well supported by covarying compensatory base changes. The UGUNR sequence motif (N is A, C, G or U, R is a purine) located between these helices (labeled UGU in Figure 2(a)) is part of a structurally important U-turn. Both features promote the high degree of compactness of the small SRP domain. It remains to be determined how similar the structure of the protein-free small domain of the archaea SRP is to the solved crystal structure of the mammalian Alu domain in complex with the SRP9/14 protein heterodimer [36].

As previously noted and confirmed by mining of the larger collection of archaea SRP RNA sequences, deviations from the UGUNR motif occur in several groups [37]. Conspicuous erosions of the small domain take place in the SRP RNAs of several Desulfurococcales and in *Nitrosopumilus maritimus* SCM1. Base pairs which typically participate in the formation of helices 1 and 3 are absent in these sequences, while other residues perhaps form an extended helix 4. Due to the relatively small number of available sequences within these subgroups it is not yet possible to conclusively prove or disprove plausible base pairs.

Another hydrogen-bonded tertiary interaction engages two adenosines within the apical tetraloops of helices 6 and 8

Table 1: High-resolution structures of archaeal SRP components. Indicated are the archaea subdomains (Crenarchaeota or Euryarchaeota), species names, components, and methods (X-Ray diffraction of NMR) used for structure determination. The pdb IDs allow easy retrieval of the coordinates [18]. The protein-conducting channel is abbreviated as PCC. Additional nonarchaea SRP high-resolution structures are listed at http://rnp.uthct.edu/rnp/SRPDB/srpstructures.html.

Subdomains	Species	Components	Methods	pdb	References
Crenarchaeota	*Acidianus ambivalens*	SRP54NG	X-Ray	1J8M,1J8Y	[19]
	Sulfolobus solfactaricus	SRP54 with helix 8	X-Ray	1QZW	[20]
		SRP54 dimer	X-Ray	1QZX	[20]
		SRP54 with signal peptide	X-Ray	3KL4	[21]
		SRP19 with helix 6 and helix 8	X-Ray	3KTV,3KTW	[22]
Euryarchaeota	*Archaeoglobus fulgidus*	SRP19	NMR	1KVN,1KVV	[23]
		SRP54M	NMR	2JQE	[24]
	Methanococcus jannaschii	SRP19 with SRP54 with RNA	X-Ray	2V3C	[25]
		SRP19 with helix 6 and helix 8	X-Ray	1LNG	[26]
		Helix 6 and helix 8	X-Ray	1Z43	[27]
		PCC	X-Ray	1RHZ,1RH5	[28]
	Pyrococcus furiosus	SRP19	X-Ray	3DLU,3DLV,3DM5	[29]
		SRP54	X-Ray	3DLU,3DLV,3DM5	[29]
		FtsY	X-Ray	3E70,3DM9,3DMD	[30]

Table 2: Taxonomic distribution of archaea SRP features. Indicated are the archaea subdomains, the number of species identified in each group, and a representative species. Features are the UGUNR motif (N is any nucleotide and R is a purine residue), helices (typically 1 to 4) in the small SRP domain (SD), the GNAR tetranucleotide loop (tetraloop) of helix 6, the GGAA tetraloop of helix 8, proteins SRP19 and SRP54 (SRP19/54), and the acidic (A) domain of the FtsY receptor (FtsY-A). "+" shows presence, "−" absence, and "±" indicates that this feature is present only in a subset of the group members. Sequences deviating from the top-listed motif (e.g., the GGGA in the Thermoproteales) are given. The structural alignments of the 81 identified archaea SRP RNAs and the protein alignments of SRP19, SRP54 and FtsY are provided online as listed in Supplementary Materials 2 available online at doi:10.1155/2010/485051 and are available at http://rnp.uthct.edu/rnp/SRPDB/srprna.html. Features of the SRP RNA in all three domains of life have been recently described in detail in [1].

Subdomains	Groups (n)	Prototypical Species	UGUNR	SD helices	GNAR	GGAA	SRP19/54	FtsY-A
Crenarchaeota	Desulfurococcales (5)	*Pyrodictium occultum*	−	some absent	+	+	+	−
	Sulfolobales (12)	*Sulfolobus solfataricus*	±	+	+	+	+	+, varies
	Thermoproteales (4)	*Thermoproteus neutrophilus V24Sta*	−	+	+	GGGA	+	−
Euryarchaeota	Archaeoglobales (2)	*Archaeoglobus fulgidus*	+	+	+	+	+	±
	Halobacteriales (10)	*Halobacterium species NRC-1*	−	+	+	+	+	±, some long
	Methanobacteriales (3)	*Methanobacterium thermoautotrophicum*	+	+	+	+	+	+, some long
	Methanococcales (12)	*Methanocaldococcus jannaschii DSM 2661*	+	+	varies	+	+	±
	Methanomicrobiales (1)	*Methanoculleus marisnigri JR1*	UAUAA	+	+	+	+	±
	Methanosarcinales (5)	*Methanosarcina acetivorans*	+	+	+	+	+	±
	Methanopyrales (1)	*Methanopyrus kandleri AV19*	+	+	+	GAGA	+	−
	Thermococcales (14)	*Pyrococcus horikoshii OT3*	+	+	+	+	+	±, some long
	Thermoplasmatales (3)	*Thermoplasma acidophilum*	−	+	AAAG	+	+	−
Korarchaeota	Candidatus Korarchaeum (1)	*Candidatus Korarchaeum cryptofilum OPF8*	−	+	+	+	+	−
Thaumarchaeota	Nitrosopumilales (1)	*Nitrosopumilus maritimus SCM1*	−	some absent	+	GGGA	+	−

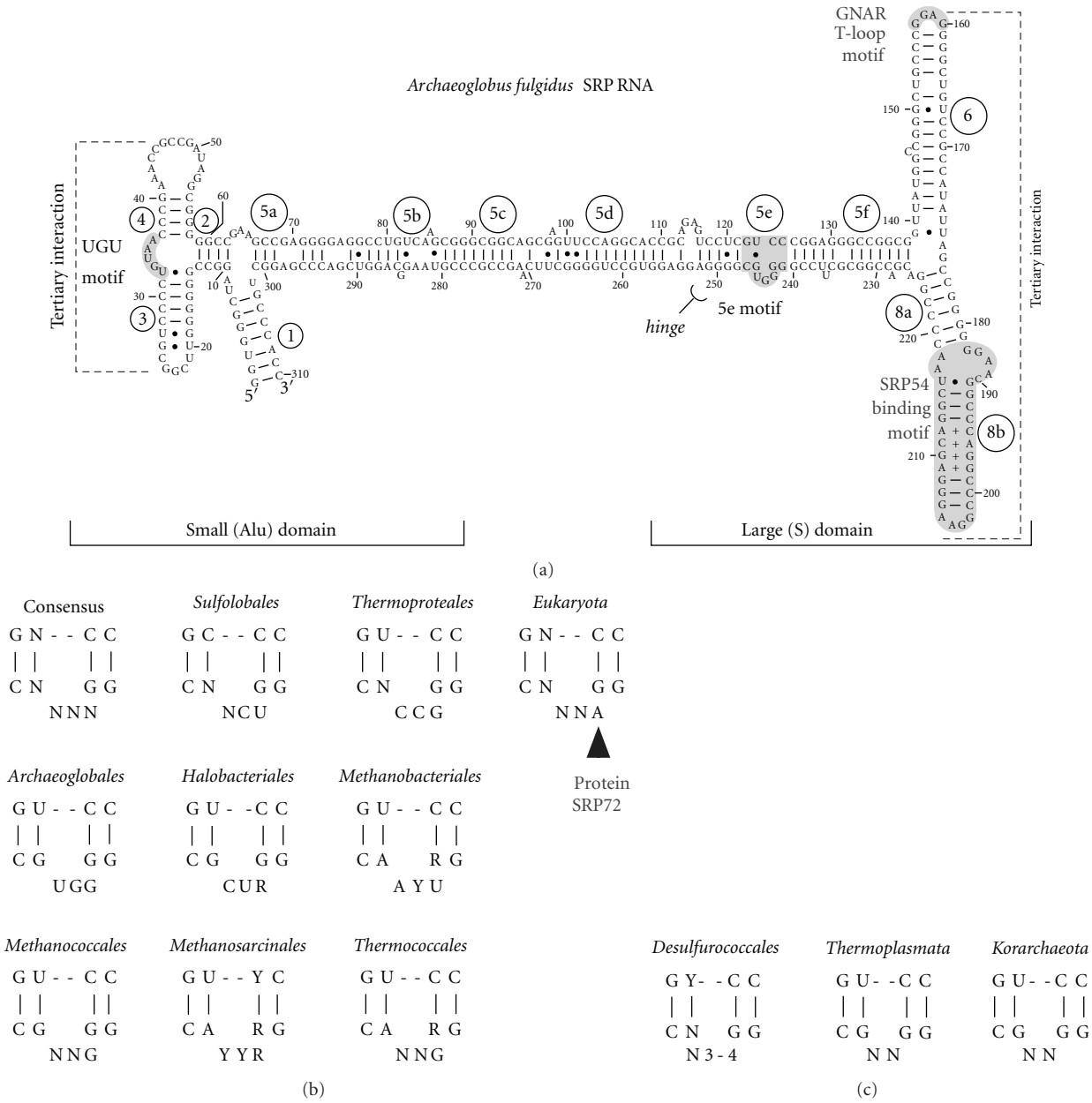


Figure 2: Archaea SRP RNA features. (a) Secondary structure of *Archaeoglobus fulgidus* SRP RNA. (b) Consensus-matching 5e motifs sorted by their taxonomic membership. (c) Mismatching 5e motifs. Helices are numbered from 1 to 8 and helical sections are labeled with letters a to f [33]. For example, helix 1 is composed of residues one to seven which are base paired with residues 303 to 310; helix 2 consists of residues at positions ten to 13 base paired with the residues at positions 59 to 62. The extended helix 5 contains six helical sections, 5a to 5f. Helix 7 is lacking in the SRP RNAs of the archaea. The 5′- and 3′-ends are shown, and residues are labeled in ten-residue increments. Base pairs were determined by comparative sequence analysis [32] and by considering high-resolution data (Table 1). The approximate extents of the large (S) and the small (Alu) domains are indicated. Shown in dark gray are the UGUNR motif (labeled UGU) in the small domain, the 5e motif within helical section 5e at the indicated hinge [34], the GNAR apical tetraloop of helix 6 and the SRP54 binding motif of helix 8 in the large domain. Dashed lines suggest tertiary interactions.

(A159 and A205 in *Archaeoglobus fulgidus*, Figure 2(a)). This long-range interaction was first seen in the crystal structures of *Methanococcus jannaschii* SRP RNA from the large domain (Table 1). The adenosine clamp severely constraints the arrangement of helices 6 and 8. It is highly conserved and likely exists in all archaea and eukaryotic SRP RNAs. The participating adenosine of helix 6 is presented within a GNAR tetranucleotide loop (tetraloop) in most archaea SRP RNAs, but deviates (AAAG) from the consensus in the four SRP RNA sequences of the Thermoplasmatales. The

interacting helix 8 has a GRRA loop with GGAA being the most frequently represented tetranucleotide. GGGA is found in the Thermoproteales and Thaumarchaeota (Nitrosopumilales), and GAGA in the Methanopyrales. These helix 8 tetraloop sequences are probably useful when attempting to identify and classify the archaea SRP RNAs (Table 2).

3. The 5e Motif: A Case for Molecular Exaptation

The 11-nucleotide 5e element is the most recently discovered SRP RNA motif and has been helpful in the prediction of SRP RNA genes [38]. The motif consists of four base pairs interrupted by a three-nucleotide loop. Two of the base pairs are symmetrically arranged G-C pairs. The comparison of 141 eukaryal and 28 archaea sequences shows that the first residue of the eukaryotic 5e loop is a conserved adenosine (A240 in human SRP RNA) in the eukarya (Figure 2(b)) [39]. In the archaea, the corresponding nucleotide can be any residue, and only two halobacterial sequences (*Haloferax volcanii*, GenBank Accession AF395888, and *Halomicrobium mukohataei*, GenBank Accession CP001688) possess an adenosine.

Systematic site-directed mutagenesis of the 5e region showed that human SRP RNA with a single A240G change was unable to form a complex with full-length human SRP72 [39]. The 5e RNA was found to bind a 56 amino acid-residue polypeptide of human SRP72 which contained the consensus sequence PDPXRWLPXXER (X is for any amino acid residue) [40]. Bioinformatic analyses identified two relatively poor consensus sequence matches in the genomes of archaea, one with a methyl coenzyme M reductase of an uncultured methanogenic archaeon (GenBank Accession ABI18429), the other with a hypothetical protein of *Pyrobaculum islandicum* DSM 4184 (GenBank Accession ABL88435). These relationships are likely coincidental and, until proven otherwise, are consistent with the notion that a functional equivalent of the eukaryotic SRP72 is lacking in the archaea.

The conserved adenosine in the 5e motif of the eukaryal, but not the archaeal SRP RNAs suggests that the 5e element was recruited in evolution to supply a new function to the protein-rich eukaryotic SRP thereby providing a striking example for molecular exaptation, defined as the utilization of a feature for a function which differs from what it was originally developed for [41, 42]. Because human SRP72 binds strongly to the *Haloferax volcanii* SRP RNA [40], the structures of the 5e region of archaea and eukaryotes are apparently very similar.

The 5e RNA fragment is remarkably resistant towards ribonucleolytic attack [39] indicating that it is compactly folded and may resemble the structure of an RNA kink-turn [43]. Although 5e conforms only loosely to the K-turn consensus secondary structure, 3D molecular modeling demonstrates compatible structures (Zwieb, unpublished). This suggest that 5e is part of the bend or hinge which allows the elongated SRP to adjust to the curvature of the ribosome and bind simultaneously to separate ribosomal sites [34]. Such an interpretation is supported by the finding that 5e is present in SRP RNAs with a standard set of helices in their small SRP domain [38]. Conversely, Figure 2(c) and the data shown in Table 2 suggest that SRP RNAs deviate from the 5e consensus when they lack the UGUNR motif or when the small SRP domain is eroded. These hinge-impaired archaea SRP RNAs may function in a mode which resembles the SRP-mediated protein targeting of the majority of bacteria which lack the small SRP domain.

4. Protein SRP19, Is It Required?

Although protein SRP19 was thought to be absent in certain archaea genomes [3], its genes (91 sequences) have now been identified in all archaea subgroups (Table 2). SRP19 coexist with SRP RNA helix 6 as part of the large SRP domain. Mainly due to the reduced size of its loop 4, the archaea SRP19 is generally somewhat shorter than its eukaryotic homolog (Figure 3, top, gray triangle). The NMR structure of *Archaeoglobus fulgidus* has been solved [23], and several crystal structures of the free and RNA-bound SRP19 have been determined (Table 1) revealing a single-domain compactly folded protein.

Certain conserved amino acid residues (Y/W and GR in loop 1; Figure 3, top) participate in the binding to the SRP RNA through induced fit mechanisms involving both the protein and the RNA. For example, loop 3 (Figure 3, top) of *Archaeoglobus fulgidus* SRP19 reorders and adopts a single conformation upon binding to RNA [23]. In the Thermococcales, loop 3 is enlarged and disordered and, upon binding, rearranges to assist in the proper folding of the SRP RNA [29]. This mechanism of mutual conformational adjustment has been observed in several other protein-RNA complexes [44].

In eukaryotic cells, SRP is assembled in the nucleolus and transported to the cytosol where it associates with SRP54 [45, 46]. Archaea SRPs contain only two proteins, SRP19 and SRP54, and assemble in the cytosol. The mammalian SRP19 is required to position SRP RNA helices 6 and 8 in a side-by-side fashion and expose the SRP54 binding site through a conformational collapse in helix 8. In contrast, archaea SRP RNA binds SRP54 even in the absence of SRP19 [47, 48]. RNase susceptibility measurements of wild-type and mutant *Archaeoglobus fulgidus* SRP RNAs show that the conserved adenosine of the GNAR tetraloop in helix 6, and not SRP19, is responsible for a compactly arranged large SRP domain [49]. Indeed, helices 6 and 8 are closely packed in the protein-free crystal structures of *Methanococcus jannaschii* and *Sulfolobus solfataricus* SRP RNAs [22, 27].

Figure 4 indicates that helix 6 and helix 8 interact with each other not only through their distal tetraloop adenosines but also via internal looped-out residues. However, the asymmetric internal loop of helix 8 engages in distinctly different ways. In the human SRP RNA, two adenosines protrude from the short strand of the asymmetric loop to form A-minor motifs with helix 6 [50]. In contrast, in the *Methanococcus jannaschii* RNA structures, two adenosines of

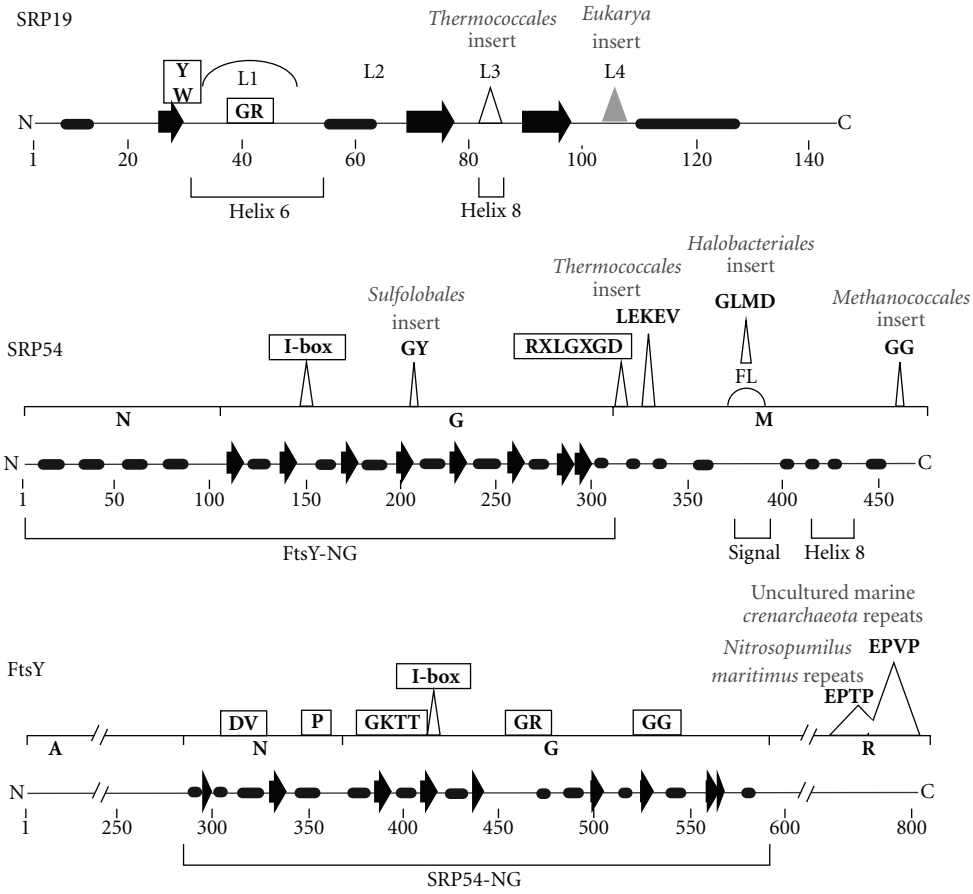


Figure 3: Features of SRP19 (top), SRP54 (center) and the FtsY SRP receptor (bottom). Indicated are the N- and C-termini. Helices are shown as cylinders, beta-sheets as arrows, and some loops are labeled with an arch and the letter L. Invariant or highly conserved residues are shown boxed, and several amino acid residue inserts which are characteristic for the indicated taxonomic groups are shown as bold letters. Numbering is according to the column positions of each protein alignment accessible as listed in Supplementary Materials 2 and at the SRP database at http://rnp.uthct.edu/rnp/SRPDB/SRPDB.html). Regions and sites which interact with other SRP components or the signal sequence are marked with brackets below each panel.

helix 6 are bulged out and interact in the minor groove of helix 8 [25].

Deletion of the yeast SRP19 homolog Sec65 was shown to be lethal to the eukaryote *Yarrowia lipolytica* [51]. In the archaea, structural and biochemical data as well as the deviation from the GNAR tetraloop motif observed within the Thermoplasmatales (Table 2) suggest that SRP19 is not required for SRP assembly and dispensable for protein sorting and survival. In fact, deletion of SRP19 from the *Haloferax volcanii* genome had no effect on protein translocation or membrane insertion. Increased levels of membrane bacterioruberin were detected in the deletion mutant and significant amounts of SRP19 mRNA were observed in nonmutated cells [52] suggesting a relatively minor possibly regulatory function for SRP19. Although the protein might participate in a more substantive way when *Haloferax volcanii* is challenged to survive in external environments, the data demonstrate the diminished importance of the archaea SRP19 when compared to its significant role for the survival of eukaryotic cells.

5. SRP54

SRP54, or its bacterial homolog Ffh, is present in all organisms, including the chloroplast SRPs which lack an SRP RNA [2]. Deletion of the *Haloferax volcanii* SRP54 gene results in the loss of cell viability as proof of the central role of SRP54 in archaea protein targeting [52, 53]. Sequence and three-dimensional structure (Table 1) of the protein are highly conserved. These properties are readily explained by the numerous interactions which engage SRP54 in the binding not only to the SRP RNA, but also the signal sequence and the FtsY SRP receptor. The observed exceptionally high level of conservation likely reflects the need to carry out multiple binding reactions in a coordinated dynamically GTP-regulated way to ensure proper and efficient delivery of a wide variety of signal sequence-tagged proteins into the PCC.

The functions of SRP54 are brought about by three domains. The N-terminal (N) domain is composed of a bundle of four alpha helices, the GTPase (G) domain

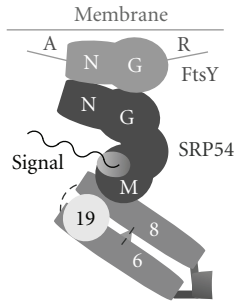


Figure 4: Interactions between the components of archaea SRP-mediated protein targeting. Schematic drawing of the coaxially-arranged SRP RNA helices 6 and 8 bound together by SRP19 and two tertiary interactions (dashed lines). The M-domain of the SRP54 protein (dark gray) binds to SRP RNA helix 8 as well as the signal sequence (black). The NG-domains of SRP54 and the FtsY SRP receptor are arranged quasisymmetrically and poised to separate upon the hydrolysis of two G-domain-bound GTP molecules. The N-terminal region labeled A (for acidic) and the C-terminal repeat region (R) of FtsY are variable or may be absent (see Table 2).

contains a unique insertion (I-box) which serves as a guanine nucleotide-exchange factors (GEFs) and stabilizes the nucleotide free protein [54, 55], and the methionine-rich (M) domain binds to the SRP RNA and the signal sequence (Figure 3, center). The predominantly alpha helical M domain contains an extended segment (the so-called fingerloop) which delineates or is folded into a groove which accepts signal sequences [20, 21, 56]. This wide and short hydrophobic groove was observed also in the crystal structure of the RNA-bound *Escherichia coli* Ffh [57]. The NMR structure of the *Archaeoglobus fulgidus* SRP54 M domain [24] is similar to these crystal structures and disfavors another proposed mode whereby the signal sequence binds within a long and narrow groove of SRP54M [28, 50]. The conformations of the fingerloop in solution suggest that it adaptively binds and stabilizes the signal sequences. Binding is weak [24] and likely reversible in order to permit signal sequence release upon the binding of SRP54 to the SRP receptor. The molecular details of the contacts made by a signal peptide with the *Sulfolobus solfataricus* SRP54 have been revealed recently and suggest that portions of the fingerloop may adopt an alpha helical conformation [21].

Adding to the intricacy of signal sequence recognition, the M domain and the NG region of SRP54 are joined together via a flexible linker. This region has the consensus sequence RXLGXGD and allows the RNA-bound SRP54 to undergo substantial structural rearrangements upon binding to a signal sequence [20, 22]. Consistent with this assertion, site-directed mutagenesis experiments of mammalian SRP [58] and a recent crosslinking study of the *Escherichia coli* SRP [59] demonstrate the involvement of the signal sequence not only with the M domain, but also the NG region. No evidence for the binding of NG to signal sequences has been provided in the archaea. However, the exceptional conservation of SRP54 throughout all domains of life suggests that archaea employ a similar if not identical signal recognition mechanism. The NG region can be in close proximity to SRP RNA helix 8 and, in archaea, appears to engage also helix 6 [25].

The alignment of 103 archaea SRP54 sequences reveals several group-specific amino acid residue insertions, for example a GY in the G domain of Sulfolobales which might modulate the GTPase activity. Into the M domain, Thermococcales insert the sequence LEKEV, Halobacteriales GLMD, and Methanococcales GG (Figure 3). These amino acid residues have the potential to contribute to the binding of the protein to the SRP RNA, to signal peptide recognition or other yet to be specified enhanced functions. Regardless of their potential significance, these short peptide sequences are useful for assigning SRP54 sequences to their proper taxonomic group.

6. FtsY: The SRP Receptor

The SRP receptor (SR) of the eukarya is composed of the peripheral membrane SRα and the integral membrane SRβ proteins. Bacteria and archaea possess only FtsY, a homolog of SRα [60]. Sequence comparisons of FtsY with SRP54 suggest a gene duplication event [61] and support the classification into the three domains of life as well as the close rooting of archaea and eukarya [62]. Archaea FtsY shares its conserved NG region with NG of SRP54, including the I-box, but differs from SRP54 with respect to several short amino acid stretches as revealed by the alignment of 95 archaea FtsY sequences (Figure 3, Supplementary Material 1). The NG regions are symmetrically arranged in three dimensions to constitute the structural and functional core of signal sequence release and nascent polypeptide delivery into the cell membrane (Figure 4) by mutually catalyzing the hydrolysis of GTP [63–65].

As has been observed within the bacterial genomes [66, 67] several archaea FtsY sequences consist only of the NG domain and lack an N-terminal acidic (A) domain. Diversity with respect to the A domain is observed even within a single archaea subgroup (Table 2). Full-length *Haloferax volcanii* FtsY as well as polypeptides lacking the A domain were shown to bind to inverted membrane vesicles indicating that the A domain is dispensable for attaching FtsY to the membrane. Instead, the A domain may play a role in recruiting SRP to the haloarchaeal membrane [68, 69]. Assuming a pool of free FtsY in the cytosol [70, 71] (Figure 1) these findings are particularly relevant. On the other hand, fluorescence microscopy showed that almost all of the *Escherichia coli* FtsY associates in vivo with the inner membrane, and any soluble FtsY is unlikely to contribute to protein targeting [72]. Although archaea FtsY might interact with the membrane in similar manners as has been observed in bacteria and chloroplast [2, 73–76], the molecular details of the binding could be quite different given the differences in membrane lipid composition. FtsY might also interact directly with a cytosolically exposed portion of the PCC [77, 78]. In either case, one would

expect functional synchronicity between GTP hydrolysis and delivery of protein into the PCC [79].

In the FtsY sequences of the uncultured marine Crenarchaeota we discovered a C-terminal proline-rich extension, named R for its motif repetitions (see Figures 3 and 4). Up to 12 EPVP repeats (accession numbers ABZ10052, ABZ08863, ABZ09152, ABZ09615) and five EPVV repeats (ABZ098531) were present in the R region. Similar multiple repeats with the sequence EPTP were seen also in the FtsY of the Thaumarchaeotum *Nitrosopumilus maritimus* SCM1. Details of the R-regions can be inspected in an updated archaeal FtsY alignment provided at the SRPDB [37]. As with much of our limited understanding of the role of FtsY in the archaea, it remains to be determined if these repeats are expressed and have a function in protein export.

7. Archaea SRP Function and Evolution

During the past years, several interesting puzzle pieces with respect to SRP-mediated protein translocation and membrane insertion in the archaea have been assembled. The SRPs of the Crenarchaeotum *Acidianus ambivalens* and the Euryarchaeota *Archaeoglobus fulgidus*, *Pyrococcus furiosus* and *Haloferax volcanii* have been reconstituted [47, 48, 80–82], and the ability of an archaea SRP54 to participate in signal sequence recognition has been demonstrated [81]. Nevertheless, the role of SRP within the archaeal cell is still poorly understood. Examples of both protein synthesis-linked (cotranslational) and posttranslational translocation have been provided [83–87], but to what degree these findings are representative remains to be investigated further [16].

The proposal that signal sequences might interact with the SRP RNA has fed the imagination that the primitive SRP was composed only of RNA [57, 88]. However, because of the proteinaceous nature of the signal, a scenario in which SRP RNA coemerged with evolutionary precursors of SRP54/Ffh/FtsY appears to be more plausible. Furthermore, the recent structure of the signal peptide-bound *Sulfolobus solfataricus* SRP54 (Ffh) shows that the signal peptide is too far removed from the SRP RNA to make direct contact [21].

If the small (Alu) SRP domain was a feature of the primitive SRP which subsequently was lost in evolutionary time; the majority of the bacteria is more difficult to discern. As another possibility archaea and certain bacteria may have been faced independently with the need to enlarge a small primitive SRP, maybe to slow down translation rates and provide more time for ensuring the delivery of proteins to the membrane as has been observed in eukarya [89].

8. Future Directions

With respect to the RNA-rich archaea SRP it would be desirable to better understand the structure and function of the protein-free small SRP domain. For example, what, in molecular detail, allows the small domain to fold back onto helix 5 in order to approximate the shape and dimensions of the eukaryal SRP [36]? What is the functional significance of the conserved 5e motif and its relationship to a flexible hinge or a bend in the elongated SRP? It will also be important to further elucidate the role of the archaea FtsY, its role in the cytosol as well as the molecular features which promote its association with archaea membranes. As in the past, the studies of the archaea SRP are expected to contribute in many ways to our grasp of SRP-mediated protein targeting in all organism.

9. Methods

Sets of representative sequences were used as input to Perl scripts written to identify sequence homologs in the NCBI databases [90]. RNA sequences were aligned semiautomatically with SARSE [91]; protein sequences were aligned using MUSCLE [92] followed by manual adjustments in Jalview [93]. The alignments are available through the links listed in Supplementary Material 2. In addition, the SRP database provides tables of alphabetically and phylogenetically sorted sequences at http://rnp.uthct.edu/rnp/SRPDB/SRPDB.html.

Acknowledgments

The authors are grateful to the American Heart Association, South Central Affiliate, for support (no. 09GRNT2080038). S.Bhuiyan is supported by a UNIMET grant from the National Aeronautics and Space Administration and the United Negro College Fund Special Programs Corporation NASA Science and Technology Institute for Minority Institution Project NNA06CB14H.

References

[1] M. A. Rosenblad, N. Larsen, T. Samuelsson, and C. Zwieb, "Kinship in the SRP RNA family," *RNA Biology*, vol. 6, no. 5, pp. 508–516, 2009.

[2] K. F. Stengel, I. Holdermann, P. Cain, C. Robinson, K. Wild, and I. Sinning, "Structural basis for specific substrate recognition by the chloroplast signal recognition particle protein cpSRP43," *Science*, vol. 321, no. 5886, pp. 253–256, 2008.

[3] C. Zwieb and J. Eichler, "Getting on target: the archaeal signal recognition particle," *Archaea*, vol. 1, no. 1, pp. 27–34, 2002.

[4] R. G. Moll, "The archaeal signal recognition particle: steps toward membrane binding," *Journal of Bioenergetics and Biomembranes*, vol. 36, no. 1, pp. 47–53, 2004.

[5] H. G. Koch, M. Moser, and M. Müller, "Signal recognition particle-dependent protein targeting, universal to all kingdoms of life," *Reviews of Physiology, Biochemistry and Pharmacology*, vol. 146, pp. 55–94, 2003.

[6] K. Nagai, C. Oubridge, A. Kuglstatter, E. Menichelli, C. Isel, and L. Jovine, "Structure, function and evolution of the signal recognition particle," *EMBO Journal*, vol. 22, no. 14, pp. 3479–3485, 2003.

[7] J. A. Doudna and R. T. Batey, "Structural insights into the signal recognition particle," *Annual Review of Biochemistry*, vol. 73, pp. 539–557, 2004.

[8] M. Halic and R. Beckmann, "The signal recognition particle and its interactions during protein targeting," *Current Opinion in Structural Biology*, vol. 15, no. 1, pp. 116–125, 2005.

[9] S.-O. Shan and P. Walter, "Co-translational protein targeting by the signal recognition particle," *FEBS Letters*, vol. 579, no. 4, pp. 921–926, 2005.

[10] G. D. Sprott, "Structures of archaebacterial membrane lipids," *Journal of Bioenergetics and Biomembranes*, vol. 24, no. 6, pp. 555–566, 1992.

[11] J. L. C. M. Van de Vossenberg, A. J. M. Driessen, and W. N. Konings, "The essence of being extremophilic: the role of the unique archaeal membrane lipids," *Extremophiles*, vol. 2, no. 3, pp. 163–170, 1998.

[12] G. von Heijne, "The signal peptide," *Journal of Membrane Biology*, vol. 115, no. 3, pp. 195–201, 1990.

[13] M. Pohlschröder, M. I. Giménez, and K. F. Jarrell, "Protein transport in Archaea: Sec and twin arginine translocation pathways," *Current Opinion in Microbiology*, vol. 8, no. 6, pp. 713–719, 2005.

[14] S. Y. M. Ng, B. Chaban, D. J. VanDyke, and K. F. Jarrell, "Archaeal signal peptidases," *Microbiology*, vol. 153, no. 2, pp. 305–314, 2007.

[15] S.-V. Albers, Z. Szabó, and A. J. M. Driessen, "Protein secretion in the Archaea: multiple paths towards a unique cell surface," *Nature Reviews Microbiology*, vol. 4, no. 7, pp. 537–547, 2006.

[16] D. Calo and J. Eichler, "Crossing the membrane in Archaea, the third domain of life," *Biochimica et Biophysica Acta*. In press.

[17] J. Yuan, J. C. Zweers, J. M. van Dijl, and R. E. Dalbey, "Protein transport across and into cell membranes in bacteria and archaea," *Cellular and Molecular Life Sciences*, vol. 67, no. 2, pp. 179–199, 2010.

[18] A. Kouranov, L. Xie, J. de la Cruz et al., "The RCSB PDB information portal for structural genomics," *Nucleic Acids Research*, vol. 34, pp. D302–305, 2006.

[19] G. Montoya, K. te Kaat, R. Moll, G. Schäfer, and I. Sinning, "The crystal structure of the conserved GTPase of SRP54 from the archaeon *Acidianus ambivalens* and its comparison with related structures suggests a model for the SRP-SRP receptor complex," *Structure*, vol. 8, no. 5, pp. 515–525, 2000.

[20] K. R. Rosendal, K. Wild, G. Montoya, and I. Sinning, "Crystal structure of the complete core of archaeal signal recognition particle and implications for interdomain communication," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 100, no. 25, pp. 14701–14706, 2003.

[21] C. Y. Janda, J. Li, C. Oubridge, H. Hernández, C. V. Robinson, and K. Nagai, "Recognition of a signal peptide by the signal recognition particle," *Nature*, vol. 465, no. 7297, pp. 507–510, 2010.

[22] K. Wild, G. Bange, G. Bozkurt, B. Segnitz, A. Hendricks, and I. Sinning, "Structural insights into the assembly of the human and archaeal signal recognition particles," *Acta Crystallographica Section D*, vol. 66, no. 3, pp. 295–303, 2010.

[23] O. N. Pakhomova, S. Deep, Q. Huang, C. Zwieb, and A. P. Hinck, "Solution structure of protein SRP19 of *Archaeoglobus fulgidus* signal recognition particle," *Journal of Molecular Biology*, vol. 317, no. 1, pp. 145–158, 2002.

[24] U. Ilangovan, S. H. Bhuiyan, C. S. Hinck et al., "A. fulgidus SRP54 M-domain," *Journal of Biomolecular NMR*, vol. 41, no. 4, pp. 241–248, 2008.

[25] T. Hainz, S. Huang, and A. E. Sauer-Eriksson, "Interaction of signal-recognition particle 54 GTPase domain and signal-recognition particle RNA in the free signal-recognition particle," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 104, no. 38, pp. 14911–14916, 2007.

[26] T. Hainzl, S. Huang, and A. E. Sauer-Eriksson, "Structure of the SRP19-RNA complex and implications for signal recognition particle assembly," *Nature*, vol. 417, no. 6890, pp. 767–771, 2002.

[27] T. Hainzl, S. Huang, and A. E. Sauer-Eriksson, "Structural insights into SRP RNA: an induced fit mechanism for SRP assembly," *RNA*, vol. 11, no. 7, pp. 1043–1050, 2005.

[28] W. M. Clemons Jr., K. Gowda, S. D. Black, C. Zwieb, and V. Ramakrishnan, "Crystal structure of the conserved subdomain of human protein SRP54M at 2.1 Å resolution: evidence for the mechanism of signal peptide binding," *Journal of Molecular Biology*, vol. 292, no. 3, pp. 697–705, 1999.

[29] P. F. Egea, J. Napetschnig, P. Walter, and R. M. Stroud, "Structures of SRP54 and SRP19, the two proteins that organize the ribonucleic core of the signal recognition particle from *Pyrococcus furiosus*," *PLoS One*, vol. 3, no. 10, Article ID e3528, 2008.

[30] P. F. Egea, H. Tsuruta, G. P. de Leon, J. Napetschnig, P. Walter, and R. M. Stroud, "Structures of the signal recognition particle receptor from the Archaeon *Pyrococcus furiosus*: Implications for the targeting step at the membrane," *PLoS One*, vol. 3, no. 11, Article ID e3619, 2008.

[31] J. M. Kollman and R. F. Doolittle, "Determining the relative rates of change for prokaryotic and eukaryotic proteins with anciently duplicated paralogs," *Journal of Molecular Evolution*, vol. 51, no. 2, pp. 173–181, 2000.

[32] N. Larsen and C. Zwieb, "SRP-RNA sequence alignment and secondary structure," *Nucleic Acids Research*, vol. 19, no. 2, pp. 209–215, 1991.

[33] C. Zwieb, R. W. Van Nues, M. A. Rosenblad, J. D. Brown, and T. Samuelsson, "A nomenclature for all signal recognition particle RNAs," *RNA*, vol. 11, no. 1, pp. 7–13, 2005.

[34] M. Halic, T. Becker, M. R. Pool et al., "Structure of the signal recognition particle interacting with the elongation-arrested ribosome," *Nature*, vol. 427, no. 6977, pp. 808–814, 2004.

[35] D. A. Benson, I. Karsch-Mizrachi, D. J. Lipman, J. Ostell, and E. W. Sayers, "GenBank," *Nucleic Acids Research*, vol. 37, database issue, pp. D26–D31, 2009.

[36] O. Weichenrieder, K. Wild, K. Strub, and S. Cusack, "Structure and assembly of the Alu domain of the mammalian signal recognition particle," *Nature*, vol. 408, no. 6809, pp. 167–173, 2000.

[37] E. S. Andersen, M. A. Rosenblad, N. Larsen et al., "The tmRDB and SRPDB resources," *Nucleic acids research.*, vol. 34, pp. D163–168, 2006.

[38] M. Regalia, M. A. Rosenblad, and T. Samuelsson, "Prediction of signal recognition particle RNA genes," *Nucleic Acids Research*, vol. 30, no. 15, pp. 3368–3377, 2002.

[39] E. Iakhiaeva, J. Wower, I. K. Wower, and C. Zwieb, "The 5e motif of eukaryotic signal recognition particle RNA contains a conserved adenosine for the binding of SRP72," *RNA*, vol. 14, no. 6, pp. 1143–1153, 2008.

[40] E. Iakhiaeva, J. Yin, and C. Zwieb, "Identification of an RNA-binding domain in human SRP72," *Journal of Molecular Biology*, vol. 345, no. 4, pp. 659–666, 2005.

[41] C. Darwin, "Modes of transition," in *The Origin of Species*, chapter 6, John Murray, London, UK, 1859.

[42] S. J. Gould and E. S. Vrba, "Exaptation—a missing term in the science of form," *Paleobiology*, vol. 8, no. 1, pp. 4–15, 1982.

[43] S. Nolivos, A. J. Carpousis, and B. Clouet-d'Orval, "The K-loop, a general feature of the Pyrococcus C/D guide RNAs, is an RNA structural motif related to the K-turn," *Nucleic Acids Research*, vol. 33, no. 20, pp. 6507–6514, 2005.

[44] R. Stefl, L. Skrisovska, and F. H.-T. Allain, "RNA sequence- and shape-dependent recognition by proteins in the ribonucleoprotein particle," *EMBO Reports*, vol. 6, no. 1, pp. 33–38, 2005.

[45] J. C. Politz, S. Yarovoi, S. M. Kilroy, K. Gowda, C. Zwieb, and T. Pederson, "Signal recognition particle components in the nucleolus," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 97, no. 1, pp. 55–60, 2000.

[46] T. S. Maity and K. M. Weeks, "A threefold RNA-protein interface in the signal recognition particle gates native complex assembly," *Journal of Molecular Biology*, vol. 369, no. 2, pp. 512–524, 2007.

[47] H. Maeshima, E. Okuno, T. Aimi, T. Morinaga, and T. Itoh, "An archaeal protein homologous to mammalian SRP54 and bacterial Ffh recognizes a highly conserved region of SRP RNA," *FEBS Letters*, vol. 507, no. 3, pp. 336–340, 2001.

[48] I. Tozik, Q. Huang, C. Zwieb, and J. Eichler, "Reconstitution of the signal recognition particle of the halophilic archaeon Haloferax volcanii," *Nucleic Acids Research*, vol. 30, no. 19, pp. 4166–4175, 2002.

[49] J. Yin, Q. Huang, O. N. Pakhomova, A. P. Hinck, and C. Zwieb, "The conserved adenosine in helix 6 of Archaeoglobus fulgidus signal recognition particle RNA initiates SRP assembly," *Archaea*, vol. 1, no. 4, pp. 269–275, 2004.

[50] A. Kuglstatter, C. Oubridge, and K. Nagai, "Induced structural changes of 7SL RNA during the assembly of human signal recognition particle," *Nature Structural Biology*, vol. 9, no. 10, pp. 740–744, 2002.

[51] M. Sánchez, J.-M. Beckerich, C. Gaillardin, and A. Domínguez, "Isolation and cloning of the Yarrowia lipolytica SEC65 gene, a component of the yeast signal recognition particle displaying homology with the human SRP19 gene," *Gene*, vol. 203, no. 1, pp. 75–84, 1997.

[52] S. Yurist, I. Dahan, and J. Eichler, "SRP19 is a dispensable component of the signal recognition particle in Archaea," *Journal of Bacteriology*, vol. 189, no. 1, pp. 276–279, 2007.

[53] R. W. Rose and M. Pohlschröder, "In vivo analysis of an essential archaeal signal recognition particle in its native host," *Journal of Bacteriology*, vol. 184, no. 12, pp. 3260–3267, 2002.

[54] C. Moser, O. Mol, R. S. Goody, and I. Sinning, "The signal recognition particle receptor of Escherichia coli (FtsY) has a nucleotide exchange factor built into the GTPase domain," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 94, no. 21, pp. 11339–11344, 1997.

[55] P. J. Rapiejko and R. Gilmore, "Empty site forms of the SRP54 and SRα GTPase mediate targeting of ribosome-nascent chain complexes to the endoplasmic reticulum," *Cell*, vol. 89, no. 5, pp. 703–713, 1997.

[56] R. J. Keenan, D. M. Freymann, P. Walter, and R. M. Stroud, "Crystal structure of the signal sequence binding subunit of the signal recognition particle," *Cell*, vol. 94, no. 2, pp. 181–191, 1998.

[57] R. T. Batey, R. P. Rambo, L. Lucast, B. Rha, and J. A. Doudna, "Crystal structure of the ribonucleoprotein core of the signal recognition particle," *Science*, vol. 287, no. 5456, pp. 1232–1239, 2000.

[58] J. A. Newitt and H. D. Bernstein, "The N-domain of the signal recognition particle 54-kDa subunit promotes efficient signal sequence binding," *European Journal of Biochemistry*, vol. 245, no. 3, pp. 720–729, 1997.

[59] E. M. Clérico, A. Szymanska, and L. M. Gierasch, "Exploring the interactions between signal sequences and E. coli SRP by two distinct and complementary crosslinking methods," *Biopolymers*, vol. 92, no. 3, pp. 201–211, 2009.

[60] J. D. Miller, H. D. Bernstein, and P. Walter, "Interaction of E. coli Ffh/4.5S ribonucleoprotein and FtsY mimics that of mammalian signal recognition particle and its receptor," *Nature*, vol. 367, no. 6464, pp. 657–659, 1994.

[61] S. Althoff, D. Selinger, and J. A. Wise, "Molecular evolution of SRP cycle components: functional implications," *Nucleic Acids Research*, vol. 22, no. 11, pp. 1933–1947, 1994.

[62] S. Gribaldo and P. Cammarano, "The root of the universal tree of life inferred from anciently duplicated genes encoding components of the protein-targeting machinery," *Journal of Molecular Evolution*, vol. 47, no. 5, pp. 508–516, 1998.

[63] P. F. Egea, S.-O. Shan, J. Napetschnig, D. F. Savage, P. Walter, and R. M. Stroud, "Substrate twinning activates the signal recognition particle and its receptor," *Nature*, vol. 427, no. 6971, pp. 215–221, 2004.

[64] A. Zelazny, A. Seluanov, A. Cooper, and E. Bibi, "The NG domain of the prokaryotic signal recognition particle receptor, FtsY, is fully functional when fused to an unrelated integral membrane polypeptide," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 94, no. 12, pp. 6025–6029, 1997.

[65] E. Bibi, A. A. Herskovits, E. S. Bochkareva, and A. Zelazny, "Putative integral membrane SRP receptors," *Trends in Biochemical Sciences*, vol. 26, no. 1, pp. 15–16, 2001.

[66] S. A. Ladefoged and G. Christiansen, "A GTP-binding protein of Mycoplasma hominis: a small sized homolog to the signal recognition particle receptor FtsY," *Gene*, vol. 201, no. 1-2, pp. 37–44, 1997.

[67] I. V. Shepotinovskaya and D. M. Freymann, "Conformational change of the N-domain on formation of the complex between the GTPase domains of *Thermus aquaticus* Ffh and FtsY," *Biochimica et Biophysica Acta*, vol. 1597, no. 1, pp. 107–114, 2002.

[68] T. Lichi, G. Ring, and J. Eichler, "Membrane binding of SRP pathway components in the halophilic archaea Haloferax volcanii," *European Journal of Biochemistry*, vol. 271, no. 7, pp. 1382–1390, 2004.

[69] A. Haddad, R. W. Rose, and M. Pohlschröder, "The Haloferax volcanii FtsY homolog is critical for haloarchaeal growth but does not require the A domain," *Journal of Bacteriology*, vol. 187, no. 12, pp. 4015–4022, 2005.

[70] J. Luirink, C. M. Ten Hagen-Jongman, C. C. van der Weijden et al., "An alternative protein targeting pathway in Escherichia coli: Studies on the role of FtsY," *EMBO Journal*, vol. 13, no. 10, pp. 2289–2296, 1994.

[71] H.-J. Dong, J.-Y. Jiang, and Y.-Q. Li, "The distinct anchoring mechanism of FtsY from different microbes," *Current Microbiology*, vol. 59, no. 3, pp. 336–340, 2009.

[72] M. Mircheva, D. Boy, B. Weiche, F. Hucke, P. Graumann, and H.-G. Koch, "Predominant membrane localization is an essential feature of the bacterial signal recognition particle receptor," *BMC Biology*, vol. 7, article no. 76, 2009.

[73] E. de Leeuw, D. Poland, O. Mol et al., "Membrane association of FtsY, the *E. coli* SRP receptor," *FEBS Letters*, vol. 416, no. 3, pp. 225–229, 1997.

[74] E. de Leeuw, K. te Kaat, C. Moser et al., "Anionic phospholipids are involved in membrane association of FtsY and stimulate its GTPase activity," *EMBO Journal*, vol. 19, no. 4, pp. 531–541, 2000.

[75] R. Parlitz, A. Eitan, G. Stjepanovic et al., "Escherichia coli signal recognition particle receptor FtsY contains an essential and autonomous membrane-binding amphipathic helix," *Journal of Biological Chemistry*, vol. 282, no. 44, pp. 32176–32184, 2007.

[76] N. J. Marty, D. Rajalingam, A. D. Kight et al., "The membrane-binding motif of the chloroplast signal recognition particle receptor (cpFtsY) regulates GTPase activity," *Journal of Biological Chemistry*, vol. 284, no. 22, pp. 14891–14903, 2009.

[77] S. Angelini, S. Deitermann, and H.-G. Koch, "FtsY, the bacterial signal-recognition particle receptor, interacts functionally and physically with the SecYEG translocon," *EMBO Reports*, vol. 6, no. 5, pp. 476–481, 2005.

[78] B. Weiche, J. Bürk, S. Angelini, E. Schiltz, J. O. Thumfart, and H.-G. Koch, "A cleavable N-terminal membrane anchor is involved in membrane binding of the *Escherichia coli* SRP receptor," *Journal of Molecular Biology*, vol. 377, no. 3, pp. 761–773, 2008.

[79] L. Bahari, R. Parlitz, A. Eitan et al., "Membrane targeting of ribosomes and their release require distinct and separable functions of FtsY," *Journal of Biological Chemistry*, vol. 282, no. 44, pp. 32168–32175, 2007.

[80] R. Moll, S. Schmidtke, and G. Schäfer, "Domain structure, GTP-hydrolyzing activity and 7S RNA binding of *Acidianus ambivalens* Ffh-homologous protein suggest an SRP-like complex in archaea," *European Journal of Biochemistry*, vol. 259, no. 1-2, pp. 441–448, 1999.

[81] S. H. Bhuiyan, K. Gowda, H. Hotokezaka, and C. Zwieb, "Assembly of archaeal signal recognition particle from recombinant components," *Nucleic Acids Research*, vol. 28, no. 6, pp. 1365–1373, 2000.

[82] J. L. Diener and C. Wilson, "Role of SRP19 in assembly of the *Archaeoglobus fulgidus* signal recognition particle," *Biochemistry*, vol. 39, no. 42, pp. 12862–12874, 2000.

[83] R. Gropp, F. Gropp, and M. C. Betlach, "Association of the halobacterial 7S RNA to the polysome correlates with expression of the membrane protein bacterioopsin," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 89, no. 4, pp. 1204–1208, 1992.

[84] H. Dale and M. P. Krebs, "Membrane insertion kinetics of a protein domain in vivo. The bacterioopsin N terminus inserts co-translationally," *Journal of Biological Chemistry*, vol. 274, no. 32, pp. 22693–22698, 1999.

[85] H. Dale, C. M. Angevine, and M. P. Krebs, "Ordered membrane insertion of an archaeal opsin in vivo," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 97, no. 14, pp. 7847–7852, 2000.

[86] G. Ring and J. Eichler, "Extreme secretion: protein translocation across the archaleal plasma membrane," *Journal of Bioenergetics and Biomembranes*, vol. 36, no. 1, pp. 35–45, 2004.

[87] R. Ortenberg and M. Mevarech, "Evidence for post-translational membrane insertion of the integral membrane protein bacterioopsin expressed in the heterologous halophilic archaeon Haloferax volcanii," *Journal of Biological Chemistry*, vol. 275, no. 30, pp. 22839–22846, 2000.

[88] K. Wild, M. Halic, I. Sinning, and R. Beckmann, "SRP meets the ribosome," *Nature Structural and Molecular Biology*, vol. 11, no. 11, pp. 1049–1053, 2004.

[89] A. K. K. Lakkaraju, C. Mary, A. Scherrer, A. E. Johnson, and K. Strub, "SRP keeps polypeptides translocation-competent by slowing translation to match limiting ER-targeting sites," *Cell*, vol. 133, no. 3, pp. 440–451, 2008.

[90] S. McGinnis and T. L. Madden, "BLAST: at the core of a powerful and diverse set of sequence analysis tools," *Nucleic Acids Research*, vol. 32, pp. W20–W25, 2004.

[91] E. S. Andersen, A. Lind-Thomsen, B. Knudsen et al., "Semiautomated improvement of RNA alignments," *RNA*, vol. 13, no. 11, pp. 1850–1859, 2007.

[92] R. C. Edgar, "MUSCLE: multiple sequence alignment with high accuracy and high throughput," *Nucleic Acids Research*, vol. 32, no. 5, pp. 1792–1797, 2004.

[93] M. Clamp, J. Cuff, S. M. Searle, and G. J. Barton, "The Jalview Java alignment editor," *Bioinformatics*, vol. 20, no. 3, pp. 426–427, 2004.

Hindawi Publishing Corporation
Archaea
Volume 2010, Article ID 410975, 11 pages
doi:10.1155/2010/410975

Research Article

Mutational and Bioinformatic Analysis of Haloarchaeal Lipobox-Containing Proteins

Stefanie Storf,[1] Friedhelm Pfeiffer,[2] Kieran Dilks,[1] Zhong Qiang Chen,[1,3] Saheed Imam,[4] and Mechthild Pohlschröder[1]

[1] *Department of Biology, University of Pennsylvania, Philadelphia, PA 19104-6018, USA*
[2] *Department of Membrane Biochemistry, Max-Planck-Institute of Biochemistry, Martinsried 82152, Germany*
[3] *Graduate Group in Genomics and Computational Biology, University of Pennsylvania, Philadelphia, PA 19104-6018, USA*
[4] *Program in Cellular and Molecular Biology, University of Wisconsin-Madison, Madison, WI 53706, USA*

Correspondence should be addressed to Mechthild Pohlschröder, pohlschr@sas.upenn.edu

Received 18 May 2010; Revised 4 July 2010; Accepted 12 July 2010

Academic Editor: Jerry Eichler

A conserved lipid-modified cysteine found in a protein motif commonly referred to as a lipobox mediates the membrane anchoring of a subset of proteins transported across the bacterial cytoplasmic membrane via the Sec pathway. Sequenced haloarchaeal genomes encode many putative lipoproteins and recent studies have confirmed the importance of the conserved lipobox cysteine for signal peptide processing of three lipobox-containing proteins in the model archaeon *Haloferax volcanii*. We have extended these *in vivo* analyses to additional *Hfx. volcanii* substrates, supporting our previous *in silico* predictions and confirming the diversity of predicted *Hfx. volcanii* lipoproteins. Moreover, using extensive comparative secretome analyses, we identified genes encodining putative lipoproteins across a wide range of archaeal species. While our *in silico* analyses, supported by *in vivo* data, indicate that most haloarchaeal lipoproteins are Tat substrates, these analyses also predict that many crenarchaeal species lack lipoproteins altogether and that other archaea, such as nonhalophilic euryarchaeal species, transport lipoproteins via the Sec pathway. To facilitate the identification of genes that encode potential haloarchaeal Tat-lipoproteins, we have developed TatLipo, a bioinformatic tool designed to detect lipoboxes in haloarchaeal Tat signal peptides. Our results provide a strong foundation for future studies aimed at identifying components of the archaeal lipoprotein biogenesis pathway.

1. Introduction

Most precursors of secreted prokaryotic proteins are transported across cytoplasmic membranes via either the universally conserved Sec pathway or the Twin-arginine translocation (Tat) pathway [1, 2]. The targeting of secreted protein precursors to these translocation pathways is dependent upon the recognition of pathway-specific signal peptides [1, 3]. In bacteria, most substrates transported via these pathways contain a signal peptide processing site that is recognized by signal peptidase I (SPase I) after transfer through the cytoplasmic membrane [3, 4]. However, one type of secreted protein, the bacterial lipoprotein precursors, is processed by signal peptidase II (SPase II), which specifically recognizes a conserved "lipobox" motif at the C-terminus of the signal peptide [4, 5]. The lipobox contains a cysteine residue to which a glyceride-fatty acid lipid is attached by a prolipoprotein diacylglyceryl transferase (Lgt) [6, 7]. SPase II cleaves the precursor immediately upstream of this lipid-modified cysteine. In Gram-negative and some Gram-positive bacteria, the conserved lipobox cysteine residue is also acylated by apolipoprotein *N*-acyltransferase [8, 9].

Sequence analyses of genomes isolated from archaea that thrive in high salt environments have identified a large number of open reading frames that encode putative Tat substrates containing potential lipoboxes [10–13]. Mass spectrometry results of the haloalkaliphilic halocyanin from *Natronomonas pharaonis* was consistent with the presence of an N-terminal cysteine residue modified by a lipid consisting of two C20 phytanyl groups linked to a glycerol group and also being acetylated at the amino group [13].

When the conserved lipobox cysteine residue is replaced with a serine residue, putative *Hfx. volcanii* lipoprotein precursors are not processed, suggesting that these haloarchaeal Tat substrates are in fact lipoproteins [12]. Interestingly, while unprocessed precursors of similar replacement mutants remain cell-associated in bacteria, the three mutant *Hfx. volcanii* precursors thus far tested are secreted into the supernatant [12]. Homologs of bacterial lipid-modification enzymes have not been detected in archaeal genomes. Considering this, release of the unprocessed mutant constructs into the extracellular environment supports the hypothesis that the molecular mechanisms underlying archaeal lipoprotein biosynthesis are distinct from their bacterial counterparts. It should also be noted that recently published data has revealed that some Gram-positive bacterial lipoproteins are also Tat substrates [14, 15]; however, little is known about the lipid-anchoring process in these bacteria. One possible interpretation of this result was that the release of the haloarchaeal Tat precursor proteins into the supernatant might reflect a difference in the mechanisms involved in lipid-modification of Sec and Tat substrates.

In this study, we carried out additional *in vivo* and *in silico* analyses to gain further insight into the processes involved in archaeal Tat and Sec lipoprotein biosynthesis and the diversity of archaeal lipoproteins. Although replacing the lipobox cysteine of additional putative Tat lipoproteins with a serine blocked precursor processing, these unprocessed mutant proteins remained membrane associated, indicating that secretion of unprocessed cysteine to serine replacement mutants is not a universal phenomenon, for either archaeal lipoprotein precursors or for Tat substrate lipoproteins. Conversely, complementary *in silico* analyses suggest that an extensive use of lipid-anchoring by membrane-associated Tat substrates is restricted to haloarchaea while it appears to be rare, or even non-existent, in other archaeal phyla. To identify genes that encode putative lipoprotein precursors, we employed existing lipoprotein prediction programs primarily trained on sets of bacterial Sec lipoproteins. This allowed us to compile a set of 484 putative lipoproteins from six haloarchaeal genomes. These data were used to develop a novel Tat-specific lipoprotein prediction tool for halophilic archaea.

2. Materials and Methods

2.1. Reagents. All enzymes used in standard molecular biology procedures were purchased from New England Biolabs, except for *iProof* High-Fidelity DNA Polymerase, which was purchased from Biorad. The ECL Plus Western blotting system, horseradish peroxidase-linked sheep anti-mouse antibody was purchased from Amersham Biosciences. The anti-myc monoclonal antibody and polyvinylidene difluoride membrane were purchased from Millipore. DNA and plasmid purification kits, and the anti-Penta-His antibody were purchased from QIAGEN. NuPAGE gels, buffers, and reagents were purchased from Invitrogen. Difco agar and Bacto yeast extract were purchased from Becton, Dickinson and Company. Peptone was purchased from Oxoid. All other chemicals and reagents were purchased from either Fisher or Sigma.

2.2. Strains and Growth Conditions. The plasmids and strains used in this study are listed in Supplementary Table 1 (available online at doi:10.1155/2010/410975). *Hfx. volcanii* strains were routinely grown at 45°C in 18% MGM [16]. MGM was supplemented with novobiocin (2.0 μg ml^{-1}) and thymidine (40 μg ml^{-1}), when necessary. *Escherichia coli* strains were grown at 37°C in NZCYM and supplemented with ampicillin (100 μg ml^{-1}).

2.3. Construction of Expression Vectors. The Hvo_*B0139*, Hvo_*1242*, Hvo_*1808*, Hvo_*1609*, and Hvo_*1580* genes were placed under the control of P_{fdx} [17] and cloned into pMLH3 [18] with a C-terminal Myc tag generating plasmids listed in Supplementary Table 1. P_{fdx} was amplified from pGB70 using the FdxFor and FdxRev primers, and the substrate genes were amplified using the relevant primers (see Supplementary Table 2). Substrate genes were then placed under the control of P_{fdx} by overlap PCR, using both the substrate genes and the *fdx* promoter fragment PCR products as template. The resulting P_{fdx}, P_{fdx}-*B0139*, P_{fdx}-*1242*, P_{fdx}-*1808*, P_{fdx}-*1609*, and P_{fdx}-*1580* fragments were digested with BamHI and HindIII, and ligated into pMLH3 that had been digested with BamHI and HindIII and treated with Calf Intestinal Phosphatase. The sequence of the construct inserts was confirmed by DNA sequencing.

In addition to the listed Myc-tagged wild-type substrate constructs, several mutant constructs were generated. In all cases, wild-type substrate constructs were used as DNA templates for PCRs. The twin-arginine residues of these gene products were altered to twin lysines via overlap PCR, using the above constructs as DNA template and primers that resulted in replacement of the arginine codons with lysine codons (AAG) (Supplementary Table 2). For the genes encoding putative lipoprotein substrates, likely lipobox cysteine codons were altered to serine codons via overlap PCR, using primers that resulted in replacement of the cysteine codon with a serine codon (TCG) (Supplementary Table 2). For signal sequence deletion mutants, the potential signal sequence located prior to the N-terminal cysteine was deleted via PCR using primers listed in Supplementary Table 2. The fidelity of these mutant constructs was confirmed by DNA sequencing. Similar methods were used to clone Hvo_*0494* and Hvo_*0494C20S*, except that these constructs were cloned into pRV1-p*tna* plasmid under the control of the tryptophan-inducible promoter p*tna* [19]. Genes were amplified using the primers listed in Supplementary Table 2 and inserted into pRV1-p*tna* digested with NdeI and EcoRI.

Wild-type and mutant constructs were extracted from DH5α (Invitrogen) and passed through DL739 [20] to obtain nonmethylated DNA, which was subsequently transformed into *Hfx. volcanii* strain H99 [21] or KD5 [22] using the standard PEG method [16].

2.4. Expression and Localization of Tat Substrates. Liquid cultures of relevant strains were grown until mid-log

(OD_{600} ~ 0.5). Subsequently, cells were collected by centrifugation at 4300 g for 10 min at 4°C. The supernatants were recentrifuged at 4300 g for 10 min to remove cellular contamination, and secreted proteins were precipitated with cold TCA (100% v/v) then washed twice with cold acetone (80% v/v). The cell pellets were washed once with MGM then pelleted again as described and resuspended in 1× NuPAGE lithium dodecyl sulfate (LDS) Sample buffer.

2.5. Immunoblotting. All protein samples were stored at −20°C in 1× NuPAGE LDS sample buffer supplemented with 50 mM dithiothreitol (DTT). Samples were run on Bis-Tris NuPAGE gels under denaturing conditions using either morpholinepropanesulfonic acid (MOPS) or 2-(N-morpholino)ethanesulfonic acid (MES) running buffer, or on Tris-Acetate NuPAGE gels using TA running buffer. Proteins were transferred to polyvinylidene difluoride (PVDF) membrane using the Bio-Rad Transblot-SD Semi-Dry Transfer Cell at 15 V for 30 minutes. Three buffers were used in semidry transfer: anode I [300 mM Tris, 10% (v/v) methanol, pH 10.4], anode II [25 mM Tris, 10% (v/v) methanol, pH 10.4], and cathode [25 mM Tris, 40 mM glycine, and 10% (v/v) methanol, pH 9.4]. PVDF membranes were probed with the primary antibodies anti-Penta-His (1:1000) or anti-Myc (1:1000) and secondary anti-mouse antibody (1:10,000). All Western blots shown are representative of at least two independent experiments.

2.6. Archaeal Secretome Analyses. Various commonly available bioinformatic tools were used to analyze the secreted proteins of six haloarchaeal species, nine non-halophilic euryarchaeal species, nine crenarchaeal species, as well as three other archaeal species (see Supplementary Table 3). For lipoprotein predictions, we used three independent prediction programs: (i) the Prosite position-specific matrix PS51257 (PROKAR_LIPOPROTEIN) [23]; (ii) LipoP [24]; and (iii) predLipo [25]. To keep false positive predictions minimal, candidate proteins are considered lipobox-positive only when they are recognized by at least two of these predictors (also see Supplementary Text and Supplementary Figure 1). To identify Tat substrates, we used TatFind [26]. Predicted Tat substrates that were not predicted to be lipoproteins were designated Tat substrates with an SPase I cleavage site in their signal peptides. All predicted lipoproteins not identified by TatFind were designated Sec substrates with signal peptides processed by SPase II.

Phobius [27] was used to identify Sec substrate signal peptides cleaved by SPase I. However, we used caution when considering this data, as there is a strong tendency for Phobius to predict false positives for both Sec signal peptides and SPase I cleavage sites. Since both the Sec and Tat pathway-specific signal peptides contain a charged region followed by a hydrophobic core, and many contain an SPase I cleavage site, Phobius is inclined to misclassify many Tat signal peptides as Sec signal peptides [28]. Consequently, we disregarded Phobius predictions of Sec signal peptides in TatFind positives. Similarly, since SPase I has a relaxed specificity, which therefore necessitates relaxed cleavage site prediction constraints, many lipoproteins cleaved by SPase II are predicted to have signal peptides processed by SPase I by Phobius. Consequently, we disregarded Phobius predictions of SPase I cleavage sites for predicted lipoproteins, as predicted by the method described above (also see Supplementary Figure 1).

2.7. Tat Lipoprotein Prediction. To extract position-specific amino acid statistics, amino acid sequences of 484 lipoproteins predicted from six haloarchaeal genomes were "aligned" as follows: for 400 lipoproteins predicted to be Tat substrates, the sequences of the lipobox and the Tat motif, as predicted by the lipoprotein prediction programs and TatFind, respectively, were aligned. These alignments were attained by introducing a gap of variable length between the fifth and sixth amino acid residue after the second arginine of the twin arginine motif.

The lipoprotein set contains 50 TatFind negative sequences that contain two consecutive arginines in the charged region of the signal peptide. However, specific amino acid residues at positions +1, +4, +5, and/or +6, relative to these arginines are not allowed by TatFind. Some of these sequences may be false negatives caused by the stringent rules applied by TatFind. Therefore, the twin arginines in these potential false negatives were aligned with the twin arginines in the Tat motifs of the TatFind positives.

Most of the 34 remaining sequences that were examined were found to contain a single arginine. We aligned this arginine with the second arginine of the twin arginines of the Tat motifs identified by TatFind unless the preceding residue was the initiator methionine, in which case we aligned this arginine with the first arginine of the Tat motifs (see Supplementary Table 4).

The 484 aligned haloarchaeal lipoproteins (Supplementary Table 4) were used to compute position-specific amino acid frequencies. These frequencies were used to develop an algorithm using a rule-based approach to detect haloarchaeal lipobox motifs. For the positions with the strongest composition bias, amino acids were categorized as "required", "frequent", "normal", "tolerated", or "excluded". Required and excluded amino acids were used in regular expressions, that is, no exceptions were allowed. Tolerated amino acids are not excluded, although they have not been found in the set of 484 aligned haloarchaeal lipoproteins at the corresponding position. However, they are similar to other amino acids, which have been found (e.g., tyrosine or tryptophane in positions where phenylalanine is found) (see also Supplementary Table 5). Moreover, tolerated amino acid residues are allowed only when other positions are occupied by frequent amino acids. The rule-based algorithm is the basis of TatLipo (available at SignalFind.org), which is an extension of TatFind, a program which identifies haloarchaeal Tat substrates with a high degree of accuracy [26, 29].

3. Results

3.1. Some Unprocessed Hfx. volcanii Lipobox Replacement Mutant Proteins Remain Cell-Associated. To further investigate the diversity of haloarchaeal lipoproteins secreted via

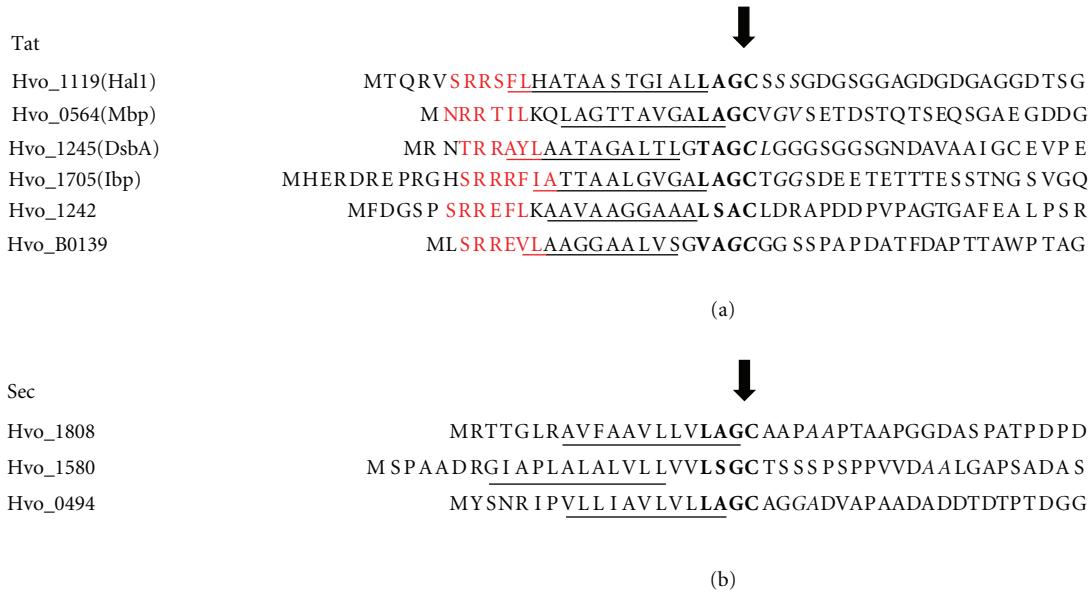


FIGURE 1: Tat and Sec signal peptides containing lipobox motifs. N-terminal regions of the precursors of (a) Tat; and (b) Sec substrates with lipoboxes (bold) predicted by at least two of the three lipoprotein prediction programs (PredLipo, LipoP and Prosite PS51257), with the exception of Hvo_1242, which was only LipoP-positive. Tat motifs (red) were predicted by TatFind, hydrophobic stretches (underlined), and SPase I cleavage sites (*italics*) were predicted by Phobius. An arrow indicates the predicted SPase II cleavage sites.

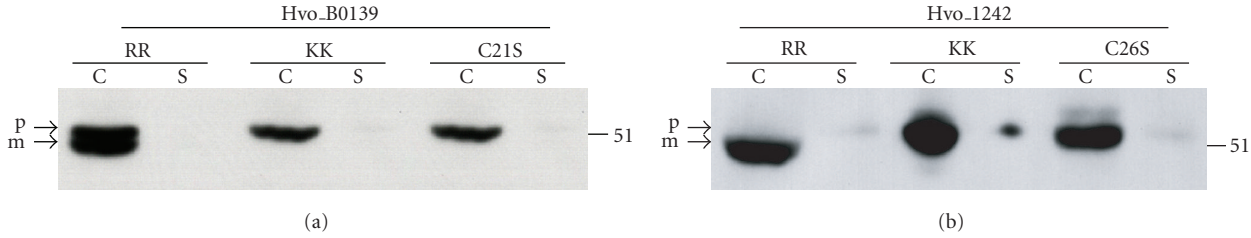


FIGURE 2: *Hfx. volcanii* proteins Hvo_B0139 and Hvo_1242 are Tat substrates that require the lipobox cysteine for processing but not for anchoring to the cytoplasmic membrane. Western blot analyses of the wild-type (RR), twin lysine replacement mutants (KK), and cysteine to serine replacement mutants (C21S and C26S for Hvo_B0139, and Hvo_1242, resp.). All proteins expressed had C-terminal Myc-tags and were detected using anti-Myc antibodies. Comparable amounts of protein were loaded in each lane. The migration of molecular weight standards is indicated on the right. Predicted positions of precursor (p) and mature (m) proteins are indicated.

the Tat pathway, we chose to characterize two additional proteins with distinct lipobox motifs, Hvo_B0139 (VAGC), predicted by three lipoprotein prediction programs (Prosite, predLipo and LipoP) and Hvo_1242 (LSGC), which was only predicted by LipoP (Figure 1(a) and see below). Consistent with these putative Tat substrates being lipoproteins, we determined that both are cell-associated (Figures 2(a) and 2(b)). Protein extracts isolated from *Hfx. volcanii* that express C-terminally Myc-tagged versions of these proteins were subjected to denaturing polyacrylamide gel electrophoresis (PAGE), followed by Western blot analyses. When genes are overexpressed from a plasmid, incomplete processing of the encoded protein is frequently observed. Consistent with the presence of a precursor and a processed protein, anti-Myc antibody detected two distinct protein bands in Western blots of extracts containing the Myc-tagged version of Hvo_B0139. Only a single protein band was detected in extracts containing Myc-tagged Hvo_1242. To provide evidence that these proteins are Tat substrates, we mutated the essential twin arginines that have been shown to be essential for protein export by the Tat translocase. Mutant proteins in which twin arginine residues in the signal peptide were replaced with twin lysines (Hvo_B0139KK and Hvo_1242KK, resp.) were overexpressed in wild-type *Hfx. volcanii*, and protein extracts containing these mutant proteins were subjected to denaturing PAGE, followed by Western blot analyses. Hvo_B0139KK migrated as a single protein band, at a position on the gel similar to that of the Hvo_B0139 precursor protein (Figure 2(a)). Moreover, Western blots of protein fractions isolated from *Hfx. volcanii* that overexpress Hvo_1242KK detected one protein band, which migrates at a slower rate than Hvo_1242, indicating that this mutant protein is not processed (Figure 2(b)).

To determine the importance of putative archaeal protein lipoboxes in signal peptide processing and membrane-anchoring, we replaced the conserved lipobox cysteines in

these archaeal Tat substrates with serines. Western blot analysis of purified protein extracts subjected to denaturing PAGE showed that Hvo_B0139C21S and Hvo_1242C26S replacement mutants migrate at a position on the gel similar to that of the corresponding precursor, indicating that the cysteine is required for signal peptidase processing (Figures 2(a) and 2(b)). Interestingly, although Giménez et al. showed that DsbA, Mbp, and Ibp cysteine to serine replacement mutants are released into the supernatant [12], the precursor forms of Hvo_B0139C21S and Hvo_1242C26S remain cell associated, which might be due to the unprocessed signal serving as a membrane anchor (Figure 2).

3.2. The Transport of Lipobox-Containing Proteins Is Not Dependent on a Particular Hfx. volcanii TatA Paralog. Although TatAt is essential for *Hfx. volcanii* survival under standard laboratory conditions, TatAo, the *Hfx. volcanii* paralog of TatAt, is not [22]. To determine whether the membrane association of unprocessed cysteine to serine replacement mutants transported via the Tat pathway depends upon a TatA paralog distinct from that involved in transporting similar mutants that are released into the extracellular environment, we investigated the transport of the Hvo_B0139 and Hvo_1242, as well as DsbA, in *Hfx. volcanii* △*tatAo* mutants. We determined that these lipobox-containing proteins are transported with similar efficiencies in wild-type cells and △*tatAo* mutants (Figures 3(a)–3(c)). Since the soluble Tat substrate arabinanase is also secreted in a TatAo-independent manner (Figure 3(d)), it is clear that in *Hfx. volcanii* the TatAt role in transport is not limited to the secretion of lipoproteins.

3.3. Putative Hfx. volcanii Sec Substrate Lipobox C-S Replacement Mutants Are Processed, but Are Relatively Unstable. To determine whether archaeal Sec substrates containing lipobox motifs require the conserved cysteine for processing, we characterized three putative *Hfx. vocanii* Sec substrates with potential lipoboxes: Hvo_1808 (LAGC), Hvo_1580 (LSGC), and Hvo_0494 (LAGC). Consistent with Hvo_1808 and Hvo_1580 being lipoproteins, they are primarily retained in cell-associated protein fractions (Figures 4(a) and 4(b)). Although the supernatant fraction of Hvo_1808 expressing cultures contained a minor protein band that migrated with one of the bands seen in the membrane-associated fraction, it is likely that this results from protein shedding during the isolation of protein fractions, as is the case for many bacterial lipoproteins (also observed for the *Hfx. volcanii* DsbA, [12]). Shedding may also be the reason why a large portion of Hvo_0494 is found in the supernatant fraction (Figure 4(c)). To confirm that these Sec substrates are processed, we attempted to overexpress signal peptide deletion mutants of these three proteins (Hvo_1808Δss, Hvo_1580Δss, and Hvo_0494Δss), which should have the same molecular weights as the processed proteins. Although Hvo_1508Δss and Hvo_0494Δss appear to be unstable, as no corresponding protein was identified by Western blot analysis (data not shown), Hvo_1808Δss migrated at the same position on the gel as the wild-type protein when separated by denaturing PAGE, indicating that the wild-type protein is processed (Figure 4(a)).

Interestingly, when the conserved lipobox cysteine in these Sec substrates was replaced with a serine, Western blot analysis showed that, when separated by denaturing PAGE, Hvo_1808C19S and Hvo_1580C24S migrate at the same position on the gel as the wild-type proteins and are found in the same relative amounts in the cell-associated and supernatant protein fractions, indicating that neither the processing nor the membrane anchoring of these mutant proteins is dependent on the lipobox cysteine (Figures 4(a) and 4(b)). In fact, all three Sec substrates also contain potential SPase I processing sites, as predicted by Phobius [27] (Figure 1(b)). However, Western blot analyses also revealed that all three mutant proteins are less abundant than the corresponding wild-type proteins, in both the cell associated and the supernatant protein fractions, suggesting that these replacement mutants are less stable than the wild-type proteins. In fact, in addition to the mutant proteins being less abundant than the wild-type proteins, Western blot analysis of protein fractions containing Hvo_0494C20S revealed a faster migrating band in the supernatant fraction, possibly a degradation product (Figure 4(c)).

3.4. Prediction of Archaeal Lipoproteins. Upon determining that many predicted *Hfx. volcanii* Tat signal peptides contain putative lipoboxes, and confirming that the conserved lipobox cysteine in these substrates is important for precursor processing, we resolved to determine (i) whether Tat substrates of other haloarchaea frequently contain SPase II processed signal peptides—as has been suggested for at least two additional haloarchaea species [10, 11]; (ii) whether non-halophilic archaeal genomes also encode predicted lipoproteins; and (iii) whether these putative non-halophilic lipoproteins are typically predicted Sec or Tat substrates.

As there was no archaeal-specific lipoprotein prediction program available, we used other existing tools: (i) the Prosite position-specific matrix PS51257 (PROKAR_LIPOPROTEIN) [23] and LipoP [24], which have both been mainly trained on lipoproteins encoded by Gram-negative bacterial genomes; and (ii) predLipo [25], which has been trained on lipoproteins encoded by Gram-positive bacterial genomes. The results of these three predictors did not correlate well for archaeal genomes (see the Supplementary Text). Only 43% of the initial predictions were common to all three programs, while 24% were specific to only one of them. To increase the reliability of the prediction, we requested that at least two of the three prediction programs must be positive for a protein to consider it a lipoprotein (i.e., as having an SPase II cleavage site). Excluding results obtained with only a single predictor resulted in a set of 484 predicted lipoproteins from six halophilic archaea (Supplementary Table 4). Of these, 56% were positive by all three predictors and 44% were positive by only two of the three predictors. In the following, we refer to these as lipoproteins although experimental confirmation in archaea is still lacking except for halocyanin from *N. pharaonis* [13]. However, we consider the requirement for

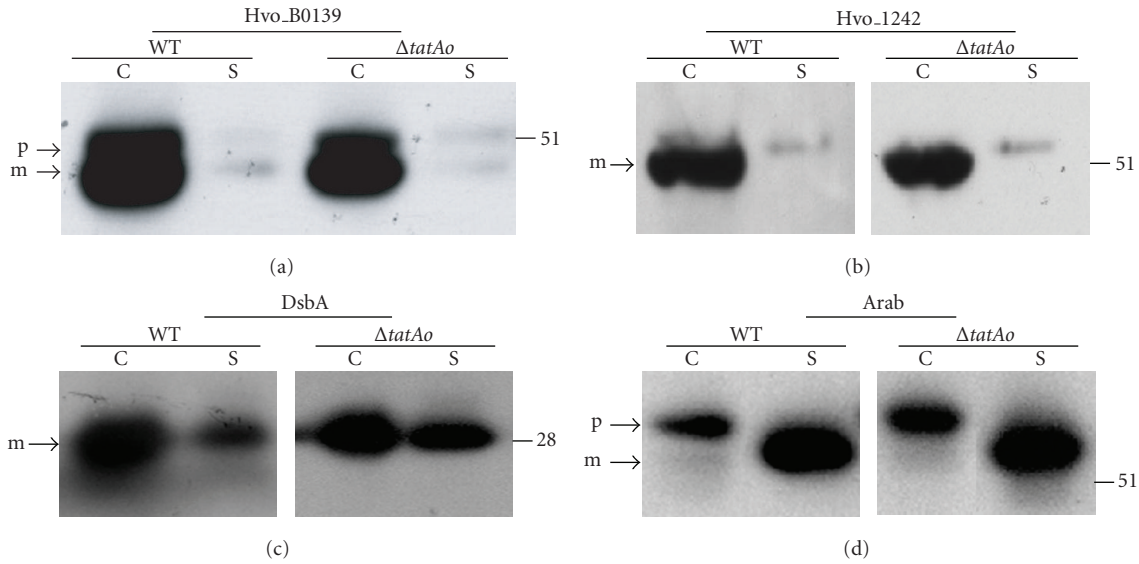


Figure 3: Putative Tat substrate lipoproteins Hvo_B0139, Hvo_1242, DsbA and arabinanase are translocated independent of TatAo in *Hfx. volcanii*. Western blot analyses of wild-type (RR) Hvo_B0139, Hvo_1242, and DsbA (Hvo_1245) and arabinanase (Hvo_B0232) expressed in *Hfx. volcanii* wild-type (WT) or TatAo deletion mutants (△*tatAo*). All proteins expressed had C-terminal Myc-tags, and were detected using anti-Myc antibodies. Comparable amounts of protein were loaded in each lane. The migration of molecular weight standards is indicated on the right. Predicted positions of precursor (p) and mature (m) proteins are indicated. For comparisons of the unprocessed and processed substrate migration see Figure 2.

positive results from two independent predictors a strong constraint so that we expect only a small fraction will be false positives.

The lipoproteins predicted by this method were used in two ways. Predictions for a total of 27 archaeal genomes were used for statistical analyses. The set of 484 haloarchaeal lipoproteins was used to determine position-specific amino acid frequencies, results of which were used to develop TatLipo, a haloarchaea-specific prediction program for lipoproteins that are secreted by the Tat pathway.

3.5. Extensive Anchoring of Tat Substrates via Lipid Anchor Appears to Be Unique to Haloarchaea. Consistent with previous analyses, we determined that 39–55% of the haloarchaeal proteins are secreted via the Tat pathway (Figure 5(a)). However, although most archaea appear to secrete some of their proteins via the Tat pathway, Tat substrates account for no more than 8% of any non-halophilic archaeal secretome thus far analyzed (Figure 5(a), Supplementary Table 3) [29]. The lack of predicted Tat substrates in the euryarchaea *M. hungatei, M. jannaschii,* and *M. kandleri*, or in *N. equitans* and *N. maritimus*, is consistent with the absence of homologs of Tat pathway components in these organisms ([29] and data not shown).

Conversely, while proteins that contain signal peptides with lipobox motifs are common in some euryarchaea, we found that they are absent in many crenarchaea (Figure 5(b), Supplementary Table 3). In the euryarchaea, the fraction of predicted lipoproteins is somewhat higher among haloarchaea (26–36%) than it is among non-halophilic species (7–20%). Six of the nine analyzed crenarchaeota completely lack predicted lipoproteins. *Ignicoccus hospitalis,* the only archaeon known to have an outer membrane [30]), is an unusual crenarchaeon, having a secretome in which 4.2% of the proteins have a lipobox-containing signal peptide (Figure 5(b), Supplementary Table 3).

Although roughly half of the haloarchaeal Tat substrates contain a lipobox motif (46–65%) (Figure 5(c), Supplementary Table 3), we determined that only about 10% of haloarchaeal Sec substrates contain one (Figure 5(d), Supplementary Table 3). However, while haloarchaea are unique among archaea with respect to the portion of secreted proteins containing the Tat/lipobox combination, they are relatively similar to other euryarchaea in regard to the frequency of secreted proteins having the Sec/lipobox combination (Figure 5(d)).

Furthermore, a large majority (78–87%) of putative haloarchaeal lipoproteins are secreted via the Tat pathway (Figure 5(e), Supplementary Table 3). In contrast, this combination is extremely rare in other archaeal species. In our *in silico* analysis, only a single exception was identified, protein AF2235 of *Archaeoglobus fulgidus* has a Tat/lipobox motif (Figure 5(e), Supplementary Table 3).

The difference in usage of the Tat pathway and lipid-anchoring cannot be attributed to a general disparity in the level of protein secretion in these organisms. While the fraction of secreted proteins varies considerably among archaeal species, ranging from 3.6% in *S. acidocaldarius* to 10.7% in *M. hungatei*, no major differences in the range of the portion of proteins secreted are evident between the various phyla of archaea (Figure 5(f), Supplementary Table 3).

3.6. Development of TatLipo for Prediction of Haloarchaeal Tat Lipoproteins. Although a large number of potential

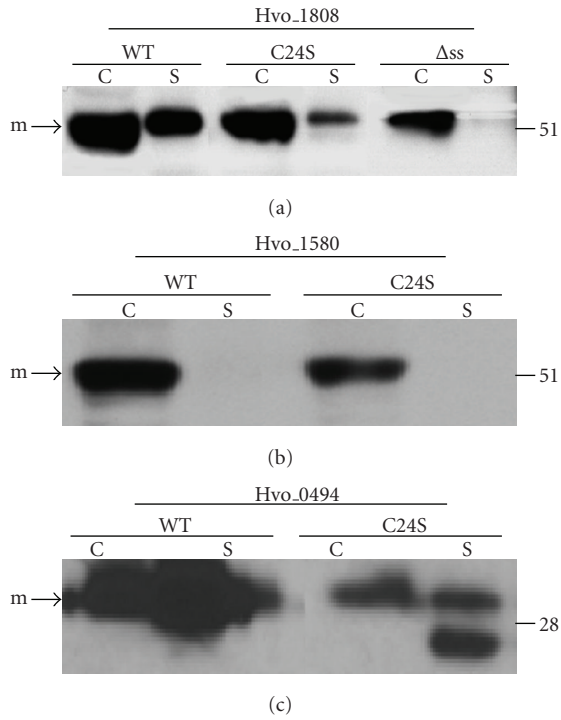


Figure 4: Cysteine to serine mutants of Hvo_1808, Hvo_1580, and Hvo_0494 are less stable than wild-type constructs but appear to be processed. Western blot analyses of wild-type proteins (WT), cysteine to serine replacement mutants (Hvo_1808C19S, Hvo_1580C24S and Hvo_0494C20S), and signal sequence deletion mutants (△ss) of Hvo_1808. All proteins expressed had C-terminal Myc-tags except for Hvo_0494 and Hvo_0494C20S, which were C-terminally His-tagged. Myc and His-tagged proteins were detected using anti-Myc and anti-His antibodies, respectively. Comparable amounts of protein were loaded in each lane. The migration of molecular weight standards is indicated on the right. Predicted positions of precursor (p) and mature (m) proteins are indicated.

Tat substrate lipoproteins were detected using the available prediction programs, our *in vivo* results show that our requirement, that at least two of the three programs used make a positive prediction, may be overly stringent (Figures 1(a) and 2(b)). This may be because these programs were trained to search for lipobox motifs within the context of Sec signal peptides, and although Tat and Sec signal peptides have significant similarities, Tat signal peptides, in addition to containing a unique twin arginine motif, are generally longer and contain a less hydrophobic h-domain than Sec signal peptides. Therefore, to identify additional potential haloarchaeal lipoproteins, we modified TatFind, which detects haloarchaeal Tat substrates, to include haloarchaeal lipobox predictions. Since the vast majority of haloarchaeal lipoproteins are secreted via the Tat pathway, this program should identify Tat lipoproteins missed by the other prediction programs used in this study.

Although there are no confirmed Tat lipoproteins to use as a training set, the stringent lipoprotein prediction as described above, provided 484 lipoprotein candidates to be used in defining a consensus haloarchaeal lipobox motif (Figure 6). Manual alignments of the N-termini of these proteins, as described in Materials and Methods, were used to compute position-specific amino acid composition and determined that the most frequent haloarchaeal lipobox motif is LAGC and that sequence conservation increases along the motif (leucine: 60.5%, alanine: 83.5%, glycine: 98.8%, cysteine: 100%) (Figure 6 and Supplementary Table 5). Furthermore, we identified a bias in the amino acid residue composition both upstream and downstream of the LAGC sequence. A number of charged or hydrophilic amino acids (D, E, R, H, K, N, Q) are strictly forbidden in front of the lipobox motif, and at the position immediately following the conserved cysteine, leucine and serine are frequent while aspartic acid and cysteine are excluded.

Moreover, we determined the distance between the twin arginine and the lipobox motif. The calculated number of amino acid residues between the second arginine and the conserved cysteine in the lipobox motif was in the range of 12–21.

To specifically predict Tat substrates with class II signal peptides, we incorporated the lipobox algorithm into TatFind, generating the lipoprotein prediction program TatLipo. Thus, TatLipo defines the prevalent haloarchaeal lipobox motif in Tat substrates (Figure 6). TatLipo was applied to the set of 400 predicted haloarchaeal lipoproteins that are secreted via the Tat pathway. Only three proteins in this set were not predicted by TatLipo: two of these slightly exceed the distance constraint. TatLipo predicts 113 additional lipoprotein candidates. More than two-thirds of the additional TatLipo assignments (78 of 113, 69.0%) were also predicted by one of the three other lipobox prediction programs, including Hvo_1242, which was LipoP-positive. This may indicate that many of the additional predictions are correctly called as lipoproteins. These partial confirmations are relatively evenly distributed among the three prediction programs (41 predLipo, 19 Prosite, 18 LipoP). Within the subset of lipoprotein candidates secreted via the Tat pathway, TatLipo confirmed most of the lipobox-containing proteins that were predicted by only one of the three bacterial lipoprotein prediction programs.

In conclusion, TatLipo is able to identify nearly all of the Tat-secreted lipobox proteins in halophilic archaea. While the stringent rules applied for lipobox assignment may have resulted in a number of false negatives, experimental confirmation in the future, will allow us to further improve the lipobox prediction algorithms.

4. Discussion

Haloarchaea transport a large fraction of their secreted proteins via the Tat pathway, possibly as an adaptation to the high salt environments they inhabit. Computational analyses of putative Tat substrates identified in several halophilic archaea have revealed that many of these precursor proteins also contain potential lipoboxes [10, 11, 22]. These findings are supported by the results obtained in the studies presented here, which also show that this predominance of SPase II cleavage sites in Tat substrates, like the extensive use

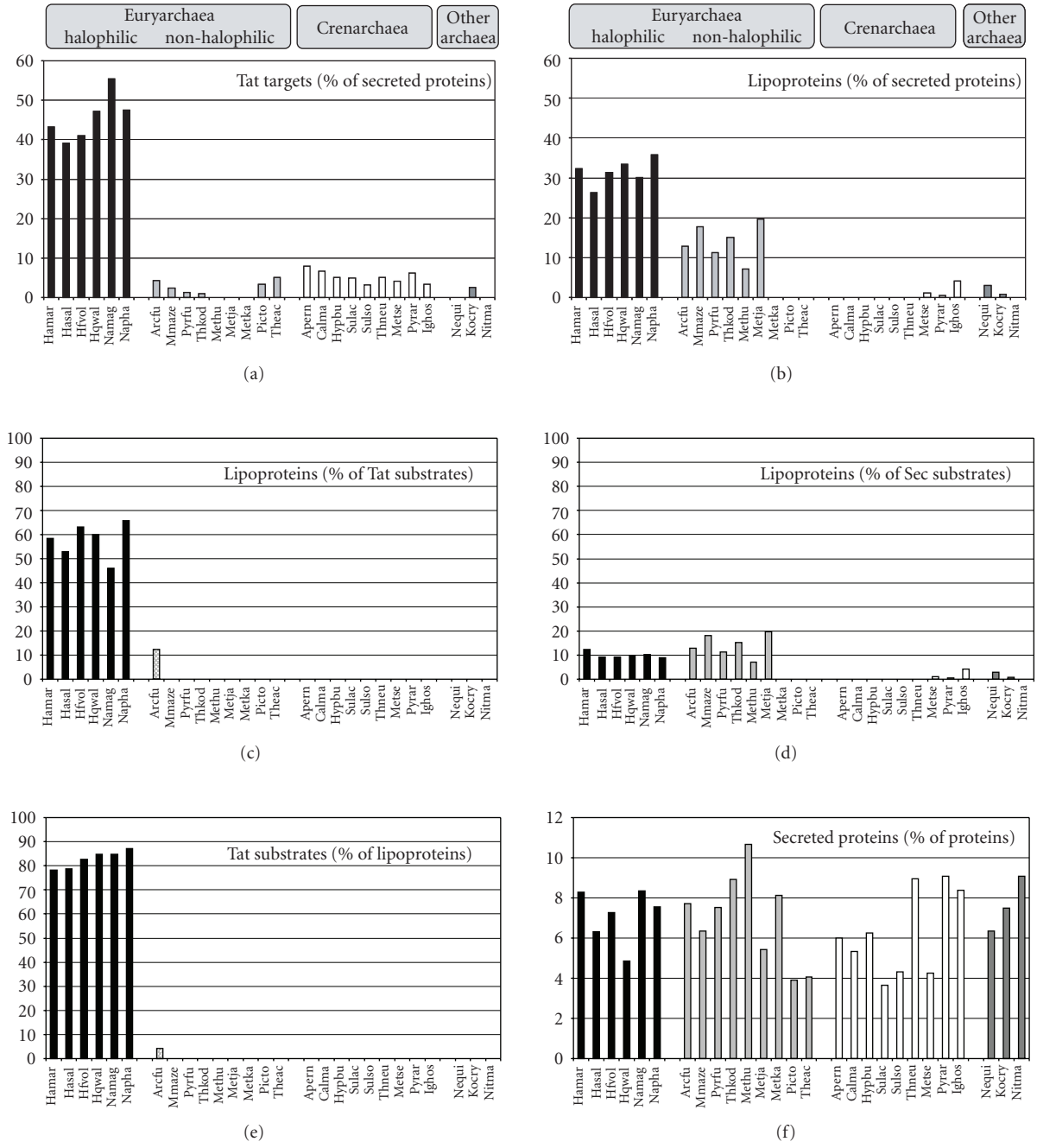


Figure 5: Only haloarchaeal Tat substrates are predicted to be predominantly lipoproteins. The predicted relative proportions are shown for several types of secreted proteins for halophilic and non-halophilic euryarchaeal species as well as crenarchaeal species and three species belonging to other archaeal phyla. (a) The percentage of secreted proteins that are Tat substrates. (b) The percentage of secreted proteins that are lipoproteins. (c) The percentage of Tat substrates that are lipoproteins. (d) The percentage of Sec substrates that are lipoproteins. (e) The percentage of lipoproteins that are secreted via the Tat pathway. (f) The percentage of proteins that are secreted (predicted Tat and Sec substrates with SPase I or SPase II cleavae sites). Raw data are available in Supplementary Table 4. Organisms are abbreviated as follows: haloarchaea (Hamar: *Haloarcula marismortui*; Hasal: *Halobacterium salinarum*; Hfvol: *Haloferax volcanii*; Hqwal: *Haloquadratum walsbyi*; Namag: *Natrialba magadii*; Napha: *Natronomonas pharaonis*), other euryarchaea (Arcfu: *Archaeoglobus fulgidus*; Mmaze: *Methanosarcina mazei*; Pyrfu: *Pyrococcus furiosus*; Thkod: *Thermococcus kodakarensis*; Methu: *Methanospirillum hungatei*; Metja: *Methanocaldococcus jannaschii*; Metka: *Methanopyrus kandleri*; Picto: *Picrophilus torridus*; Theac: *Thermoplasma acidophilum*), crenarchaea (Apern: *Aeropyrum pernix*; Calma: *Caldivirga maquilingensis*; Hypbu: *Hyperthermus butylicus*; Sulac: *Sulfolobus acidocaldarius*; Sulso: *Sulfolobus solfataricus*; Thneu: *Thermoproteus neutrophilus*; Metse: *Metallosphaera sedula*; Pyrar: *Pyrobaculum arsenaticum*; Ighos: *Ignicoccus hospitalis*), and other archaea (Nequi: *Nanoarchaeum equitans*; Kocry: *Korarchaeum cryptofilum*; Nitma: *Nitrosopumilus maritimus*).

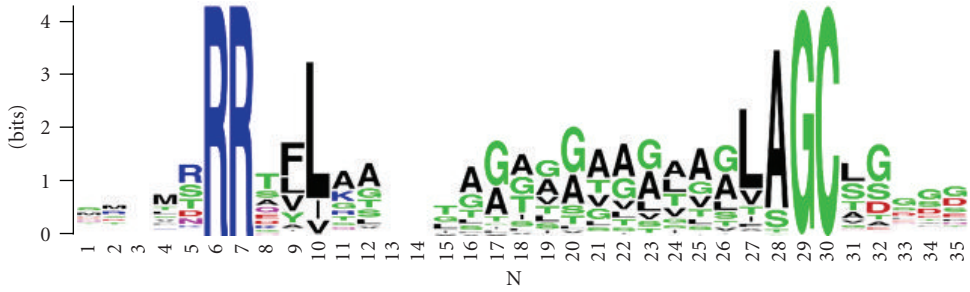


Figure 6: Conserved lipobox motifs of haloarchaea. Consensus motif was generated using the lipobox motifs of 400 predicted haloarchaeal Tat lipoproteins (see Supplementary Table 4). The alignment depicts the cleavage site G/A at position −1. Logos were generated using weblogo (http://weblogo.berkeley.edu/).

of the Tat pathway to translocate secreted proteins, is a feature that is likely unique to haloarchaea. Moreover, in addition to revealing that a large portion of haloarchaeal Tat substrates appear to be lipoprotein precursors, our *in silico* analyses have also demonstrated that relatively few archaeal Sec substrates contain a genuine lipobox.

In a previous study, we determined that some haloarchaeal Tat substrates contain functional lipoboxes, with core amino acid sequences LAGC or TAGC, required for proper processing of the precursor protein and for membrane-association of the processed protein. While our *in silico* and *in vivo* data indicate that the lipobox cysteine is strictly conserved, we have determined that in addition to lipobox glycine, which is found at position −1 in 98% of the putative lipoboxes that we have identified, alanine is occasionally found at position −1. We have now shown that a haloarchaeal Tat substrate containing an alanine at lipobox position −1 requires the putative lipobox cysteine for proper processing. Moreover, a valine at the first position or a serine at the second position of the lipobox, as observed in the lipoboxes VAGC and LSAC, respectively, can also be present in haloarchaeal lipoboxes. Similar to previously tested replacement mutants, constructs lacking the lipobox cysteine were not processed. However, while previously tested mutant substrates were released into the supernatant [12], in this study, precursors containing lipoboxes in which the cysteine was replaced with a serine remained cell associated. The cell association of these mutant proteins is not that surprising since this is similar to what has been observed for the majority of bacterial lipoproteins containing a cysteine to serine lipobox substitution where unprocessed bacterial Sec substrate lipoprotein replacement mutants remain tethered to the cytoplasmic membrane via the hydrophobic stretch of the signal peptide. Analogously, it is possible that the unprocessed signal peptides of the mutant Tat substrate lipoprotein precursors investigated in this study serve a similar purpose: anchoring these unprocessed mutant proteins to the archaeal cytoplasmic membrane. It is not clear why some unprocessed lipoproteins are released into the supernatant.

We have also determined that membrane-association of an unprocessed Tat substrate containing a lipobox is not dependent on a specific Tat pore component; in fact, we have only identified a single archaeal Tat substrate that specifically requires TatAo for successful translocation (data not shown). Moreover, we have also shown that the soluble secreted protein arabinanase is secreted via Tat pores containing TatAt, demonstrating that no Tat pore is specifically dedicated to the secretion of haloarchaeal lipoproteins (Figure 3(d)).

Although our analyses clearly indicate that predicted haloarchaeal lipoprotein precursors are generally transported to the cytoplasmic membrane via the Tat pathway, the specific mechanisms involved in lipid modification and signal peptide cleavage of these substrates are currently unknown. Lnt is the evolutionarily conserved N-acyltransferase that catalyzes the acylation of the lipobox cysteine in Gram-negative bacteria. The acylation of the conserved lipobox cysteine of lipoproteins has also been confirmed in the Gram-positive bacteria *Bacillus subtilis* and *Staphylococcus aureus,* but Lnt homologs have not been identified in either of these species, indicating that an unrelated enzyme acylates lipoproteins in these organisms [31, 32]. Given that archaea also lack Lnt homologs, archaeal species that produce lipoproteins might also express a novel, archaeal-specific N-acyltransferase, or perhaps the lipoprotein acylation in archaea and Gram-positive bacteria is performed by an enzyme that is conserved between them. On the other hand, although both Gram-negative and Gram-positive bacterial species express a conserved prolipoprotein diacylglyceryl transferase, an archaeal homolog of this enzyme has not been identified, indicating that archaea are likely to express a unique enzyme that performs an analogous function.

In addition to putative haloarchaeal Tat substrate lipoproteins, previously reported preliminary evidence has indicated that some Gram-positive bacterial lipoproteins may also be Tat substrates [14, 15]. Elucidating the mechanisms involved in modifying and processing Tat substrate lipoproteins in bacteria and archaea will almost certainly reveal important similarities and key differences in the processing of Sec substrate and Tat substrate lipoproteins.

In this study, we determined that cysteine to serine replacement mutants corresponding to two Sec substrates containing a putative lipobox are processed, and also showed that these processed mutants remain membrane-associated,

perhaps by forming protein complexes with other membrane-bound proteins. The fact that the putative lipobox in these Sec substrates is conserved in homologous proteins encoded by other haloarchaeal species is interesting (data not shown). Although it is not clear whether these conserved sequences are lipoboxes, their conserved nature suggests that they may serve an important function. On the other hand, in bacteria, some lipobox cysteine to serine replacement mutants are processed by a bacterial SPase I [33]. Interestingly, all three Sec substrates investigated here contain predicted SPase I cleavage sites, as determined by Phobius (Figure 1(b)). Consistent with the hypothesis that SPase I may have processed these mutant Sec substrates, the cysteine to serine replacement mutant of the Tat substrate DsbA also contains a predicted SPase I processing site that appears to be processed, albeit inefficiently [12]. Moreover, Western blot analyses indicate that the cysteine to serine replacement mutants are less stable than the corresponding wild-type proteins, and in the case of Hvo_0494, a smaller product was detected in the protein fraction isolated from the supernatant.

Our *in silico* data suggests that haloarchaea are unique in anchoring Tat substrates to the membrane via a lipid anchor. However, in addition, while the vast majority of the lipobox-containing proteins are secreted via the Tat pathway, haloarchaea as well as non-halophilic euryarchaeal genomes contain open reading frames that code for putative lipoproteins, which are secreted via the Sec pathway. Therefore, with regard to determining the relative importance of lipoproteins in the various archaeal phyla, further investigations of putative archaeal Sec substrate lipoproteins are necessary. These include investigation of *Hfx. volcanii* Sec substrates that contain lipoboxes, such as N-terminal amino acid sequencing of wild-type substrates and their corresponding cysteine to serine replacement mutants to determine their processing sites as well as mass spectrometry of these substrates to determine whether they are lipid modified. Furthermore, *in vivo* analyses of putative Sec substrate lipoproteins in non-haloarchaeal species, such as those predicted to be encoded by the genome of *M. mazei*, a genetically amenable methanogen, may be useful in shedding light on the significance of putative lipobox motifs in archaeal Sec substrates.

TatLipo, the first lipoprotein prediction program primarily trained on the sequences of haloarchaeal Tat lipobox motifs, predicts a vast array of haloarchaeal Tat substrates missed by prediction programs trained solely on bacterial Sec substrates. In fact, our *in silico* analyses predicted an additional 113 haloarchaeal Tat substrate lipoproteins when results generated by TatLipo were included with those of programs trained on bacterial Sec substrates, making the utility of lipoprotein prediction programs that predict archaeal- and Tat-specific lipoproteins abundantly clear. Moreover, TatLipo detected nearly all of the proteins identified by integrating the results of three bacteria-based predictors. In addition, for the subset of Tat-secreted proteins, TatLipo confirmed two-thirds of the predictions supported by only one of these three programs, including Hvo_1242, a prediction that was supported by *in vivo* mutagenesis data.

The inability of prediction programs trained on Sec substrates to recognize a significant portion of the Tat substrate lipoproteins may be due to key structural differences that exist between Tat and Sec signal peptides. In particular, Tat signal peptide hydrophobic stretches are less hydrophobic and their highly charged regions are longer than the corresponding regions of Sec signal peptides [28]. Future *in vivo* analyses of predicted archaeal Sec lipoproteins may help clarify the diversity of archaeal lipoproteins and may also allow the development of an archaeal Sec lipoprotein prediction program, analogous to TatLipo. Moreover, considering the fact that Gram-negative and Gram-positive bacterial Tat substrates that contain lipoboxes have also been identified, programs that will specifically determine the presence of lipoboxes in bacterial Tat substrate signal peptides will also be invaluable. TatLipo provides a solid foundation for the development of such a program.

Acknowledgments

The support was provided to S. Storf by a DFG postdoctoral fellowship, (reference no. Sto 804/1-1), to K. Dilks by an American Heart predoctoral fellowship (reference no. 0415376U), and to S. Imam and M. Pohlschröder by a National Science Foundation grant (Reference no. MCB02-39215). The authors thank Mike Dyall-Smith for helpful discussions. The first and second authors contributed equally to the paper.

References

[1] A. J. M. Driessen and N. Nouwen, "Protein translocation across the bacterial cytoplasmic membrane," *Annual Review of Biochemistry*, vol. 77, pp. 643–667, 2008.

[2] M. Pohlschröder, E. Hartmann, N. J. Hand, K. Dilks, and A. Haddad, "Diversity and evolution of protein translocation," *Annual Review of Microbiology*, vol. 59, pp. 91–111, 2005.

[3] P. Natale, T. Brüser, and A. J. M. Driessen, "Sec- and Tat-mediated protein secretion across the bacterial cytoplasmic membrane—distinct translocases and mechanisms," *Biochimica et Biophysica Acta*, vol. 1778, no. 9, pp. 1735–1756, 2008.

[4] M. Paetzel, A. Karla, N. C. J. Strynadka, and R. E. Dalbey, "Signal peptidases," *Chemical Reviews*, vol. 102, no. 12, pp. 4549–4579, 2002.

[5] M. I. Hutchings, T. Palmer, D. J. Harrington, and I. C. Sutcliffe, "Lipoprotein biogenesis in Gram-positive bacteria: knowing when to hold 'em, knowing when to fold 'em," *Trends in Microbiology*, vol. 17, no. 1, pp. 13–21, 2009.

[6] K. Gan, S. D. Gupta, K. Sankaran, M. B. Schmid, and H. C. Wu, "Isolation and characterization of a temperature-sensitive mutant of *Salmonella typhimurium* defective in prolipoprotein modification," *The Journal of Biological Chemistry*, vol. 268, no. 22, pp. 16544–16550, 1993.

[7] K. Sankaran and H. C. Wu, "Lipid modification of bacterial prolipoprotein. Transfer of diacylglyceryl moiety from phosphatidylglycerol," *The Journal of Biological Chemistry*, vol. 269, no. 31, pp. 19701–19706, 1994.

[8] S. D. Gupta and H. C. Wu, "Identification and subcellular localization of apolipoprotein N-acyltransferase in *Escherichia coli*," *FEMS Microbiology Letters*, vol. 62, no. 1, pp. 37–41, 1991.

[9] A. Tschumi, C. Nai, Y. Auchli et al., "Identification of apolipoprotein N-acyltransferase (Lnt) in mycobacteria," *The Journal of Biological Chemistry*, vol. 284, no. 40, pp. 27146–27156, 2009.

[10] A. Bolhuis, "Protein transport in the halophilic archaeon Halobacterium sp. NRC-1: a major role for the twin-arginine translocation pathway?" *Microbiology*, vol. 148, no. 11, pp. 3335–3346, 2002.

[11] M. Falb, F. Pfeiffer, P. Palm et al., "Living with two extremes: conclusions from the genome sequence of *Natronomonas pharaonis*," *Genome Research*, vol. 15, no. 10, pp. 1336–1343, 2005.

[12] M. I. Giménez, K. Dilks, and M. Pohlschröder, "*Haloferax volcanii* twin-arginine translocation substates include secreted soluble, C-terminally anchored and lipoproteins," *Molecular Microbiology*, vol. 66, no. 6, pp. 1597–1606, 2007.

[13] S. Mattar, B. Scharf, S. B. H. Kent, K. Rodewald, D. Oesterhelt, and M. Engelhard, "The primary structure of halocyanin, an archaeal blue copper protein, predicts a lipid anchor for membrane fixation," *The Journal of Biological Chemistry*, vol. 269, no. 21, pp. 14939–14945, 1994.

[14] K. F. Chater, S. Biró, K. J. Lee, T. Palmer, and H. Schrempf, "The complex extracellular biology of Streptomyces," *FEMS Microbiology Reviews*, vol. 34, no. 2, pp. 171–198, 2010.

[15] D. A. Widdick, K. Dilks, G. Chandra et al., "The twin-arginine translocation pathway is a major route of protein export in *Streptomyces coelicolor*," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 103, no. 47, pp. 17927–17932, 2006.

[16] M. Dyall-Smith, *Halohandbook: Protocols for Haloarchaeal Genetics*, 2008, http://www.haloarchaea.com/resources/halohandbook/Halohandbook_2008_v7.pdf.

[17] G. Bitan-Banin, R. Ortenberg, and M. Mevarech, "Development of a gene knockout system for the halophilic archaeon *Haloferax volcanii* by use of the *pyrE* gene," *Journal of Bacteriology*, vol. 185, no. 3, pp. 772–778, 2003.

[18] M. Holmes, F. Pfeifer, and M. Dyall-Smith, "Improved shuttle vectors for *Haloferax volcanii* including a dual-resistance plasmid," *Gene*, vol. 146, no. 1, pp. 117–121, 1994.

[19] A. Large, C. Stamme, C. Lange et al., "Characterization of a tightly controlled promoter of the halophilic archaeon *Haloferax volcanii* and its use in the analysis of the essential cct1 gene," *Molecular Microbiology*, vol. 66, no. 5, pp. 1092–1106, 2007.

[20] L. B. Blyn, B. A. Braaten, and D. A. Low, "Regulation of pap pilin phase variation by a mechanism involving differential Dam methylation states," *EMBO Journal*, vol. 9, no. 12, pp. 4045–4054, 1990.

[21] T. Allers, H.-P. Ngo, M. Mevarech, and R. G. Lloyd, "Development of additional selectable markers for the halophilic Archaeon *Haloferax volcanii* based on the leuB and trpA genes," *Applied and Environmental Microbiology*, vol. 70, no. 2, pp. 943–953, 2004.

[22] K. Dilks, M. I. Giménez, and M. Pohlschröder, "Genetic and biochemical analysis of the twin-arginine translocation pathway in halophilic archaea," *Journal of Bacteriology*, vol. 187, no. 23, pp. 8104–8113, 2005.

[23] N. Hulo, A. Bairoch, V. Bulliard et al., "The 20 years of PROSITE," *Nucleic Acids Research*, vol. 36, no. 1, pp. D245–D249, 2008.

[24] A. S. Juncker, H. Willenbrock, G. Von Heijne, S. Brunak, H. Nielsen, and A. Krogh, "Prediction of lipoprotein signal peptides in Gram-negative bacteria," *Protein Science*, vol. 12, no. 8, pp. 1652–1662, 2003.

[25] P. G. Bagos, K. D. Tsirigos, T. D. Liakopoulos, and S. J. Hamodrakas, "Prediction of lipoprotein signal peptides in Gram-positive bacteria with a Hidden Markov Model," *Journal of Proteome Research*, vol. 7, no. 12, pp. 5082–5093, 2008.

[26] R. W. Rose, T. Brüser, J. C. Kissinger, and M. Pohlschröder, "Adaptation of protein secretion to extremely high-salt conditions by extensive use of the twin-arginine translocation pathway," *Molecular Microbiology*, vol. 45, no. 4, pp. 943–950, 2002.

[27] L. Käll, A. Krogh, and E. L. L. Sonnhammer, "A combined transmembrane topology and signal peptide prediction method," *Journal of Molecular Biology*, vol. 338, no. 5, pp. 1027–1036, 2004.

[28] B. C. Berks, T. Palmer, and F. Sargent, "Protein targeting by the bacterial twin-arginine translocation (Tat) pathway," *Current Opinion in Microbiology*, vol. 8, no. 2, pp. 174–181, 2005.

[29] K. Dilks, R. W. Rose, E. Hartmann, and M. Pohlschröder, "Prokaryotic utilization of the twin-arginine translocation pathway: a genomic survey," *Journal of Bacteriology*, vol. 185, no. 4, pp. 1478–1483, 2003.

[30] D. J. Näther and R. Rachel, "The outer membrane of the hyperthermophilic archaeon *Ignicoccus*: dynamics, ultrastructure and composition," *Biochemical Society Transactions*, vol. 32, no. 2, pp. 199–203, 2004.

[31] S. Hayashi, S.-Y. Chang, and S. Chang, "Modification and processing of internalized signal sequences of prolipoprotein in Escherichia coli and in *Bacillus subtilis*," *The Journal of Biological Chemistry*, vol. 260, no. 9, pp. 5753–5759, 1985.

[32] W. W. Navarre, S. Daefler, and O. Schneewind, "Cell wall sorting of lipoproteins in *Staphylococcus aureus*," *Journal of Bacteriology*, vol. 178, no. 2, pp. 441–446, 1996.

[33] D. Cavard, D. Baty, and S. P. Howard, "Lipoprotein nature of the colicin A lysis protein: effect of amino acid substitutions at the site of modification and processing," *Journal of Bacteriology*, vol. 169, no. 5, pp. 2187–2194, 1987.

Hindawi Publishing Corporation
Archaea
Volume 2010, Article ID 608243, 13 pages
doi:10.1155/2010/608243

Review Article

Shaping the Archaeal Cell Envelope

Albert F. Ellen,[1,2] Behnam Zolghadr,[2,3] Arnold M. J. Driessen,[2] and Sonja-Verena Albers[3]

[1] *Centre for Integrative Biology, Microbial Genomics, Via delle Regole 101, 38123 Mattarello, Italy*
[2] *Department of Molecular Microbiology, Groningen Biomolecular Sciences and Biotechnology Institute, The Zernike Institute for Advanced Materials, University of Groningen, Kerklaan 30, 9751 NN Haren, The Netherlands*
[3] *Molecular Biology of Archaea, Max Planck Institute for terrestrial Microbiology, Karl-von-Frisch-Straße 10, 35043 Marburg, Germany*

Correspondence should be addressed to Sonja-Verena Albers, albers@mpi-marburg.mpg.de

Received 14 April 2010; Accepted 29 May 2010

Academic Editor: Jerry Eichler

Although archaea have a similar cellular organization as other prokaryotes, the lipid composition of their membranes and their cell surface is unique. Here we discuss recent developments in our understanding of the archaeal protein secretion mechanisms, the assembly of macromolecular cell surface structures, and the release of S-layer-coated vesicles from the archaeal membrane.

1. The Archaeal Cell Envelope

The ability of many archaea to endure extreme conditions in hostile environments intrigues researchers to study the molecular mechanisms and specific adaptations involved. Very early, it was realized that the structure of the archaeal cell envelope differs substantially from that of bacteria [1]. With the only exception of *Ignicoccus* which exhibits an outer membrane enclosing a huge periplasmic space [2], known archaea possess only a single membrane. This cytoplasmic membrane is enclosed by an S-layer, a two-dimensional protein crystal that fully covers the cells (see review Jarrell et al. in this issue). In contrast to bacterial ester lipids, archaeal lipids consist of repeating isoprenyl groups linked to a glycerol backbone through an ether linkage [3, 4]. These lipids typically form diether bilayer membranes similar to membranes of eukarya and bacteria. Hyperthermoacidophiles contain tetraether lipids that consist of C_{40} isoprenoid acyl chains that span the membrane entirely forming a monolayer membrane [5]. These membranes are extremely proton impermeable and enable these organisms to survive under conditions that the extracellular pH is up to 4 units below that of the cytoplasm [6]. Another peculiarity is that most of the extracellular proteins of archaea are glycosylated via N- and O-glycosylation. Finally, Archaea do not produce any murein, and only some methanogenic species are known to produce pseudomurein [7].

As the archaeal cell surface is so different from that of bacteria and eukarya, unique mechanisms must exist to form and shape it. Until recently most of our knowledge of protein secretion and on the assembly of the cell surface components in archaea was obtained by comparative genomic studies. However, in recent years tremendous progress has been made in our understanding of the assembly and function of cell surface structures and both the structural and functional basis of protein translocation across the archaeal membrane. Here we will discuss these topics with an emphasis on the cell surface structures.

2. Protein Secretion

2.1. Transport of Unfolded Proteins Across the Cytoplasmic Membrane. The ability to transport proteins across membranes is vital for cell viability. In general, the systems found in archaea that mediate protein transport across the cytoplasmic membrane are similar to those of bacteria. In archaea most proteins are secreted across the cytoplasmic membrane by the general secretion (Sec) or Twin arginine translocase (Tat) route (see Figure 1). The Sec pathway consists of a universally conserved translocation complex embedded in the membrane, which is termed SecYEG in bacteria and Sec61p in the endoplasmic reticulum (ER) of eukaryotes. The Sec system handles the transport of

unfolded proteins but is also required for the integration of membrane proteins into the cytoplasmic membrane [8]. In bacteria, the SecYEG complex either associates with the ribosome for cotranslational membrane protein insertion or with the motor protein SecA, to catalyze posttranslational protein translocation. In the ER, Sec61p associates with the ribosome for co-translational protein translocation and membrane protein insertion and Sec61p associates with the Sec63p complex and the ER luminal chaperone BiP for post-translational protein translocation. The core of the protein-conducting channel is composed of two essential components, SecY and SecE in bacteria and Sec61α and Sec61γ in eukaryotes [9]. Both proteins are found in all archaea but the third, nonessential component, that is, SecG in bacteria or Sec61β in eukaryotes, was identified only after extensive bioinformatic analyses [10, 11]. In this respect, the archaeal SecG homolog is more related to the eukaryotic Sec61β than to the bacterial SecG. Therefore, the archaeal translocon is often referred to as the SecYEβ complex [12]. The exact composition of the minimal protein translocase of Archaea has, however, remained unclear. Archaea lack a homolog of the bacterial SecA motor protein, a protein that is well conserved among bacteria and the chloroplast thylakoid [8]. Likewise, Archaea also do not contain homologs of the eukaryal Sec63p complex, but they do contain DnaK (or Hsp70) chaperones homologous to BiP. These chaperones fulfill general functions in protein folding but in analogy with the ER, a BiP homolog involved in protein translocation would need to be extracellular. However, no archaeal Hsp70 homolog has been detected extracellularly and of course the energy source ATP would be absent. Therefore, it is generally assumed that protein translocation is co-translationally coupled to chain elongation at the ribosome [13]. However, in the euryarchaeon *Haloferax volcanii,* it was noted that some proteins are present as fully synthesized signal peptide bearing precursors in the cytoplasm before they are secreted. Based on this finding, it has been proposed that post-translational protein secretion also exists in archaea [14]. Interestingly, euyarchaeota contain a homolog of the bacterial SecDF protein complex [15], whereas this protein is absent from crenarchaeota. The exact role of SecDF is unknown, but it has been implicated in the proton motive force-dependent release of translocated proteins from the periplasmic face of the membrane. SecDF is not essential for translocation per se, but it enhances the rate of translocation. Other suggested roles of SecDF are that it may act on the SecA ATPase catalytic cycle but since SecA is absent from archaea such a role seems unlikely.

The structural analysis of the *Methanocaldococcus jannashii* SecYEβ heterotrimer [12] has provided important insights in how this channel may function in protein translocation. The main subunit SecY consists of two halves with an internal pseudo-twofold symmetry. These two halves comprise transmembrane segments (TMSs) 1–5 and 6–10, respectively, and are connected by a hinge region. In this organization, the channel resembles a clamshell that encompasses a central hourglass-shaped pore with a narrow constriction ring in the middle of the membrane. This ring is lined by hydrophobic amino acid residues and is proposed to prevent leakage of ions in the "closed" state. SecE embraces the SecY clamshell at the hinge side in a V-shaped manner. The third subunit, Sec61β is peripherally associated with the SecYE complex. The pore-like opening in the center is obstructed by a plug-like domain also termed TMS 2a that resides at the periplasmic side of the constriction ring. Thereby, it closes the pore on the extracellular face of the membrane. In the clamshell organization of SecY, the two halves contact each other via TMS 2, TMS 7, and TMS 8. The opening between TMS2 and TMS7/8 is termed the lateral gate and localizes at the front of the SecY pore. When opened, it may provide an exit path for hydrophobic polypeptide segments to enter the membrane. The lateral gate also fulfills an important role in the channel opening mechanism during protein translocation [16]. It is believed that insertion of the signal sequence into the lateral gate region results in a widening of the central constriction and an opening of the channel. This in turn will destabilize the plug domain that once released from the extracellular funnel will vacate a central aqueous path for polar polypeptides to cross the membrane. Because of the high conservation of the core subunits of the translocon, the proposed mechanism of channel opening is likely conserved in all domains of life [8]. In this respect, it is remarkable that the structural work with the archaeal SecYEβ complex has been instrumental to define a unifying mechanism of protein translocation despite the fact that the exact details of this process have not been

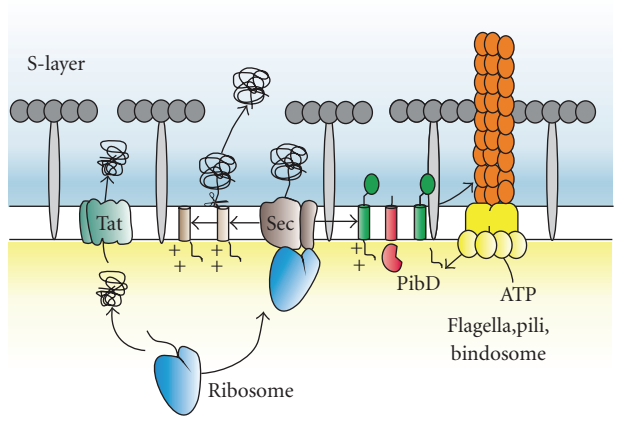


Figure 1: Model of the archaeal cell envelope showing different characterized secretion pathways. Proteins synthesized at the ribosome can follow several routes to the exterior of the cell. During co-translational translocation, the ribosome-nascent chain complex is targeted to the SecYEβ complex by the signal recognition particle. At the SecYEβ complex protein synthesis and translocation across the cytoplasm membrane occurs simultaneously. In the case of a preprotein with a class I signal peptide, the signal peptide is removed during translocation and the protein is released and folds at the external face of the membrane. Class III signal peptide containing proteins translocated via the SecYEβ complex are processed by PibD and subsequently assembled into a flagellum, pilus, bindosome or so far unknown cell surface structures. Alternatively, folded proteins are transported across the cytoplasmic membrane via the Twin arginine translocase pathway.

resolved in archaea as so far no *in vitro* translocation system has been established.

2.2. Transport of Folded Proteins Across the Cytoplasmic Membrane. The Tat pathway mediates the transport of protein in their folded state. This in particular, but not only, concerns cofactor containing proteins that fold and assemble in the cytoplasm. Typically, the bacterial Tat-pathway consists of three integral membrane proteins, TatA, TatB, and TatC. In archaea and in most Gram-positive bacteria, the Tat complex consists of only two components, TatA and TatC, whereas the third component TatB is missing [11]. In current models, TatBC is involved in the initial recruitment of a substrate while TatA, probably in concert with TatC, forms the pore through which the folded protein is transported across the membrane [17]. In most bacteria and archaea, the number of Tat substrates is relatively small as compared to the number of substrates that are translocated by the Sec pathway. However, in halophilic archaea the Tat pathway is the predominant route for protein secretion [18]. This requirement for the Tat-pathway is thought to be an adaptation to the high-salt environment that may interfere with protein folding inside of the cell. However, the halophilic bacterium *Salinibacter ruber* mostly secretes proteins via the Sec route [19] suggesting that the requirement for Tat is not an adaptation to high salt per se. Another unique feature of the Tat pathway in haloarchaea is that translocation is driven by the sodium motive force whereas in many other microorganisms, the proton motive force is used as a driving force [20]. It should be noted that in the bacterium *Streptomyces coelicolor*, many of the proteins that are typically secreted by the Sec-pathway utilize the Tat pathway instead [21].

Proteins are routed to either the Sec or Tat pathway by an N-terminal signal peptide that upon secretion is removed by a signal peptidase. The basic tripartite organization of the signal peptides utilized by these two pathways is very similar. The Sec and Tat signal peptides have a three-domain structure: a positively charged amino-terminal n-domain, a central hydrophobic h-domain, and a polar c-domain which contains a cleavage site for the signal peptidase [22]. Apart from the presence of a pair of arginines in a SRRXFLK (X = any amino acid) motif in the N-region of Tat signal peptides [23], there is no sequence homology in the other regions. The signal peptides of the three domains of life are functionally interchangeable [24]. Remarkably, about 60% of the Tat signal sequences in *Escherichia coli* are able to route proteins to the Sec translocation machinery as well [23]. In this respect, unfolded proteins are rejected by the Tat pathway [25], although some other studies suggest that the Tat pathway can handle intrinsically unfolded proteins [26].

2.3. Transport Across the Outer Envelope. The most outer border of the archaeal cell is usually a layer of crystalline protein, that is, the surface (S-) layer. The S-layer contains pore-like openings that have suggested to allow free passage of nutrients and other small molecules [1]. However, little is known on how proteins cross this barrier during secretion. Protein secretion across the outer envelope, the outer membrane, has been studied in great detail in didermic bacteria. A total of seven different systems have been recognized in these organisms and the protein secretion processes associated with these systems are termed type I-VII secretion. Archaea share components of some of these systems, but since types III, V, VI, and VII secretion seem to be absent from archaeal genomes, these will not be further discussed here.

Type I secretion involves an ATP-binding cassette (ABC) transporter that via a cytoplasmic membrane bound fusion (or adaptor) protein (MFP) associates with an outer membrane pore [27]. These systems secrete proteins directly from the cytoplasm to the exterior of the cell. ABC type transporters are relatively abundant in archaea but most are involved in substrate uptake [11]. It is not clear if type I secretion exists in archaea. However, no homologues have been identified of the membrane fusion proteins and porin proteins are absent because of the lack of an outer membrane. Proteomic studies in thermophilic crenarchaea show that a significant portion of the exoproteomes concerns proteins devoid of signal sequences. For instance, in the thermoacidophile *Sulfolobus solfataricus* secretion of a superoxide dismutase has been reported [28], but the gene encoding this protein does not specify a signal sequence and thus it remains unknown how this protein is released from the cells. Therefore, it remains to be established whether the presence of signal sequenceless proteins in the external medium is the result of a specific protein secretion process or cell lysis [29–31].

Type II secretion systems of didermic bacteria consist of 12 to 16 proteins that assemble into a secretion apparatus that spans both the cytoplasmic and outer membrane. The genes coding for the secretion system are often arranged into a large operon. With type II secretion, substrate proteins are first translocated to the periplasm by either the Sec- or Tat pathway [32, 33]. These proteins fold into their native state in the periplasm and may even assemble into multisubunit protein complexes. Next, these folded proteins are translocated across the OM through a large pore termed the secretin. The targeting of proteins to the secretin is poorly understood. For example, *Pseudomonas aeruginosa* secretes various proteins, such as a lipase, an elastase, and exotoxin A, via its type II secretion systems but these substrates share no common recognition motif and it is generally believed that the secretin recognizes structural folds rather than amino acid sequences [32]. Transport through the secretin is believed to involve a pseudopilus, a short filament that assembles from subunits at the cytoplasmic membrane. It has been proposed that the pseudopilus acts as a kind of piston to push substrates through the secretin across the outer membrane [32]. Although archaea do not possess an outer membrane, their flagella and pili assembly systems contain subunits reminiscent to proteins in the type II secretion systems of bacteria and will be discussed in more detail below.

Type IV secretion systems are involved in the transport of effector proteins and of DNA, but are considered to be

primarily protein exporters that secrete DNA through its attachment to a secreted protein [34]. Very recently the structure of the type IV secretion channel was solved. This structure that contains 4 different subunits spans the entire periplasmic space and resides in the cytoplasmic and outer membrane [35]. Conjugative plasmids containing some subunits of type IV secretion systems have been identified in crenarchaea only [36–39]. In these homologs of the cytoplasmic ATPase VirB4, the polytopic membrane protein VirB6 and the coupling protein VirD4 were identified, but these are significantly different than their bacterial counterparts. No details are known about their involvement in conjugative transfer of DNA in archaea. In the euryarchaeote *Haloferax volcanii,* it was reported that bidirectional chromosomal DNA transfer occurred during conjugation, and large structures (2 μm long and 0.1 μm wide) bridging cells were postulated to mediate DNA transfer [40]. However, the system mediating this transfer has not been identified.

Yet another well-studied system in bacteria is the assembly machinery of type IV pili that are involved in a multitude of functions such as surface adhesion, cell-cell contact, autoaggregation, twitching motility, and DNA uptake [41]. Type IV pilins contain the so-called class III signal peptides that prior to the pilus assembly reaction are processed by PilD, a processing peptidase that also methylates the N-terminal phenylalanine of the mature pilin [42]. Up to 15 proteins are involved in the correct assembly of the pilins into the pilus structure, but the driving force for its assembly is provided by the cytoplasmic ATPase PilB. This process is antagonized by the action of the ATPase PilT causing the disassembly of the pilus. Interestingly, the archaeal flagellum biogenesis apparatus resembles a simplified type IV assembly machinery and different archaeal surface structures have been identified which belong to the same class [43] (more details will be discussed in the section about archaeal surface structures).

All type II/IV secretion and type IV pili assembly systems contain a cytosolic ATPase that functions as a motor to drive secretion or assembly. Because of the similarity, these ATPases likely function by similar mechanisms and are evolutionary related [44]. Secretion ATPases assemble into a hexameric ring. The structure of the secretion ATPase GspE2 of *A. fulgidus* shows that the N-terminal domain alternates between a standing and laying down position, and it has been suggested that this process is driven by ATP and needed to deliver a piston-like movement that would drive the movement (or assembly) of a pilus [45]. The relative shift of the N-terminal domain is 10 Å which fits to the required movement of 10.5 Å for pilus assembly [45]. The genomes of most archaea contain genes specifying several type II/IV secretion ATPases [45]. These are often arranged in an operon together with genes encoding pilin-like proteins and a membrane protein. Therefore, it appears that the archaeal assembly systems are of a lower complexity than their didermic bacterial counterparts, at least lacking the outer membrane protein components. In this respect, they are more similar to those observed in monodermic bacteria.

3. Signal Peptides and Secretomes

Three different classes of signal peptides which are processed by their own designated signal peptidase have been recognized [46]. Class I signal peptides are cleaved at the C-domain by type I signal peptidases. Proteins containing class I signal peptides are typically released as soluble proteins or are, if they contain a C-terminal transmembrane helix, C-terminally embedded in the membrane [47]. Class II signal peptides are exclusively found in lipoproteins. Characteristic of class II signal peptides is a conserved cysteine that is present at the cleavage site. After cleavage of the signal peptide, the cysteine forms the N-terminal residue of the mature protein where it serves as a lipid attachment site to anchor the protein to the membrane [48]. In bacteria, several steps are involved in processing of the class II signal peptide. First, a diacylglyceryl group is attached to the cysteine. This reaction is catalyzed by prolipoprotein diacylglyceryl transferase. After this modification the signal peptide is cleaved by the type II signal peptidase. The final step, that is, the attachment of a lipid, is then executed by an apolipoprotein N-acyltransferase. Peculiarly, none of the proteins involved in processing of class II signal peptides have been identified in archaea, despite the presence of functional class II signal peptides [49]. In archaea, Sec and Tat signal peptides can be found in both class I or class II signal peptides [46, 48]. Class III signal sequences are processed at the N-domain by a specific membrane-integrated peptidase that eliminates the positively charged amino acids, thus, leaving the H-domain of signal peptide attached to the protein. This processing event occurs at the inner face of the cytosolic membrane, and because of the removal of the positive charges the translocation block is removed allowing the subsequent translocation of the pilin subunit for downstream assembly. The latter involves the H-domain that functions as an assembly scaffold to support the formation of a pilus or pseudopilus on the outside of the cell [41, 42]. In archaea, the best example of a class III signal peptide bearing substrate is flagellin, the subunit of the archaeal flagellum that is used for motility. The class III signal peptides are processed by a specialized peptidase, that is, the preflagellin peptidase that utilizes the same catalytic mechanism as the bacterial prepilin peptidases [50, 51]. However, in archaea, class III signal peptides are not only confined to flagellins, pilins, and/or pseudopilins but are also found in a variety of other extracellular proteins such as substrate-binding proteins or proteases [52].

The signal peptide plays a decisive role in initiating the secretion process. In co-translational protein secretion, the protein synthesizing ribosome is brought to the transport machinery by a protein-RNA complex called Signal Recognition Particle (SRP). The SRP binds to the signal peptide of the protein being synthesized and to the ribosome. The ribosome-SRP complex interacts with a membrane-associated SRP receptor and upon entry of the signal peptide into the Sec translocon the SRP and SRP receptor are released [53]. In eukaryotes, the SRP contains six proteins together with a 300 nucleotide RNA molecule, whereas the bacterial version is much simpler as it consists of one protein, Ffh,

and a 113 nucleotide RNA molecule. The archaeal SRP is similar to the eukaryote SRP albeit much smaller. It consists of two essential components; the SRP54 protein and a~ 300-nucleotide-long RNA molecule and the nonessential accessory protein SRP19 [54]. The archaeal SRP receptor is more similar to the bacterial SRP receptor FtsY than to the eukaryotic SRP receptor that consists of two subunits, SRα and SRβ [55].

3.1. The Secretome. Current knowledge of protein secretion and the advancement of proteomics led researchers to define the secretome [56] which is the collection of proteins that is secreted by the cell. Essentially, these are the proteins that contain a signal peptide and that are actively transported across the cytoplasmic membrane, but proteomic studies have also identified sets of secreted proteins that do not contain an identifiable signal peptide but still can be regarded as secreted. In principle any program able to detect the presence of signal peptides can be used to create an *in silico* secretome. For example, PSORTb predicts the cellular localization of a protein and SignalP predicts the likelihood that a protein contains a signal peptide [57, 58]. By means of these prediction programs, various *in silico* secretomes of archaea have been drafted [30, 46, 59–61]. These vary from 1.2 up to 19% of the total proteome depending on the specific program, stringency of criteria, and the archaeal species analyzed. Of special interest are the programs PRED-SIGNAL and Flafind [52, 62]. PRED-SIGNAL has been designed exclusively for the prediction of archaeal signal peptides, while it also distinguishes between signal peptides and amino-terminal transmembrane helices. Analysis of 48 archaeal genomes by PRED-SIGNAL predicts that 5%–14% of the proteome specifies signal peptide-containing proteins, while no significant differences between crenarchaea and euryarchaea were found [62]. The program Flafind recognizes class III signal peptides, which in archaea are believed to be particularly important for the biogenesis of cell surface appendages. Flafind indicated the presence of 308 class III signal peptide-bearing proteins amongst 22 archaeal proteomes [52]. The majority of the Flafind positives are hypothetical proteins that are associated with pilus assembly systems.

A critical issue is the experimental validation of the *in silico* secretomes. In the supernatant of the psychrophile *Methanococcoides burtonii* only 7 signal peptide-containing proteins have been identified [47]. In a later study, this number was increased to 16 proteins by applying a whole proteome analysis [63]. In *S. solfataricus,* attempts to cover the whole proteome resulted in the identification of 32 proteins exclusively present in the supernatant [31]. When an inventory was made of supernatant proteomes and cell surface subproteomes of three *Sulfolobus* species, a total of 64 proteins was reported [29]. In these *Sulfolobus* species, cell surface proteins dominated the supernatant proteome suggesting that actual secretion is a rare event and that the majority of the secreted proteins originate from cell surface released proteins. This notion was further strengthened by the observation that an extracellular α-amylase mostly resides at the cell surface [29]. Similar observations were made in the crenarchaeon *Aeropyrum pernix* in which 107 proteins were identified from both the cell surface and the supernatant [30]. The proteomic studies demonstrate that there are significant differences between predicted and experimental secretomes. For example, proteins devoid of an identifiable signal peptide are not predicted by the *in silico* methods but appear in large numbers extracellularly. An important source of proteins without signal peptides are those associated with extracellular membrane vesicles that appear to result from a specific secretion phenomenon (discussed below). It has been suggested that cytosolic proteins are secreted via yet unknown secretion systems [30], but this phenomenon appears general in proteomic studies in both bacteria and archaea and often concerns different proteins. Overall, these cytosolic proteins may be highly resistant against proteolysis and, therefore, show a long retention time in the external medium after cell lysis. None of the proteomic studies has achieved a full coverage of the *in silico* secretome. The latter is due to various limitations in the analysis. Often only one growth condition is used, and thus only a subset of proteins is expressed. Also, the methods are not optimized for the isolation of the extracellular cell surface associated proteins, and only those are observed that are released. By isolating the glycosylated cell surface proteins using lectin columns [29, 64], the set of identified extracellular proteins may be significantly expanded.

4. Membrane Vesicles as a Novel Secretion Vehicle

A rather unusual and poorly understood protein secretion mechanism is the release of proteins packaged into small membrane vesicles that emerge from the cell surface. Many didermic bacteria are known to release outer membrane vesicles from their surface [65], but this process also seems to occur in archaea where the membrane vesicles are coated with S-layer proteins. In a screen for viruses amongst the euryarchaeal order of *Thermococcales* it was discovered that most of the strains tested released small spherical vesicles [66]. These vesicles do not resemble viruses and often have genomic DNA associated to their surface [66]. Membrane vesicle release has been reported for many different archaea, such as the thermophilic euryarchaeon *Aciduliprofundum boonei* isolated from hydrothermal deep-sea vents [67], and various crenarchaeota, in particular *Sulfolobus* [68, 69]. With *S. islandicus* [70] and *S. tokodaii* [68] (Ellen et al, unpublished), the membrane vesicles appear to contain an antimicrobial protein(s) that inhibits the growth of related *Sulfolobus* species. The antimicrobial activity involves a proteinaceous component, but its identity has not yet been elucidated. Overall, it seems that in *S. tokodaii,* the antimicrobial protein(s) is specifically sorted to the membrane vesicles, but it is unknown if membrane vesicle formation is mechanistically linked to the secretion of the antimicrobial protein factors. Also *Ignicoccus* species are vigorous producers of membrane vesicles. These organisms lack a cell wall and instead contain an outer membrane-like structure. Electron microscopic investigations indicate

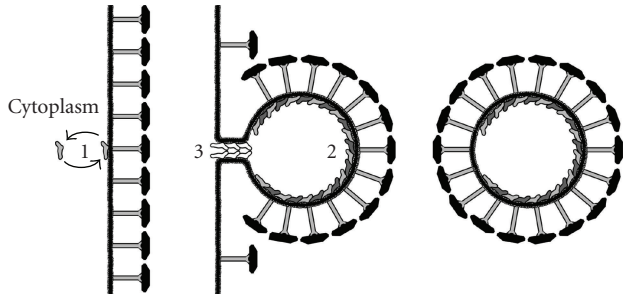


Figure 2: Model for vesicle budding in crenarchaea. Archaeal homologues of eukaryote ESCRT-III subunits are in equilibrium between a freely diffusible state in the cytoplasm and a membrane-bound state (1). If the equilibrium shifts towards the membrane associated state a heterocomplex (2) of different ESCRT-III subunits is formed leading to the creation of an outwardly growing bud that is covered by S-layer protein. Recruitment of the last group of ECRT-III subunits (3) creates the "neck" through which the bud is attached to the cytoplasmic membrane just before the membrane vesicle is pinched off and released into the medium.

that membrane vesicles are released from the cytoplasmic membrane and released in the spacious periplasmic space [2]. It has been suggested that these vesicles fuse with the outer membrane and that they are either part of a specific secretion system or involved in the biogenesis of the outer membrane.

To date, only for the *Sulfolobus* derived vesicles a proteomic analysis has been performed. The protein composition of these membrane vesicles is markedly different from that of the cytoplasmic membrane [68] suggesting that they may emerge from a specific release event. However, the vesicles do not seem to contain a specific cargo that would point to a specific role, except for the presence of archaeal homologues of the eukaryotic endosomal sorting complex required for transport-I (ESCRT) proteins [68]. This has led to the hypothesis that the membrane vesicles emerge from the cytoplasmic membrane through an outward budding event similar to the inward budding of vesicles in the endosomal compartment of eukaryotes (see Figure 2). The *Sulfolobus* vesicles vary in size from 50 to 200 nm and are surrounded by a S-layer, as verified by proteomic analysis and electron diffraction [70]. The presence of the S-layer coat indicates that the membrane vesicles are pushed through the cell envelope, which would be consistent with an assumed flexibility of the S-layer. The ESCRT-III proteins have also been implicated in cell division [71], and another possibility would be that the membrane vesicles are remnants of the cellular constriction and released during the cell division processes. Intriguingly, ESCRT-III proteins are not present in euryarchaea, although membrane vesicle formation has also been observed in these archaea.

The release of membrane vesicles appears a general feature observed in all three domains of life. In this respect, despite the presence of a cell wall, membrane vesicle release has also been reported for monodermic bacteria and fungi [72, 73]. In didermic bacteria, release of outer membrane vesicles is commonly observed feature and some indirect genetic evidence suggests that this is an essential process [74]. The protein composition of the outer membrane vesicles (or blebs) differs significantly from that of the outer membrane, suggesting that proteins are specifically sorted to the vesicles [75]. The exact function of membrane vesicle release has remained obscure as they have been implicated in a variety of processes. The membrane vesicles may function as a protein secretion system to provide a protected environment for the cargo. For instance, in *E. coli* α-haemolysin is secreted via a type I secretion system. However, the majority of the α-haemolysin remains tightly associated with outer membrane vesicles that also contain TolC, the outer membrane porin associated with the haemolysin type I secretion system. This suggests a link between the secretion of a membrane active toxin and membrane vesicle formation [76]. Membrane vesicle release may be a stress phenomenon providing a means to get rid of excess membrane material. In many cases, DNA seems to be associated with the membrane vesicles. For *Thermococcales*, it has been suggested that the associated DNA is not specifically packaged into the membrane vesicles but rather associates with the membrane vesicles after their release into the medium [66]. The DNA may originate from lysed cells, and because of the membrane association, it may become resistant to nuclease activity and, thus, show a greater persistence. Finally, membrane vesicle release may provide a means to secrete insoluble hydrophobic substances that partition into the lipid membrane. For example, many microorganisms produce quorum-sensing molecules with hydrophobic acyl chains of varying lengths. In *Pseudomonas aeruginosa* such quorum-sensing molecules are packaged into outer membrane vesicles [77]. The release of membrane vesicles could also serve to restore cellular imbalances caused by aggregates of denatured proteins as suggested for *E. coli* [78]. Future studies should reveal the exact function of the secreted membrane vesicles in archaea and provide clues on their mechanism of biogenesis.

5. Assembly of Archaeal Surface Structures

5.1. Archaeal Flagella: Structure and Function. Archaeal flagella have been studied at the genetic, structural, and functional level for several archaeal strains. Early observations of these pili-like filaments by electron microscopy led to the suggestion that they are functionally analogous of bacterial flagella performing similar tasks in swimming motility and biofilm formation. Cell motility by flagella has been demonstrated for the archaea *Halobacterium salinarum*, *M. voltae*, *S. acidocaldarius* and *S. solfataricus [79–83]*. In *H. salinarum*, the bidirectional rotation of the flagellum creates a motion to forward or reverse direction by instant switching of the flagellum rotation which appears to be similar to the rotation of bacterial flagellum [82]. Such a rotational motion has not yet been observed for other archaeal flagella. The flagella are also essential for surface attachment and colonization as demonstrated for *Pyroccocus furiosus* and *S. solfataricus* [84–86].

The subunit composition, structure, and assembly mechanism of the archaeal flagellum is very different from that of

the bacterial flagellum [87, 88]. The archaeal flagellum has a right-handed helical subunit packaging with a diameter of approximate 10–14 nm which is much thinner than the bacterial flagellum [80, 89]. Only in few cases thicker filaments were found depending on the flagellins assembled [90]. The archaeal flagellum is not hollow and the inner space is most probably formed by coiled-coil interaction of the N-terminal hydrophobic domains of the flagellins similar to the assembled type IV pilus [91]. Moreover, recent studies suggest that the energy required for the rotation of the *H. salinarum* flagellum is directly gained from ATP hydrolysis and not from the proton motive force. Therefore, the mechanism of the *H. salinarum* flagellum rotation is fundamentally different from that of the bacterial system [92]. The archaeal flagellum is encoded by the *fla* operon, a single locus of 8–10 genes present in many Crenarchaeota and Euryarchaeota. The overall composition of the *fla*-operon shares homology with bacterial type-IV pili assembly, type II and type IV secretion systems [52, 80, 93–96]. Flagellins are the subunits of the flagellum and contain a class III signal peptide that is necessary for their membrane insertion and assembly into the flagellum. Processing involves the membrane peptidase FlaK (or PibD) [51, 97], and these enzymes are homologous to the bacterial PilD but do not catalyze the N-methylation of the newly formed N-terminus of the flagellin subunit. The H-domain likely folds into an extended hydrophobic α-helix that participates in coiled-coil interactions between subunits within the inner core of the flagellum. Reconstruction studies of the *H. salinarum* and *S. shibatae* flagella suggests that the H-domains constitute a central hydrophobic core similar to that of type-IV pili, but there is no direct evidence for a structural role of the H-domain [98, 99].

Archaeal flagella differ in the number of the structural subunits, the flagellins. The *fla* operon of *M. voltae* contains 4 structural flagellin genes: *flaA*, *flaB1*, *flaB2,* and *flaB3* [100]. FlaB1 and FlaB2 are the major components of the flagellum and the deletion of their corresponding genes results in flagellum deficiency. FlaA is distributed throughout the flagellum as a minor component and deletion of *flaA* results in flagellated but less motile mutants [81]. FlaB3 is localized proximal to the cell surface forming a curved shape structure with similarity to the bacterial hook structure. Deletion of *flaB3* resulted in flagellated and motile mutants [101]. The similarity between this suggestive archaeal hook structure and the hook domain of bacterial flagella may indicate that a similar torque-driven motion is generated by the *M. voltae* flagellum. However, the mechanism of *M. voltae* motility is unknown and the role of the archaeal hook in rotation of the flagellum has not been demonstrated. In *H. salinarum*, five *fla* genes in two loci (*flaA1*, *flaA2* and *flaB1*, *flaB2, flaB3*) encode flagellum subunits [102–104]. The *flaA1* and *flaA2* genes encode the major components of the flagellum. The flagellum of *H. salinarum* does have a bi-directional rotation mechanism which drives the cells forward and backwards [82].

Possibly, the central core complex encoded by the *fla*-operon is only involved in assembly of the flagellum much akin that of bacterial type IV pilins, while another as yet unknown system functions as the rotating motor. The *Sulfolobales fla* operon contains only one structural flagellin gene, FlaB [80, 105]. In *P. furiosus*, FlaB1 is the main component of the flagellum, but the *fla* operon contains a second flagellin subunit (FlaB2) with unknown function [84]. FlaI is homologous to the bacterial type IV pili assembly and type II secretion ATPases, PilB and GspE, respectively. This further suggests a conserved mechanism for assembly of the archaeal flagellum and bacterial type IV pili assembly/type II secretion systems [89, 94–96]. ATPase activity was demonstrated for *S. solfataricus* and *S. acidocaldarius* FlaI proteins expressed and purified after overexpression in *E. coli* [94, 106]. So far, FlaI is the only identified ATPase component of the flagellum core complex and although its role in flagellation has been demonstrated with the deletion of the *flaI* gene, it remains unclear if FlaI is also involved in energizing the motility of the cell. FlaJ is the only known integral membrane component of the flagellar assembly system [79, 80, 101]. FlaJ proteins contain 9 transmembrane segments and two large cytoplasmic domains of about 25 and 15 kDa, respectively. These polar domains are thought to function as the interaction site for FlaI as shown for the membrane anchoring proteins of bacterial type II secretion systems. Structural analysis of the interacting domains of EpsE and EpsN, the assembly ATPase and the membrane protein of the toxin type II secretion system of the bacterium *Vibrio cholerae,* indicated that hydrophobic interactions and salt bridges are responsible for this interaction [107]. Alignment of archaeal FlaI/FlaJ with EpsE/EpsN suggests that this interaction might be conserved in the archaeal type IV pili assembly systems. The function of FlaJ in flagella assembly has not been examined. Although the flagellum of *S. solfataricus* is essential for motility on surfaces [80], a rotational motion and a hook-like structure in the flagellum filament remain to be demonstrated. Overall, the mechanism for twitching motility by means of the archaeal flagellum is poorly understood.

The function of the other components of the archaeal flagellum assembly operon is unknown, however, in *H. salinarum,* it was recently demonstrated that the flagella accessory proteins FlaCE and FlaD interact via two newly identified proteins with three different proteins from the Che signaling cascade (CheY,CheD, and CheC2), providing the link between the flagellum and the sensory apparatus [108]. As Che proteins are lacking in crenarchaeotes also the FlaCEDs are absent in the flagella operon implying a different mechanism for how stimuli will be transduced into a change of motility direction.

5.2. Novel Archaeal Surface Structures. Archaea exhibit a wide variety of cell surface appendages with intriguing structures and biological functions. These appear to be highly specialized due to the specific adaptation of the microorganisms to their hostile habitats. The cannulae network of *Pyrodictium abyssi* is an example of such a structure [109, 110].

P. abyssi has been isolated from hydrothermal marine environments and its optimal growth temperatures range from 80 up to 100°C [111, 112]. The cannulae network seems crucial for cell survival as it is highly abundant in the cell colonies. Cannulae tubes have an outside diameter

of 25 nm and they consist of at least three different, but homologous, glycoprotein subunits with identical N-termini but with different molecular masses (i.e., 20, 22, and 24 kDa). These proteins are highly resistant to denaturing conditions such as exposure to temperatures up to 140°C. From the three-dimensional reconstruction of the cannulae-cell connections, it appears that cannulae enter the periplasmic space but not the cytoplasm forming an intercellular connection of the periplasmic spaces between cells [109]. These connections are formed when cells divide whereupon the cells stay connected through the growing cannulae [111]. The function of the cannulae network is still unclear. It might act to anchor cells to each other or function as a means of communication, mediate nutrients exchange, or even transport of genetic material [87]. It is also not known which system(s) is (are) involved in the assembly of the cannulae network.

Another unusual archaeal cell surface appendage is the "hamus" [87, 113]. This structure represents a novel filamentous cell appendage of unexpectedly high complexity. Archaeal cells bearing these structures are found in macroscopically visible string-of-pearls-like arrangements which also entangle bacterial cells mainly *Thiothrix* (SM) or IMB1 proteobacterium (IM) that grow in cold (10°C) sulfidic springs [114]. The archaeal cells are coccoids of approximately 0.6 μm in diameter with about 100 filamentous hami attached to each cell. Hami are 1 to 3 μm in length and 7 to 8 nm in diameter and have a helical structure with three prickles (each 4 nm in diameter) emanating from the filament at periodic distances of 46 nm. The end of filament is formed by a tripartite, barbed grappling hamus-like hook. The hamus is composed mainly of a 120-kDa protein. However, the sequence of this protein is unknown. They are stable over a broad temperature (0 to 70°C) and pH range (pH 0.5 to 11.5) and mediate strong cellular adhesion to surfaces of different chemical compositions. It is proposed that the hami function in surface attachment and biofilm initiation, much like flagella and pili in bacterial biofilm formation, but in addition provides a strong means of anchoring.

A new pili type was recently isolated from *Ignicoccus hospitalis* which are 14 nm in width and up to 20 μm in lenth and constitute up to 5% of cellular protein. They are composed mainly of protein Iho670, which has a class III signal peptide [115]. As *I. hospitalis* has an outer membrane, it would be expected that the pili assembly would be located in the outer membrane instead of the inner membrane as in all other known archaea.

S. solfataricus expresses UV-induced pili at its cell surface [116]. This system is encoded by the *ups* operon and present in all S*ulfolobales* genomes [94]. This operon is strongly induced when *S. solfataricus* is exposed to UV light; subsequently the cells assemble pili at their surface and form large cellular aggregates. The Ups pili are much shorter than the wave-shaped flagella of *S. solfataricus* and are relatively thin with a diameter of 7 nm [80]. They show a right-handed helical symmetry similar to the flagellum. Mutants lacking the *upsE* gene that encodes a GspE-like ATPase are deficient in pili formation and cell aggregation. UpsE shares strong homology with FlaI and other assembly ATPases, and it likely energizes the assembly of the Ups pili. The *upsF* gene encodes the transmembrane protein of the assembly system and is it highly homologous to FlaJ. Another gene in the operon is *upsX*. UpsX shows no homology with any other protein and its function is unknown. The *ups* operon contains two genes that encode pilins, UpsA and UpsB. Both proteins contain a class III signal peptide and are processed by the general class III signal peptidase PibD. Overexpression of UpsA in *S. solfataricus* results in the formation of unusual long pili. Interestingly, the Ups pili are also essential for surface adhesion of *S. solfataricus* [86]. The Ups system and the flagellum can initiate the attachment of *S. solfataricus* to different surfaces and recent studies on *Sulfolobales* biofilm formation reveal the Ups system is essential for lateral biofilm formation (Koerdt and Albers, unpublished).

Recent studies on the flagella and novel pili structures promoted an initiative to map archaeal pili-like biogenesis clusters through bioinformatics analysis of a large number of sequenced archaeal genomes [52]. The FlaFind program was developed to search for proteins containing class III signal sequences, which therefore encode putative structural surface proteins. This *in silico* analysis identified 388 putative class III signal sequence-containing proteins in 22 archaeal genomes, from which 102 proteins were annotated with a function: 44 flagellin subunits and 33 as substrate-binding proteins. Also extra cellular proteases and redox proteins were among this list. A total of 120 of these proteins were found connected to operons similar to bacterial type IV pilus assembly systems and type IV pilin signal peptidases. The FlaFind hits were analyzed for short and highly conserved motifs. Also eight additional SBP and 19 euryarchaeal proteins containing a QXSXEXXXL motif with unknown function were identified. In the DUF361 domain, the Q residue was at +1 from the cleavage site. Several of these proteins were identified in an operon together with a novel type IV signal peptidase called EppA from euryarchaeal *Methanococcus maripaudis*. Experiments showed that EppA specifically processes proteins belonging to the DUF361 group. The cleavage was tested by coexpressing a DUF361-containing protein with FlaK and EppA. It is probable that the DUF361 proteins are functionally and structurally different than the well-known flagellin and pilin proteins due to the requirement of a homologue but yet different type IV signal peptidase for the cleavage of their signal peptide. Recently, the structure of the *M. maripaudis* pilus has been resolved with cryo electron microscopy and it revealed a novel structure assembled from two subunit packaging [117]. A one-start helical symmetry filament and a ring structure of 4 subunits were combined in the same filament.

Another intriguing archaeal type IV pilus assembly system is the bindosome assembly system (Bas) in *S. solfataricus* which is involved in assembly of sugar-binding proteins into the bindosome, a structure that is expected to be localized close to the cytoplasmic membrane or integrated within the S-layer [118]. The main evidence in support of the presence of this hypothesized structure is that the proposed structural components, the substrate-binding proteins (SBPs), contain class III signal peptide sequences, a feature typical of proteins

which are well known to form oligomeric structures in both archaea and bacteria. The oligomerization of sugar-binding proteins was studied after isolation of the sugar-binding proteins from the membrane of *S. solfataricus* on size exclusion chromatography (Zolghadr et al., unpublished). Previous studies demonstrated that the precursors of the sugar-binding proteins are processed by PibD, the archaeal type IV signal peptidase [50, 97]. The sugar binding-protein oligomer is proposed to play a role in facilitating sugar uptake, a function that enables *S. solfataricus* to grow on a broad variety of substrates.

The Bas system is unique and it has only been identified in *S. solfataricus*. The *bas* operon contains five genes that are organized into 2 smaller operons: the *basEF* genes encoding the main components of the assembly system which are homologues of FlaI/FlaJ of archaeal flagellum assembly system and UpsE/F from the Ups system of Sulfolobus [94]. A second set of genes encompasses *basABC* that encodes small pili-like proteins with class III signal peptides. BasABC is unique and has only been identified in *S. solfataricus*. Previous studies showed that they are constitutively expressed but the electron microscopic investigations did not reveal any pili structure assembled by BasABC. The uptake of glucose was strongly inhibited in a *basEF* deletion mutant and, concomitantly, growth on glucose was strongly impaired. However, the deletion of *basABC* only moderately affected the growth rate and sugar uptake. These results suggested that the Bas system is a novel assembly system involved in correct localization of sugar-binding proteins to the cell envelope, which have a pilin signal peptide. BasEF forms the core of the assembly machinery in the membrane while the BasABC assists the assembly of the binding proteins by an as yet unresolved mechanism.

6. Extracellular Polysaccharides

Bacteria secrete glycosylated proteins and exopolymer substances (EPSs) into the medium for the synthesis of extracellular structures and biofilm. EPS formation, not to be confused with protein glycosylation, is the assembly of long sugar polymers from diverse monosaccharides such as glucose, mannose, and fructose. The EPS is in most cases produced as a capsule surrounding the cell and thereby increasing the adhesion to surfaces or strengthening cell-cell contacts in cell aggregates which leads to biofilm formation [119–121]. Other roles of EPS within biofilms are mainly to provide stability for the structures of the biofilm and protection against different contaminants in media like heavy metals and toxic organic compounds. EPS production is in general increased when cells are exposed to contaminants. EPS and biofilm formation by archaea is a new research area. Using fluorescently conjugated lectins, it was demonstrated that surface attached *S. solfataricus* cells produced EPS containing a variety of different sugars (glucose, mannose, galactose, and N-acetylglucosamine) [86]. Interestingly, the extracellular network produced by PBL2025, a deletion strain appeared different to the wild-type strain *S. solfataricus* P2 strain. PBL2025 lacking a set of 50 genes, which are by BLAST-search analysis predicted to be involved in sugar metabolism/catabolism and transport of solutes across the cytoplasmic membrane. The disruption of these genes has led to the overproduction of EPS and an analysis of the expression pattern of these genes in P2 demonstrated that they are upregulated during surface attachment of the cells on mica [86], identifying the first genes involved in modulation of secreted polysaccharides. Most of the secreted archaeal proteins are glycosylated, a process that is described in detail by Eichler and Jarrell in this issue.

7. Conclusions and Outlook

Electron microscopic investigations of cultured and uncultivable archaea have revealed a remarkable variety of cell surface associated appendages. In recent years, the development of genetic systems for a number of model archaea now allows for experimental investigations on the assembly and function of these structures in at least some organisms. These studies now rapidly increase our understanding on how the archaeal cell surface is assembled. Various cell surface structures such as pili and flagella have been identified and their roles in cell-to-cell and cell-surface interactions start to be uncovered. Interestingly, also secreted vesicles have been identified in different archaeal species that contain a specific subset of proteins implied in an eukaryotic-like vesicle budding systems. This exemplifies the mosaic nature of archaea, which in many cases employ simplified eukaryotic-like mechanisms implying a similar evolutionary origin.

Acknowledgments

This work was supported by the Netherlands Proteomics Centre (NPC), and a VIDI grant from the Dutch Science Organization (NWO) to S. V. Albers who also received intramural funds from the Max Planck Society. A. F. Ellen and B. Zolghadr contributed equally to this work.

References

[1] H. Koenig, "Archaeobacterial cell envelopes," *Canadian Journal of Microbiology*, vol. 34, pp. 395–406, 1988.

[2] R. Rachel, I. Wyschkony, S. Riehl, and H. Huber, "The ultrastructure of Ignicoccus: evidence for a novel outer membrane and for intracellular vesicle budding in an archaeon," *Archaea*, vol. 1, no. 1, pp. 9–18, 2002.

[3] M. Kates, "Structural analysis of phospholipids and glycolipids in extremely halophilic archaebacteria," *Journal of Microbiological Methods*, vol. 25, no. 2, pp. 113–128, 1996.

[4] M. De Rosa, A. Gambacorta, B. Nicolaus, B. Chappe, and P. Albrecht, "Isoprenoid ethers; backbone of complex lipids of the archaebacterium *Sulfolobus solfataricus*," *Biochimica et Biophysica Acta (BBA)*, vol. 753, no. 2, pp. 249–256, 1983.

[5] M. De Rosa, A. Trincone, B. Nicolaus, A. Gambacorta, and G. di Prisco, "Archaebacteria: lipids, membrane structures, and adaptations to environmental stresses," in *Life under Extreme Conditions*, pp. 61–87, Springer, Berlin, Germany, 1991.

[6] J. L. C. M. Van de Vossenberg, T. Ubbink-Kok, M. G. L. Elferink, A. J. M. Driessen, and W. N. Konings, "Ion permeability of the cytoplasmic membrane limits the maximum

growth temperature of bacteria and archaea," *Molecular Microbiology*, vol. 18, no. 5, pp. 925–932, 1995.

[7] O. Kandler and H. Koenig, "Chemical composition of the peptidoglycan-free cell walls of methanogenic bacteria," *Archives of Microbiology*, vol. 118, no. 2, pp. 141–152, 1978.

[8] A. J. M. Driessen and N. Nouwen, "Protein translocation across the bacterial cytoplasmic membrane," *Annual Review of Biochemistry*, vol. 77, pp. 643–667, 2008.

[9] A. M. Flower, L. L. Hines, and P. L. Pfennig, "SecG is an auxiliary component of the protein export apparatus of *Escherichia coli*," *Molecular and General Genetics*, vol. 263, no. 1, pp. 131–136, 2000.

[10] L. N. Kinch, M. H. Saier Jr., and N. V. Grishin, "Sec61β—a component of the archaeal protein secretory system," *Trends in Biochemical Sciences*, vol. 27, no. 4, pp. 170–171, 2002.

[11] S.-V. Albers, Z. Szabó, and A. J. M. Driessen, "Protein secretion in the Archaea: multiple paths towards a unique cell surface," *Nature Reviews Microbiology*, vol. 4, no. 7, pp. 537–547, 2006.

[12] B. Van den Berg, W. M. Clemons Jr., I. Collinson et al., "X-ray structure of a protein-conducting channel," *Nature*, vol. 427, no. 6969, pp. 36–44, 2004.

[13] G. Ring and J. Eichler, "Extreme secretion: protein translocation across the archael plasma membrane," *Journal of Bioenergetics and Biomembranes*, vol. 36, no. 1, pp. 35–45, 2004.

[14] V. Irihimovitch and J. Eichler, "Post-translational secretion of fusion proteins in the halophilic archaea *Haloferax volcanii*," *The Journal of Biological Chemistry*, vol. 278, no. 15, pp. 12881–12887, 2003.

[15] N. J. Hand, R. Klein, A. Laskewitz, and M. Pohlschröder, "Archaeal and bacterial SecD and SecF homologs exhibit striking structural and functional conservation," *Journal of Bacteriology*, vol. 188, no. 4, pp. 1251–1259, 2006.

[16] D. J. F. du Plessis, G. Berrelkamp, N. Nouwen, and A. J. M. Driessen, "The lateral gate of SecYEG opens during protein translocation," *The Journal of Biological Chemistry*, vol. 284, no. 23, pp. 15805–15814, 2009.

[17] C. Robinson and A. Bolhuis, "Tat-dependent protein targeting in prokaryotes and chloroplasts," *Biochimica et Biophysica Acta*, vol. 1694, no. 1–3, pp. 135–147, 2004.

[18] K. Dilks, R. W. Rose, E. Hartmann, and M. Pohlschröder, "Prokaryotic utilization of the twin-arginine translocation pathway: a genomic survey," *Journal of Bacteriology*, vol. 185, no. 4, pp. 1478–1483, 2003.

[19] K. Dilks, M. I. Giménez, and M. Pohlschröder, "Genetic and biochemical analysis of the twin-arginine translocation pathway in halophilic archaea," *Journal of Bacteriology*, vol. 187, no. 23, pp. 8104–8113, 2005.

[20] D. C. Kwan, J. R. Thomas, and A. Bolhuis, "Bioenergetic requirements of a Tat-dependent substrate in the halophilic archaeon *Haloarcula hispanica*," *FEBS Journal*, vol. 275, no. 24, pp. 6159–6167, 2008.

[21] D. A. Widdick, K. Dilks, G. Chandra et al., "The twin-arginine translocation pathway is a major route of protein export in *Streptomyces coelicolor*," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 103, no. 47, pp. 17927–17932, 2006.

[22] G. von Heijne, "The signal peptide," *Journal of Membrane Biology*, vol. 115, no. 3, pp. 195–201, 1990.

[23] D. Tullman-Ercek, M. P. DeLisa, Y. Kawarasaki et al., "Export pathway selectivity of *Escherichia coli* twin arginine translocation signal peptides," *The Journal of Biological Chemistry*, vol. 282, no. 11, pp. 8309–8316, 2007.

[24] J. W. Izard and D. A. Kendall, "Signal peptides: exquisitely designed transport promoters," *Molecular Microbiology*, vol. 13, no. 5, pp. 765–773, 1994.

[25] M. P. DeLisa, D. Tullman, and G. Georgiou, "Folding quality control in the export of proteins by the bacterial twin-arginine translocation pathway," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 100, no. 10, pp. 6115–6120, 2003.

[26] S. Richter, U. Lindenstrauss, C. Lücke, R. Bayliss, and T. Brüser, "Functional tat transport of unstructured, small, hydrophilic proteins," *The Journal of Biological Chemistry*, vol. 282, no. 46, pp. 33257–33264, 2007.

[27] I. B. Holland, L. Schmitt, and J. Young, "Type 1 protein secretion in bacteria, the ABC-transporter dependent pathway," *Molecular Membrane Biology*, vol. 22, no. 1-2, pp. 29–39, 2005.

[28] R. Cannio, A. D'Angelo, M. Rossi, and S. Bartolucci, "A superoxide dismutase from the archaeon *Sulfolobus solfataricus* is an extracellular enzyme and prevents the deactivation by superoxide of cell- bound proteins," *European Journal of Biochemistry*, vol. 267, no. 1, pp. 235–243, 2000.

[29] A. F. Ellen, S.-V. Albers, and A. J. M. Driessen, "Comparative study of the extracellular proteome of *Sulfolobus* species reveals limited secretion," *Extremophiles*, vol. 14, no. 1, pp. 87–98, 2010.

[30] G. Palmieri, R. Cannio, I. Fiume, M. Rossi, and G. Pocsfalvi, "Outside the unusual cell wall of the hyperthermophilic archaeon *Aeropyrum pernix* K1," *Molecular and Cellular Proteomics*, vol. 8, no. 11, pp. 2570–2581, 2009.

[31] P. K. Chong and P. C. Wright, "Identification and characterization of the *Sulfolobus solfataricus* P2 proteome," *Journal of Proteome Research*, vol. 4, no. 5, pp. 1789–1798, 2005.

[32] A. Filloux, "The underlying mechanisms of type II protein secretion," *Biochimica et Biophysica Acta*, vol. 1694, no. 1–3, pp. 163–179, 2004.

[33] R. Voulhoux, G. Ball, B. Ize et al., "Involvement of the twin-arginine translocation system in protein secretion via the type II pathway," *EMBO Journal*, vol. 20, no. 23, pp. 6735–6741, 2001.

[34] I. Chen, P. J. Christie, and D. Dubnau, "The ins and outs of DNA transfer in bacteria," *Science*, vol. 310, no. 5753, pp. 1456–1460, 2005.

[35] R. Fronzes, E. Schäfer, L. Wang, H. R. Saibil, E. V. Orlova, and G. Waksman, "Structure of a type IV secretion system core complex," *Science*, vol. 323, no. 5911, pp. 266–268, 2009.

[36] Q. She, H. Phan, R. A. Garrett, S.-V. Albers, K. M. Stedman, and W. Zillig, "Genetic profile of pNOB8 from *Sulfolobus*: the first conjugative plasmid from an archaeon," *Extremophiles*, vol. 2, no. 4, pp. 417–425, 1998.

[37] G. Erauso, K. M. Stedman, H. J. G. van den Werken, W. Zillig, and J. van der Oost, "Two novel conjugative plasmids from a single strain of *Sulfolobus*," *Microbiology*, vol. 152, no. 7, pp. 1951–1968, 2006.

[38] D. Prangishvili, S.-V. Albers, I. Holz et al., "Conjugation in archaea: frequent occurrence of conjugative plasmids in *Sulfolobus*," *Plasmid*, vol. 40, no. 3, pp. 190–202, 1998.

[39] K. M. Stedman, Q. She, H. Phan et al., "pING family of conjugative plasmids from the extremely thermophilic archaeon *Sulfolobus islandicus*: insights into recombination and conjugation in Crenarchaeota," *Journal of Bacteriology*, vol. 182, no. 24, pp. 7014–7020, 2000.

[40] I. Rosenshine, R. Tchelet, and M. Mevarech, "The mechanism of DNA transfer in the mating system of an archaebacterium," *Science*, vol. 245, no. 4924, pp. 1387–1389, 1989.

[41] L. Craig and J. Li, "Type IV pili: paradoxes in form and function," *Current Opinion in Structural Biology*, vol. 18, no. 2, pp. 267–277, 2008.

[42] M. S. Strom, D. N. Nunn, and S. Lory, "A single bifunctional enzyme, PilD, catalyzes cleavage and N-methylation of proteins belonging to the type IV pilin family," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 90, no. 6, pp. 2404–2408, 1993.

[43] S. Y. M. Ng, B. Zolghadr, A. J. M. Driessen, S.-V. Albers, and K. F. Jarrell, "Cell surface structures of archaea," *Journal of Bacteriology*, vol. 190, no. 18, pp. 6039–6047, 2008.

[44] P. J. Planet, S. C. Kachlany, R. DeSalle, and D. H. Figurski, "Phylogeny of genes for secretion NTPases: identification of the widespread tadA subfamily and development of a diagnostic key for gene classification," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 98, no. 5, pp. 2503–2508, 2001.

[45] A. Yamagata and J. A. Tainer, "Hexameric structures of the archaeal secretion ATPase GspE and implications for a universal secretion mechanism," *EMBO Journal*, vol. 26, no. 3, pp. 878–890, 2007.

[46] S.-V. Albers and A. J. M. Driessen, "Signal peptides of secreted proteins of the archaeon *Sulfolobus solfataricus*: a genomic survey," *Archives of Microbiology*, vol. 177, no. 3, pp. 209–216, 2002.

[47] N. F. W. Saunders, C. Ng, M. Raftery, M. Guilhaus, A. Goodchild, and R. Cavicchioli, "Proteomic and computational analysis of secreted proteins with type I signal peptides from the antarctic archaeon *Methanococcoides burtonii*," *Journal of Proteome Research*, vol. 5, no. 9, pp. 2457–2464, 2006.

[48] J. Eichler and M. W. W. Adams, "Posttranslational protein modification in Archaea," *Microbiology and Molecular Biology Reviews*, vol. 69, no. 3, pp. 393–425, 2005.

[49] M. I. Giménez, K. Dilks, and M. Pohlschröder, "*Haloferax volcanii* twin-arginine translocation substates include secreted soluble, C-terminally anchored and lipoproteins," *Molecular Microbiology*, vol. 66, no. 6, pp. 1597–1606, 2007.

[50] S.-V. Albers, Z. Szabó, and A. J. M. Driessen, "Archaeal homolog of bacterial type IV prepilin signal peptidases with broad substrate specificity," *Journal of Bacteriology*, vol. 185, no. 13, pp. 3918–3925, 2003.

[51] S. L. Bardy and K. F. Jarrell, "FlaK of the archaeon *Methanococcus maripaludis* possesses preflagellin peptidase activity," *FEMS Microbiology Letters*, vol. 208, no. 1, pp. 53–59, 2002.

[52] Z. Szabó, A. O. Stahl, S.-V. Albers, J. C. Kissinger, A. J. M. Driessen, and M. Pohlschröder, "Identification of diverse archaeal proteins with class III signal peptides cleaved by distinct archaeal prepilin peptidases," *Journal of Bacteriology*, vol. 189, no. 3, pp. 772–778, 2007.

[53] S.-O. Shan and P. Walter, "Co-translational protein targeting by the signal recognition particle," *FEBS Letters*, vol. 579, no. 4, pp. 921–926, 2005.

[54] P. F. Egea, S.-O. Shan, J. Napetschnig, D. F. Savage, P. Walter, and R. M. Stroud, "Substrate twinning activates the signal recognition particle and its receptor," *Nature*, vol. 427, no. 6971, pp. 215–221, 2004.

[55] A. Haddad, R. W. Rose, and M. Pohlschröder, "The *Haloferax volcanii* FtsY homolog is critical for haloarchaeal growth but does not require the A domain," *Journal of Bacteriology*, vol. 187, no. 12, pp. 4015–4022, 2005.

[56] H. Tjalsma, A. Bolhuis, J. D. H. Jongbloed, S. Bron, and J. M. Van Dijl, "Signal peptide-dependent protein transport in *Bacillus subtilis*: a genome-based survey of the secretome," *Microbiology and Molecular Biology Reviews*, vol. 64, no. 3, pp. 515–547, 2000.

[57] J. L. Gardy, M. R. Laird, F. Chen et al., "PSORTb v.2.0: expanded prediction of bacterial protein subcellular localization and insights gained from comparative proteome analysis," *Bioinformatics*, vol. 21, no. 5, pp. 617–623, 2005.

[58] J. D. Bendtsen, H. Nielsen, G. von Heijne, and S. Brunak, "Improved prediction of signal peptides: signalP 3.0," *Journal of Molecular Biology*, vol. 340, no. 4, pp. 783–795, 2004.

[59] S. L. Bardy, J. Eichler, and K. F. Jarrell, "Archaeal signal peptides—a comparative survey at the genome level," *Protein Science*, vol. 12, no. 9, pp. 1833–1843, 2003.

[60] M. Abu-Qarn and J. Eichler, "An analysis of amino acid sequences surrounding archaeal glycoprotein sequons," *Archaea*, vol. 2, no. 2, pp. 73–81, 2007.

[61] M. Saleh, C. Song, S. Nasserulla, and L. G. Leduc, "Indicators from archaeal secretomes," *Microbiological Research*, vol. 165, no. 1, pp. 1–10, 2010.

[62] P. G. Bagos, K. D. Tsirigos, S. K. Plessas, T. D. Liakopoulos, and S. J. Hamodrakas, "Prediction of signal peptides in archaea," *Protein Engineering, Design and Selection*, vol. 22, no. 1, pp. 27–35, 2009.

[63] T. J. Williams, D. W. Burg, M. J. Raftery et al., "Global proteomic analysis of the insoluble, soluble, and supernatant fractions of the psychrophilic archaeon *Methanococcoides burtonii* part I: the effect of growth temperature," *Journal of Proteome Research*, vol. 9, no. 2, pp. 640–652, 2010.

[64] D. R. Francoleon, P. Boontheung, Y. Yang et al., "S-layer, surface-accessible, and concanavalin a binding proteins of *Methanosarcina acetivorans* and *Methanosarcina mazei*," *Journal of Proteome Research*, vol. 8, no. 4, pp. 1972–1982, 2009.

[65] L. Mashburn-Warren, R. J. C. Mclean, and M. Whiteley, "Gram-negative outer membrane vesicles: beyond the cell surface," *Geobiology*, vol. 6, no. 3, pp. 214–219, 2008.

[66] N. Soler, E. Marguet, J.-M. Verbavatz, and P. Forterre, "Virus-like vesicles and extracellular DNA produced by hyperthermophilic archaea of the order *Thermococcales*," *Research in Microbiology*, vol. 159, no. 5, pp. 390–399, 2008.

[67] A.-L. Reysenbach, Y. Liu, A. B. Banta et al., "A ubiquitous thermoacidophilic archaeon from deep-sea hydrothermal vents," *Nature*, vol. 442, no. 7101, pp. 444–447, 2006.

[68] A. F. Ellen, S.-V. Albers, W. Huibers et al., "Proteomic analysis of secreted membrane vesicles of archaeal *Sulfolobus* species reveals the presence of endosome sorting complex components," *Extremophiles*, vol. 13, no. 1, pp. 67–79, 2009.

[69] R. Grimm, H. Singh, R. Rachel, D. Typke, W. Zillig, and W. Baumeister, "Electron tomography of ice-embedded prokaryotic cells," *Biophysical Journal*, vol. 74, no. 2, pp. 1031–1042, 1998.

[70] D. Prangishvili, I. Holz, E. Stieger, S. Nickell, J. K. Kristjansson, and W. Zillig, "Sulfolobicins, specific proteinaceous toxins produced by strains of the extremely thermophilic archaeal genus *Sulfolobus*," *Journal of Bacteriology*, vol. 182, no. 10, pp. 2985–2988, 2000.

[71] R. Y. Samson, T. Obita, S. M. Freund, R. L. Williams, and S. D. Bell, "A role for the ESCRT system in cell division in archaea," *Science*, vol. 322, no. 5908, pp. 1710–1713, 2008.

[72] F. Mayer and G. Gottschalk, "The bacterial cytoskeleton and its putative role in membrane vesicle formation observed in a gram-positive bacterium producing starch-degrading enzymes," *Journal of Molecular Microbiology and Biotechnology*, vol. 6, no. 3-4, pp. 127–132, 2003.

[73] M. L. Rodrigues, L. Nimrichter, D. L. Oliveira et al., "Vesicular polysaccharide export in *Cryptococcus neoformans* is a eukaryotic solution to the problem of fungal trans-cell wall transport," *Eukaryotic Cell*, vol. 6, no. 1, pp. 48–59, 2007.

[74] B. L. Deatherage, J. C. Lara, T. Bergsbaken, S. L. R. Barrett, S. Lara, and B. T. Cookson, "Biogenesis of bacterial membrane vesicles," *Molecular Microbiology*, vol. 72, no. 6, pp. 1395–1407, 2009.

[75] E.-Y. Lee, D.-S. Choi, K.-P. Kim, and Y. S. Gho, "Proteomics in Gram-negative bacterial outer membrane vesicles," *Mass Spectrometry Reviews*, vol. 27, no. 6, pp. 535–555, 2008.

[76] C. Balsalobre, J. M. Silván, S. Berglund, Y. Mizunoe, B. E. Uhlin, and S. N. Wai, "Release of the type I secreted α-haemolysin via outer membrane vesicles from *Escherichia coli*," *Molecular Microbiology*, vol. 59, no. 1, pp. 99–112, 2006.

[77] L. M. Mashburn and M. Whiteley, "Membrane vesicles traffic signals and facilitate group activities in a prokaryote," *Nature*, vol. 437, no. 7057, pp. 422–425, 2005.

[78] A. J. McBroom and M. J. Kuehn, "Release of outer membrane vesicles by Gram-negative bacteria is a novel envelope stress response," *Molecular Microbiology*, vol. 63, no. 2, pp. 545–558, 2007.

[79] N. A. Thomas, S. Mueller, A. Klein, and K. F. Jarrell, "Mutants in flaI and flaJ of the archaeon *Methanococcus voltae* are deficient in flagellum assembly," *Molecular Microbiology*, vol. 46, no. 3, pp. 879–887, 2002.

[80] Z. Szabó, M. Sani, M. Groeneveld et al., "Flagellar motility and structure in the hyperthermoacidophilic archaeon *Sulfolobus solfataricus*," *Journal of Bacteriology*, vol. 189, no. 11, pp. 4305–4309, 2007.

[81] S. L. Bardy, T. Mori, K. Komoriya, S.-I. Aizawa, and K. F. Jarrell, "Identification and localization of flagellins FlaA and FlaB3 within flagella of *Methanococcus voltae*," *Journal of Bacteriology*, vol. 184, no. 19, pp. 5223–5233, 2002.

[82] T. Nutsch, W. Marwan, D. Oesterhelt, and E. D. Gilles, "Signal processing and flagellar motor switching during phototaxis of *Halobacterium salinarum*," *Genome Research*, vol. 13, no. 11, pp. 2406–2412, 2003.

[83] M. Alam and D. Oesterhelt, "Morphology, function and isolation of halobacterial flagella," *Journal of Molecular Biology*, vol. 176, no. 4, pp. 459–475, 1984.

[84] D. J. Näther, R. Rachel, G. Wanner, and R. Wirth, "Flagella of *Pyrococcus furiosus*: multifunctional organelles, made for swimming, adhesion to various surfaces, and cell-cell contacts," *Journal of Bacteriology*, vol. 188, no. 19, pp. 6915–6923, 2006.

[85] S. Schopf, G. Wanner, R. Rachel, and R. Wirth, "An archaeal bi-species biofilm formed by *Pyrococcus furiosus* and *Methanopyrus kandleri*," *Archives of Microbiology*, vol. 190, no. 3, pp. 371–377, 2008.

[86] B. Zolghadr, A. Kling, A. Koerdt, A. J. M. Driessen, R. Rachel, and S.-V. Albers, "Appendage-mediated surface adherence of *Sulfolobus solfataricus*," *Journal of Bacteriology*, vol. 192, no. 1, pp. 104–110, 2010.

[87] S. Y. M. Ng, B. Chaban, and K. F. Jarrell, "Archaeal flagella, bacterial flagella and type IV pili: a comparison of genes and posttranslational modifications," *Journal of Molecular Microbiology and Biotechnology*, vol. 11, no. 3–5, pp. 167–191, 2006.

[88] S.-V. Albers and M. Pohlschröder, "Diversity of archaeal type IV pilin-like structures," *Extremophiles*, vol. 13, no. 3, pp. 403–410, 2009.

[89] S. Trachtenberg and S. Cohen-Krausz, "The archaeabacterial flagellar filament: a bacterial propeller with a pilus-like structure," *Journal of Molecular Microbiology and Biotechnology*, vol. 11, no. 3–5, pp. 208–220, 2006.

[90] M. G. Pyatibratov, S. N. Beznosov, R. Rachel et al., "Alternative flagellar filament types in the haloarchaeon *Haloarcula marismortui*," *Canadian Journal of Microbiology*, vol. 54, no. 10, pp. 835–844, 2008.

[91] L. Craig, M. E. Pique, and J. A. Tainer, "Type IV pilus structure and bacterial pathogenicity," *Nature Reviews Microbiology*, vol. 2, no. 5, pp. 363–378, 2004.

[92] S. Streif, W. F. Staudinger, W. Marwan, and D. Oesterhelt, "Flagellar rotation in the archaeon *Halobacterium salinarum* depends on ATP," *Journal of Molecular Biology*, vol. 384, no. 1, pp. 1–8, 2008.

[93] N. A. Thomas, S. L. Bardy, and K. F. Jarrell, "The archaeal flagellum: a different kind of prokaryotic motility structure," *FEMS Microbiology Reviews*, vol. 25, no. 2, pp. 147–174, 2001.

[94] S.-V. Albers and A. J. M. Driessen, "Analysis of ATPases of putative secretion operons in the thermoacidophilic archaeon *Sulfolobus solfataricus*," *Microbiology*, vol. 151, no. 3, pp. 763–773, 2005.

[95] N. Patenge, A. Berendes, H. Engelhardt, S. C. Schuster, and D. Oesterhelt, "The fla gene cluster is involved in the biogenesis of flagella in *Halobacterium salinarum*," *Molecular Microbiology*, vol. 41, no. 3, pp. 653–663, 2001.

[96] N. A. Thomas and K. F. Jarrell, "Characterization of flagellum gene families of methanogenic archaea and localization of novel flagellum accessory proteins," *Journal of Bacteriology*, vol. 183, no. 24, pp. 7154–7164, 2001.

[97] Z. Szabó, S.-V. Albers, and A. J. M. Driessen, "Active-site residues in the type IV prepilin peptidase homologue PibD from the archaeon *Sulfolobus solfataricus*," *Journal of Bacteriology*, vol. 188, no. 4, pp. 1437–1443, 2006.

[98] S. Cohen-Krausz and S. Trachtenberg, "The structure of the archeabacterial flagellar filament of the extreme halophile *Halobacterium salinarum* R1M1 and its relation to eubacterial flagellar filaments and type IV pili," *Journal of Molecular Biology*, vol. 321, no. 3, pp. 383–395, 2002.

[99] S. Cohen-Krausz and S. Trachtenberg, "The flagellar filament structure of the extreme acidothermophile *Sulfolobus shibatae* B12 suggests that archaeabacterial flagella have a unique and common symmetry and design," *Journal of Molecular Biology*, vol. 375, no. 4, pp. 1113–1124, 2008.

[100] M. L. Kalmokoff and K. F. Jarrell, "Cloning and sequencing of a multigene family encoding the flagellins of *Methanococcus voltae*," *Journal of Bacteriology*, vol. 173, no. 22, pp. 7113–7125, 1991.

[101] B. Chaban, S. Y. M. Ng, M. Kanbe et al., "Systematic deletion analyses of the fla genes in the flagella operon identify several genes essential for proper assembly and function of flagella in the archaeon, *Methanococcus maripaludis*," *Molecular Microbiology*, vol. 66, no. 3, pp. 596–609, 2007.

[102] S. N. Beznosov, M. G. Pyatibratov, and O. V. Fedorov, "On the multicomponent nature of *Halobacterium salinarum* flagella," *Microbiology*, vol. 76, no. 4, pp. 435–441, 2007.

[103] L. Gerl and M. Sumper, "Halobacterial flagellins are encoded by a multigene family. Characterization of five flagellin genes," *The Journal of Biological Chemistry*, vol. 263, no. 26, pp. 13246–13251, 1988.

[104] V. Y. Tarasov, M. G. Pyatibratov, S.-L. Tang, M. Dyall-Smith, and O. V. Fedorov, "Role of flagellins from A and B loci

in flagella formation of *Halobacterium salinarum*," *Molecular Microbiology*, vol. 35, no. 1, pp. 69–78, 2000.

[105] D. M. Faguy, D. P. Bayley, A. S. Kostyukova, N. A. Thomas, and K. F. Jarrell, "Isolation and characterization of flagella and flagellin proteins from the thermoacidophilic archaea *Thermoplasma volcanium* and *Sulfolobus shibatae*," *Journal of Bacteriology*, vol. 178, no. 3, pp. 902–905, 1996.

[106] S.-V. Albers, M. Jonuscheit, S. Dinkelaker et al., "Production of recombinant and tagged proteins in the hyperthermophilic archaeon *Sulfolobus solfataricus*," *Applied and Environmental Microbiology*, vol. 72, no. 1, pp. 102–111, 2006.

[107] J. Abendroth, P. Murphy, M. Sandkvist, M. Bagdasarian, and W. G. J. Hol, "The X-ray structure of the type II secretion system complex formed by the N-terminal domain of EpsE and the cytoplasmic domain of EpsL of *Vibrio cholerae*," *Journal of Molecular Biology*, vol. 348, no. 4, pp. 845–855, 2005.

[108] M. Schlesner, A. Miller, S. Streif et al., "Identification of Archaea-specific chemotaxis proteins which interact with the flagellar apparatus," *BMC Microbiology*, vol. 9, article 56, 2009.

[109] S. Nickell, R. Hegerl, W. Baumeister, and R. Rachel, "*Pyrodictium cannulae* enter the periplasmic space but do not enter the cytoplasm, as revealed by cryo-electron tomography," *Journal of Structural Biology*, vol. 141, no. 1, pp. 34–42, 2003.

[110] G. Rieger, R. Rachel, R. Hermann, and K. O. Stetter, "Ultrastructure of the hyperthermophilic archaeon *Pyrodictium-Abyssi*," *Journal of Structural Biology*, vol. 115, no. 1, pp. 78–87, 1995.

[111] C. Horn, B. Paulmann, G. Kerlen, N. Junker, and H. Huber, "In vivo observation of cell division of anaerobic hyperthermophiles by using a high-intensity dark-field microscope," *Journal of Bacteriology*, vol. 181, no. 16, pp. 5114–5118, 1999.

[112] K.O. Stetter, H. Konig, and E. Stackebrandt, "Pyrodictium gen-nov, a new genus of submarine disc-shaped sulfur reducing archaebacteria growing optimally at 105-degrees-C," *Systematic and Applied Microbiology*, vol. 4, pp. 535–551, 1983.

[113] C. Moissl, R. Rachel, A. Briegel, H. Engelhardt, and R. Huber, "The unique structure of archaeal 'hami', highly complex cell appendages with nano-grappling hooks," *Molecular Microbiology*, vol. 56, no. 2, pp. 361–370, 2005.

[114] C. Rudolph, G. Wanner, and R. Huber, "Natural communities of novel archaea and bacteria growing in cold sulfurous springs with a string-of-pearls-like morphology," *Applied and Environmental Microbiology*, vol. 67, no. 5, pp. 2336–2344, 2001.

[115] D. W. Müller, C. Meyer, S. Gürster et al., "The Iho670 fibers of *Ignicoccus hospitalis*: a new type of archaeal cell surface appendage," *Journal of Bacteriology*, vol. 191, no. 20, pp. 6465–6468, 2009.

[116] S. Fröls, M. Ajon, M. Wagner et al., "UV-inducible cellular aggregation of the hyperthermophilic archaeon *Sulfolobus solfataricus* is mediated by pili formation," *Molecular Microbiology*, vol. 70, no. 4, pp. 938–952, 2008.

[117] Y. A. Wang, X. Yu, S. Y. M. Ng, K. F. Jarrell, and E. H. Egelman, "The structure of an archaeal pilus," *Journal of Molecular Biology*, vol. 381, no. 2, pp. 456–466, 2008.

[118] B. Zolghadr, S. Weber, Z. Szabó, A. J. M. Driessen, and S.-V. Albers, "Identification of a system required for the functional surface localization of sugar binding proteins with class III signal peptides in *Sulfolobus solfataricus*," *Molecular Microbiology*, vol. 64, no. 3, pp. 795–806, 2007.

[119] D. F. Rodrigues and M. Elimelech, "Role of type 1 fimbriae and mannose in the development of *Escherichia coli* K12 biofilm: from initial cell adhesion to biofilm formation," *Biofouling*, vol. 25, no. 5, pp. 401–411, 2009.

[120] H. Laue, A. Schenk, H. Li et al., "Contribution of alginate and levan production to biofilm formation by *Pseudomonas syringae*," *Microbiology*, vol. 152, no. 10, pp. 2909–2918, 2006.

[121] S. Tsuneda, H. Aikawa, H. Hayashi, A. Yuasa, and A. Hirata, "Extracellular polymeric substances responsible for bacterial adhesion onto solid surface," *FEMS Microbiology Letters*, vol. 223, no. 2, pp. 287–292, 2003.

Hindawi Publishing Corporation
Archaea
Volume 2010, Article ID 612948, 13 pages
doi:10.1155/2010/612948

Review Article

S-Layer Glycoproteins and Flagellins: Reporters of Archaeal Posttranslational Modifications

Ken F. Jarrell,[1] Gareth M. Jones,[1] Lina Kandiba,[2] Divya B. Nair,[1] and Jerry Eichler[2]

[1] *Department of Microbiology and Immunology, Queen's University, Kingston, ON, Canada K7L 3N6*
[2] *Department of Life Sciences, Ben Gurion University, Beersheva 84105, Israel*

Correspondence should be addressed to Ken F. Jarrell, jarrellk@queensu.ca

Received 8 April 2010; Accepted 15 June 2010

Academic Editor: Joerg Soppa

Many archaeal proteins undergo posttranslational modifications. S-layer proteins and flagellins have been used successfully to study a variety of these modifications, including N-linked glycosylation, signal peptide removal and lipid modification. Use of these well-characterized reporter proteins in the genetically tractable model organisms, *Haloferax volcanii, Methanococcus voltae* and *Methanococcus maripaludis,* has allowed dissection of the pathways and characterization of many of the enzymes responsible for these modifications. Such studies have identified archaeal-specific variations in signal peptidase activity not found in the other domains of life, as well as the enzymes responsible for assembly and biosynthesis of novel N-linked glycans. In vitro assays for some of these enzymes have already been developed. N-linked glycosylation is not essential for either *Hfx. volcanii* or the *Methanococcus* species, an observation that allowed researchers to analyze the role played by glycosylation in the function of both S-layers and flagellins, by generating mutants possessing these reporters with only partial attached glycans or lacking glycan altogether. In future studies, it will be possible to consider questions related to the heterogeneity associated with given modifications, such as differential or modulated glycosylation.

1. Introduction

Carl Woese initially defined the third form of life, the Archaea, on the basis of the novel oligonucleotide signatures of their small ribosomal subunit RNA [1–3]. Specifically, by generating phylogenetic trees based on 16S rRNA sequences, Woese clearly showed that Archaea formed a unique group, distinct from Bacteria or Eukarya. However, early analysis also revealed that this unusual group of microbes shared a variety of other characteristics, most notably ether-linked membrane lipids, a variety of unusual cell walls (none of which contained murein), atypical DNA-dependent RNA polymerases and later, their own variation of flagella [4, 5]. Indeed, cell wall composition was one of the very first phenotypical traits of the Archaea considered that allowed for differentiation from Bacteria [6] and was considered in the early days of archaeal research to be "the only useful phylogenetic criterion, other than direct molecular phylogenetic measurement" to distinguish between the two prokaryotic domains [7]. A common feature of many genera of Archaea, found in representatives of all the major lineages, is the presence of an outermost component of the cell envelope termed the surface (S)-layer, comprising protein or often glycoprotein subunits that form a regularly structured array.

In addition to their distinctive cell walls, cell surface structures of Archaea are also unusual [8]. While many, such as cannulae [9] and hami [10] are unique to Archaea, even the more commonly found flagella and pili are unlike their bacterial namesakes [11, 12]. Archaeal flagella are the best-studied of the archaeal appendages and are unusual in many aspects, including the initial biosynthesis of the component flagellins with N-terminal class III signal peptides that are cleaved by a prepilin peptidase-like enzyme. As considered below, these traits are all similar to those found in bacterial type IV pili systems but absent in bacterial flagella systems.

Both S-layer proteins and flagellins are among the most abundant proteins synthesized by the archaeal cell and both can be isolated with relative ease in substantial amounts for biochemical and structural studies. In Archaea,

the majority of these proteins appear to be glycoproteins, mainly containing N-linked glycans, yet sometimes also containing glycans O-linked to threonine residues. In fact, archaeal S-layer proteins, especially those from extreme halophiles, served to identify a novel class of prokaryotic glycoproteins [13–16]. More recently, S-layer proteins and flagellins have been widely used as reporter proteins for the study of a variety of posttranslational modifications in Archaea [17–20], including both class I and class III signal peptide removal, N- and O-glycosylation and lipid modification. Despite these advances, it is only in a very small number of Archaea, including *Haloferax volcanii* and *Methanococcus* species, that genetic studies linking specific genes to a particular posttranslational modification have been performed.

2. The S-layer of Haloarchaea and Methanoarchaea

Although found in numerous archaeal species, the S-layers of haloarchaea remain the best studied. Indeed, the first description of a S-layer in Archaea was reported in 1956 when electron microscopic examination of *Halobacterium halobium* (*salinarum*) cells revealed a surface presenting morphological units organized in a hexagonal pattern [21]. Later examination of thin-sectioned haloarchaeal cells revealed the presence of a 17 nm thick cell wall beyond the plasma membrane [22–24]. Blaurock et al. [25] later relied on X-ray diffraction to demonstrate a protein layer laying beyond the haloarchaeal plasma membrane at a distance of 8 nm, with a periplasmic-like space being formed from morphological subunits assuming an "inverted-parabola shape". Enzymatic iodination of *Hbt. salinarum* surface proteins, together with proteolytic treatment, revealed this surface (S)-layer to contain the S-layer glycoprotein [13]. Kessel et al. [26] next proposed a three-dimensional reconstruction of the *Halobacterium* (later renamed *Haloferax*) *volcanii* S-layer glycoprotein and cell envelope after considering the primary sequence of the *Hbt. salinarum* S-layer glycoprotein [14], as well as the earlier X-ray diffraction data and electron microscopic images of negatively stained cell envelopes. In this model, based on reconstruction to a 2 nm resolution, six S-layer glycoproteins form a 4.5 nm thick dome-shaped pore, with the open center expanding as one approaches the membrane. The deduced C-terminal transmembrane domain of each S-layer glycoprotein is thought to anchor the structure to the membrane, while an O-glycosylated domain of the S-layer glycoprotein lying upstream of the transmembrane domain is proposed to act as a spacer unit, propping up the domed structure. While the forces responsible for maintaining the integrity of such assemblies remain unknown, divalent cations have been shown to be important [13, 26]. Given the use of intact cells maintained in their growth medium and the high degree of sample preservation afforded by the rapid freezing, reconstruction of the *Hbt. salinarum* S-layer through the use of electron tomography offered a more realistic view of this structure [27]. It was thus shown that the *Hbt. salinarum* cell envelope assumes the same basic architecture as does the *Hfx. volcanii* S-layer. Despite similarities in their S-layer architecture, *Hbt. salinarum* and *Hfx. volcanii* cells assume very different shapes, with the former appearing as rods and the latter as indented disks, pointing to factors other than the S-layer as affecting cell shape.

Like their halophilic counterparts, numerous methanoarchaeal species, such as members of the genus *Methanococcus*, are also surrounded by a glycoprotein-based S-layer [28]. For many methanogens, study of the S-layer has been relatively limited to identification of the major S-layer component, its response to glycoprotein staining procedures and determination of the lattice symmetry by electron microscopy [29–31]. In the case of *M. voltae*, the S-layer is formed from a 76 kDa protein arranged into an hexagonal lattice with a center to center spacing of 10 nm [32]. While the S-layer protein does not stain positively with the periodic acid Schiff reagent, suggesting the absence of protein glycosylation, subsequent mass spectrometry analysis has shown it to contain a N-linked glycan identical to that found on flagellins in this species [33]. Methodologies were developed to create protoplasts of *M. voltae* by removal and regeneration of the S-layer [34]. The protoplasts could be used for transformation of plasmid DNA either directly or by electroporation [35].

In some instances, such as *Methanothermus fervidus* [36], the S-layer surrounds a sacculus of pseudomurein, a peptidoglycan unique to certain methanoarchaeal species and distinct from bacterial murein. In some *Methanosarcina* species, the cell envelope is thought to consist of a protein-based S-layer surrounded by a rigid cell wall composed of methanochondroitin, a heteropolysaccharide reminiscent of eukaryotic chondroitin [37].

A model of the S-layer of *Methanolobus limicola*, formed from glycoprotein subunits, was determined using a variety of microscopy techniques. Using standard electron crystallographic techniques, it was determined that the S-layer had p6 symmetry with a lattice constant of 14.7 nm and a thickness of 4.5 nm [38]. Later examination by scanning tunneling microscopy resulted in a thickness determination of 6.5 nm [39]. As in *Hbt. salinarum*, the subunits are thought to assemble into a dome-shaped structure. Moreover, although no spacer elements have been identified linking the S-layer to the cytoplasmic membrane, negative staining does reveal a narrow space of about 5–10 nm between the two layers [38].

Methanospirillum hungatei cells present a very complicated envelope profile. Individual cells are surrounded by a cell wall consisting of an S-layer [40] and then a second unusual outer paracrystalline layer termed the sheath [41], consisting of individual and discrete hoop-like components [42]. The ensheathed cells are then separated from the surrounding environment or from neighboring cells by complicated end or spacer plugs [42].

Finally, in *Methanocorpusculum sinense*, examination of the role of the S-layer in cell-shape maintenance and cell division revealed that lattice faults in the normal p6 symmetry of the S-layer appear to be sites of incorporation of new subunits and initiation points for cell division [43].

3. Archaeal Flagella Assembly and Composition

Biochemical, genetic, and structural studies performed on flagella from several different archaeal species over the last 2 decades have demonstrated the unique nature of this motility apparatus [5, 44]. Flagella have been reported in all of the major subgroupings of cultivatable archaea, including species of extreme halophiles, haloalkaliphiles, methanogens, hyperthermophiles, and thermoacidophiles [45, 46]. Detailed biochemical, structural and/or genetic studies have been reported in a variety of archaeal genera, including *Methanococcus* [47–49], *Methanospirillum* [50–52], *Halobacterium* [53–56], *Haloarcula* [57], *Haloferax* [58], *Sulfolobus* [59], *Natrialba* [60, 61], *Thermococcus* [62], and *Pyrococcus* [63]. However, the bulk of published work on archaeal flagella is focused on *Halobacterium* and *Methanococcus*; that is, members of the Euryarchaeota, and it is not certain that findings in one organism or even within one of the archaeal domains are applicable to all Archaea or even to members of the other major archaeal domain (i.e., crenarchaeotes). There may be fundamental differences between the two domains; for instance, genes known to be essential for flagellation in *Methanococcus* are not found in crenarchaeotes [46, 64] and even something as fundamental as the presence of the hook may be variable, as hooks have not been observed in the crenarchaeote *Sulfolobus* [65].

Archaeal flagella are motility structures involved in swimming and, in the one example where this has been examined in any detail (i.e., *Hbt. salinarum*), the flagella can switch their direction of rotation [66–68]. Other than this superficial commonality, archaeal flagella do not bear other similarities to their bacterial counterparts. For example, there are no homologues of bacterial flagella structural or biosynthetic genes contained in any sequenced archaeal genome [67, 69]. Another fundamental difference between the two prokaryotic flagella organelles may be in the driving force for flagellar rotation. Proton or, more rarely, sodium gradients are used to power bacterial flagellar motion [70], while in the one instance where this has been examined in Archaea, again in *Hbt. salinarum*, flagellar motor rotation depends on ATP [71]. Structurally, archaeal flagella are similar to bacterial type IV pili, surface structures involved in a type of motility across solid surfaces called twitching [72], and, critically, lack a central channel that could allow the passage of subunits through the growing structure for assembly at the distal tip [73–75]. Indeed, the archaeal flagellum has been termed "a bacterial propeller with a pilus-like structure" [75]. Accordingly, archaeal flagella share several commonalities with bacterial type IV pili. Most strikingly, the major subunits of the archaeal flagellum, the flagellins, are made as preproteins with unusual, type IV pilin-like signal peptides (class III signal peptides) that are removed by a specific type IV prepilin signal peptidase homologue (FlaK/PibD; see below) [47, 48, 76]. In addition, both the archaeal flagella system and the type IV pili systems contain a homologous ATPase and a conserved membrane component that may serve as the platform for assembly of the structures [77, 78]. These similarities to type IV pili and the lack of a central channel indicate that assembly of the archaeal flagellum takes place by addition of subunits to the base of the structure, as is also the case in pili growth [45], and fundamentally different from the growth of bacterial flagella, where new subunits are added to the distal end after their passage through the central channel [79].

A single *fla* operon encompassing up to thirteen flagella-associated genes has been identified in various flagellated archaea, although the core composition of genes involved in flagellation and their arrangement in the genome can vary in different organisms [44]. Unlike most bacterial species, nearly all the flagellated archaeal species contain multiple (i.e., 2–6) flagellin genes, a rare exception being *Sulfolobus* spp., where only a single flagellin gene is found. Studies show that each flagellin has its own function as deletion of a single flagellin usually results in nonflagellated cells [64, 80, 81]. Interestingly, one of the flagellins forms the hook region in *Methanococcus* species [49, 64, 82] and *Hbt. salinarum* [82]. The *fla* operon typically begins with the multiple flagellin genes, followed by the conserved *fla*-associated genes, *fla*C-*flaJ*, or a subset thereof. The preflagellin peptidase gene is typically located outside this main locus. Deletion analysis has demonstrated that all of the successfully deleted *fla*-associated genes are essential for flagellation, even though some of these genes are not found in all flagellated archaea [64]. Of the *fla*-associated genes, *flaHIJ* are conserved in all flagellated archaea [5]. FlaI and FlaJ are homologous to ATPases (i.e., PilT/PilB) and conserved membrane protein (i.e., PilC/TadB) of type IV pili systems. FlaI has been shown to possess ATPase activity [83]. FlaH may also have ATPase activity, as it contains a conserved Walker box A, although a Walker box B has not been identified [84]. Thus, due to their universal presence in all flagellated archaea and their relationships to type IV pili-related proteins, FlaHIJ are most likely key components in the export and assembly of flagellin subunits. Deletions in any of these genes lead to nonflagellated cells [56, 59, 64, 84, 85]. The roles for the other *fla* associated genes are generally unknown, although recent reports indicate FlaCE and FlaD of *Hbt. salinarum* associate with various Che proteins of the chemotaxis system [86].

Finally, in addition to signal peptide removal, archaeal flagellins also undergo a second posttranslational modification, as most are glycoproteins. While Bacteria sometimes contain glycosylated flagellins, the flagellin glycan is always found in an O-linkage [87]. In Archaea, to date the glycan has always been found to be N-linked to the flagellins [33, 88, 89]. As considered below, the presence and completeness of the glycan has marked effects on the assembly and function of the archaeal flagella.

In bacteria, flagella do not function only as organelles for swimming but can also be involved in such diverse activities as swarming motility across surfaces, sensing wetness and playing a role in biofilm formation, for example [90–92]. Similarly, the flagella of Archaea have been recently shown to be involved in other important biological functions in addition to their presumably primary role in swimming. In *Pyrococcus furiosus*, flagella can form cable-like connections among cells and in adhesion to *Methanopyrus kandleri* cell surfaces [63]. Interactions of *P. furiosus* with *Methanopyrus* cells can occur through flagella, resulting in

a two-component archaeal biofilm [93]. In *Sulfolobus*, in addition to mediating swimming and swarming, flagella, along with pili, have both been shown to be involved in surface adhesion [94]. On the other hand, in *Hfx. volcanii*, flagella were shown not to be involved significantly in surface adhesion [58]. Instead, attachment is mediated by other type IV pilin-like proteins processed by a type IV prepilin peptidase-like enzyme. Swimming without flagella can occur via different mechanisms in bacteria [12] but, to date, no such nonflagellar-driven swimming modes have been reported in Archaea.

4. Posttranslational Modification of S-Layer Glycoproteins and Flagellins

In addition to serving important structural and physiological roles, S-layer glycoproteins and flagellins, in particular those from halophilic and methanogen archaea, are important reporters of posttranslational modifications, including signal sequence cleavage, glycosylation and lipid attachment (Figure 1).

4.1. Signal Peptide Cleavage. Bacteria can contain as many as three distinct signal peptidases, essential for removing the N-terminal signal peptides that target preproteins for export from the cytoplasm [95]. Signal peptidase I (SPI) is the housekeeping signal peptidase, responsible for cleaving the signal peptides from most preproteins secreted from the cell via either the Sec or TAT pathways. Signal peptidase II (SPII) removes signal peptides specifically from lipoproteins. Finally, type IV prepilin peptidases (TFPP, sometimes termed signal peptidase III, SPIII) are necessary for the cleavage of class III signal peptides from type IV pilins and related molecules. In Archaea, only SPI and TFPP have been identified [18].

Signal Peptidase I. Archaeal signal peptidase I was first identified in *Methanococcus voltae*, where the gene was cloned and the protein expressed and studied biochemically *in vitro*, using a heterologously expressed, truncated S-layer protein as substrate [96]. Site-directed mutagenesis studies on the methanogen enzyme and that of *Hfx. volcanii* reached the same conclusion; namely, that the archaeal enzyme relies on a different grouping of essential amino acids than does either the typical prokaryotic (P-type) enzyme found in Bacteria or eukaryotic (ER-type) enzyme [97, 98]. Bacterial P-type SPIs, found as well in mitochondria and chloroplasts, utilize a Ser-Lys dyad for catalysis. Site-directed mutagenesis studies revealed that Ser^{90} and Lys^{145} of *Escherichia coli* SPI are critical for enzymatic activity [99], while subsequent mutagenesis studies, based on crystal structure analysis, lead to the identification of Ser^{278} as being necessary for optimal activity [100]. In the case of ER-type SPIs, no lysine residues are essential for activity, reflecting the use of a different catalytic mechanism. Indeed, the conserved lysine of the P-type SPI is replaced by a conserved histidine in the ER-type SPI [101]. Further mutagenesis studies identified a conserved serine, histidine and two aspartic acid residues as being important for activity of the eukaryal enzyme, suggesting a potential catalytic mechanism relying on a Ser-His-Asp catalytic triad or a Ser-His catalytic dyad [101, 102]. Archaeal SPI maintains the conserved amino acids of the ER-type enzyme, notably the replacement of the conserved lysine by a histidine. However, site-directed mutagenesis again revealed archaeal-specific features. While the essential natures of the conserved serine and histidine residues were demonstrated, only one of the conserved aspartic acid residues was shown to be essential, unlike the yeast SPI, where both are essential [97, 98]. Almost all archaeal SPIs contain these four conserved amino acids [18], such that the mechanism of catalysis likely involves the Ser-His-Asp triad, as in the ER-type enzyme.

In *Hfx. volcanii*, two functional SPIs, that is, Sec11a and Sec11b, were identified, and although both are expressed, only Sec11b was deemed essential [103]. Since the two enzymes cleaved substrates differentially in an *in vitro* assay, they may serve distinct physiological roles.

FlaK/PibD: Archaeal Type IV Prepilin-Like Peptidases. The best studied of all archaeal signal peptidases are the TFPPs, represented mainly by FlaK/PibD [48, 76, 104], as well as by EppA [105]. These enzymes were first identified as TFPP homologues, able to cleave class III signal peptides from archaeal flagellins. Unlike bacterial flagellins, archaeal flagellins are synthesized as preproteins containing unusually short type IV pilin-like signal peptides. Such processing is an essential step in the assembly of archaeal flagella, as *flaK* mutants are nonflagellated [76].

Site-directed mutagenesis of both enzymes and substrate has greatly contributed to our knowledge of the mechanism of action and substrate range of archaeal TFPPs. Initial studies involving mutation of conserved amino acids in the signal peptide of model substrates, such as flagellins and sugar-binding proteins, revealed general similarities to the type IV pili system, where glycine at the -1 position of the signal peptide (i.e., the position immediately upstream of the cleavage site) was strongly preferred, with alanine shown to be an acceptable substitute. The -2 and -3 positions of the signal peptide are usually basic amino acids and, in the case of *Methanococcus* FlaK, the presence of lysine at the -2 position of the substrate is critical for cleavage [106]. *Sulfolobus* PibD is much less stringent in this regard, in keeping with the large number of potential substrates processed by this enzyme [104]. As revealed initially by signal peptide analysis of potential substrates [104] and later shown directly by *in vitro* assays [107], PibD is also able to accommodate extremely short signal peptides which are not processed by FlaK. Indeed, it has been suggested that PibD is a rarity among archaeal TFPPs in terms of its range of substrates [46], since, in *Methanococcus* for example, there are no sugar-binding proteins and type IV pilins are processed by a different dedicated TFPP, EppA [105]. Key to the activity of EppA seems to be the glutamine residue at position 1 found in all pilins but not in flagellins, since flagellins modified to include the pilin -2 to $+2$ amino acid region were cleaved [105]. In keeping with their processing of type IV prepilin-like molecules, site-directed mutagenesis

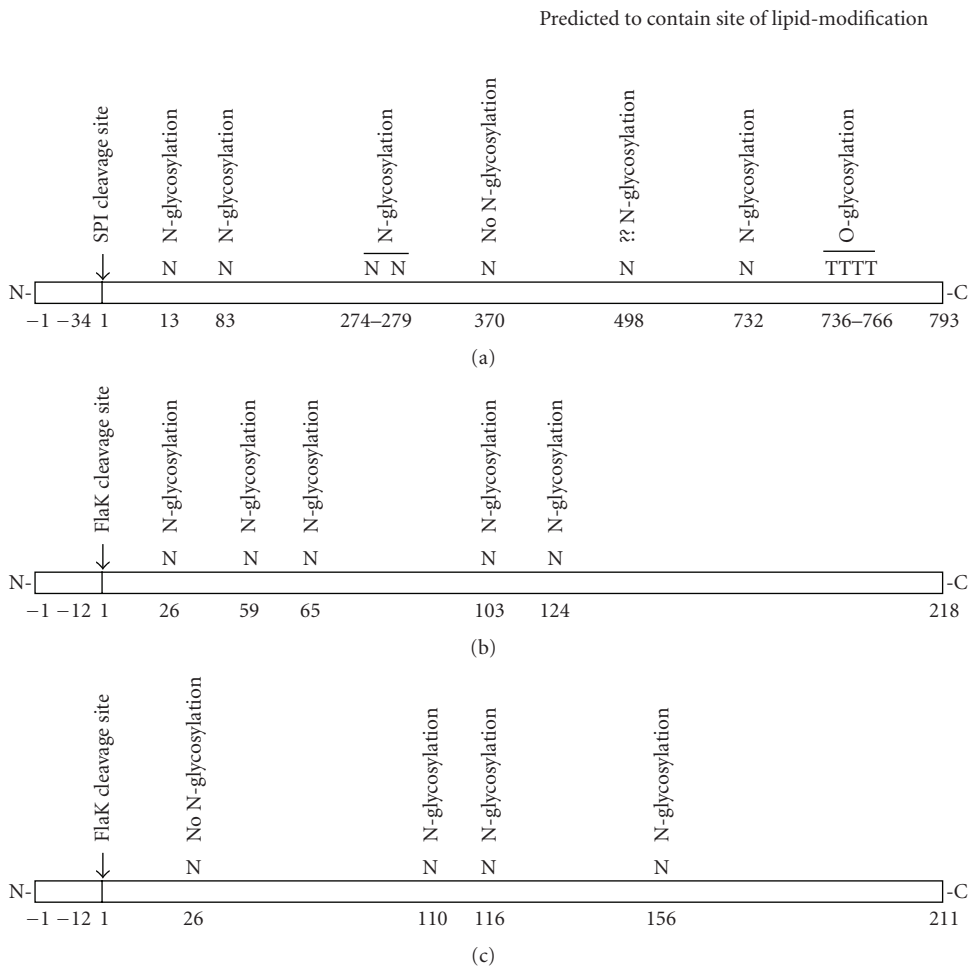


Figure 1: Schematic depiction of the posttranslational modifications experienced by the *Hfx. volcanii* S-layer glycoprotein and *Methanococcus* flagellins. A. *Hfx. volcanii* S-layer glycoprotein; B. *M. voltae* FlaB1; C. *M. maripaludis* FlaB1. Below each sequence, depicted by the elongated rectangle, amino acid residue positions are provided. Above each sequence, the residue at that position is listed, as is the posttranslational modification experienced by that residue or region of the protein. For sequon Asn residues, N-glycosylation, no N-glycosylation or unverified (??) N-glycosylation is marked. Note that the three sequences are not drawn to scale.

of both FlaK [76] and PibD [108] revealed that the pair of conserved aspartic acid residues that align with two aspartic acid residues shown to be essential for the bacterial type IV prepilin peptidase activity are also essential in the archaeal enzymes. Other conserved aspartic acid residues are not essential. Thus, both the bacterial and archaeal enzymes rely on the same catalytic mechanism and belong to the same family of novel aspartic acid proteases.

In *Methanococcus maripaludis*, both flagella and type IV-like pili are composed of major structural proteins possessing class III signal peptides. Interestingly, this species expresses two TFPPs, with FlaK specifically processing flagellins and EppA specifically cleaving the signal peptides from the prepilins [105]. This is not the case in *Sulfolobus solfataricus*, where PibD has a much broader substrate range, cleaving flagellins and type IV pilins, as well as a number of sugar binding proteins which have been hypothesized to form a pilus-like extension from the cell surface, termed the bindosome [8, 109]. In addition, the Iho670 fibers of *Ignicoccus hospitalis*, representing a novel archaeal surface structure, are also composed of subunits that contain class III signal peptides cleaved by a TFPP [110]. It thus appears that assembly of surface structures throughout the archaeal domain may rely heavily on the type IV pilus-like model.

5. N-Glycosylation

While the glycoprotein-based composition of the *Halobacterium* cell envelope had been previously suggested [111, 112], it was Mescher and Strominger [13] who purified and characterized the *Hbt. salinarum* S-layer glycoprotein, at the same time presenting the first example of a noneukaryal N-glycosylated protein. Initial efforts at describing the process of archaeal N-glycosylation revealed similarities to the parallel eukaryal process. In both cases, oligosaccharides are assembled on dolichol lipid carriers and transferred to target proteins following their delivery across a membrane; namely,

the ER membrane in Eukarya and the plasma membrane in Archaea (cf. [113]). However, it was only thirty years later, with the availability of complete genome sequences, that delineation of the biochemical pathways of archaeal N-glycosylation began in earnest. Almost simultaneously, *agl* (*a*rchaeal *gl*ycosylation) genes implicated in this posttranslational modification were identified in *Hfx. volcanii* [114] and *M. voltae* [115]. Subsequently, *agl* genes participating in the *Methanococcus maripaludis* N-glycosylation pathway were identified [116, 117].

Haloferax volcanii. In *Hfx. volcanii*, *agl* genes implicated in the assembly and attachment of a pentasaccharide to select Asn residues of the S-layer glycoprotein were first identified on the basis of their homologies to known N-glycosylation components in Eukarya or Bacteria (i.e., *Campylobacter jejuni*), the only bacterium for which a complete N-glycosylation pathway has been defined (cf. [118]). Subsequently, additional *agl* genes were identified either based upon their proximity to previously identified *agl* sequences or upon reannotation of that region of the genome where all but one of the previously identified *agl* sequences clustered [119–123]. In this manner, AglJ, AglG, AglI, AglE, and AglD, glycosyltransferases responsible for assembly of the S-layer glycoprotein-modifying pentasaccharide, comprising a hexose, two hexuronic acids, a methylester of hexuronic acid and a final hexose, were identified [123, 124]. AglB was shown to be the oligosaccharyltransferase responsible for delivery of the pentasaccharide, and apparently its precursors, to at least two residues of the S-layer glycoprotein; namely, Asn-13 and Asn-83 [123]. In addition, AglF, AglM, and AglP have also been shown to participate in the assembly of the pentasaccharide [120, 122, 124].

While involvement of each of these gene products in S-layer glycoprotein N-glycosylation has been shown through a combination of gene deletion and mass spectrometry approaches, in several instances, biochemical characterization has been carried out. In the case of AglF and AglM, the proteins have been purified and assays compatible with hypersaline conditions have been developed. Accordingly, AglF was shown to be a glucose-1-phosphate uridyltransferase, able to generate UDP-glucose from glucose-1-phosphate and UTP in a NAD^+-dependent manner, while AglM was revealed to function as a UDP-glucose dehydrogenase, generating UDP-glucuronic acid from UDP-glucose [122]. Indeed, a coupled reaction containing AglF, AglM, glucose-1-phosphate, UTP and NAD^+ led to the appearance of UDP-glucuronic acid, thus representing the first step towards in vitro reconstitution of the *Hfx. volcanii* N-glycosylation process. In the case of AglP, purification and subsequent development of an *in vitro* assay to test the function of the protein confirmed AglP to be a S-adenosyl-L-methionine-dependent methyltransferase [124], as predicted by earlier bioinformatics analysis [121]. Specifically, AglP acts on the fourth subunit of the pentasaccharide decorating the *Hfx. volcanii* S-layer glycoprotein, adding a methyl moiety to a hexuronic acid to yield the 190 Da methyl ester of hexuronic acid found at this position [124]. These results are summarized in Table 1.

Methanococcus voltae. The N-linked glycan described for *M. voltae* PS is a trisaccharide component with structure β-Man*p*NAcA6Thr-(1–4)-β-Glc*p*NAc3NAcA-(1–3)-β-Glc*p*NAc [33], although a strain harbouring a tetrasaccharide variant of this (the same trisaccharide as above with an extra 220 or 260 Da moiety attached) has been reported [125]. The glycan is linked to select asparagine residues present in flagellins and S-layer components via an N-acetylglucosamine, rather than the hexose observed in *Hfx. volcanii*. A combination of techniques, including insertional inactivation of targeted genes, immunoblot, heterologous expression studies and mass spectrometry analysis of purified flagella have identified the glycosyltransferases and the oligosaccharyltransferase required for the assembly and attachment of the glycan. AglH and AglA are responsible for the addition of the first and last sugar residues to the trisaccharide, respectively, [115, 126]. The role of AglH was elucidated by its ability to successfully complement a conditionally lethal mutation in *alg7* (N-acetylglucosamine-1-phosphate transferase) in yeast. AglC and AglK have both been implicated in the transfer of the second sugar residue [125]. Mutants in the oligosaccharyltransferase (*aglB*) contain S-layer glycoproteins and flagellins presenting molecular masses smaller than observed in any of the other *agl* mutants, consistent with this enzyme being responsible for the transfer of the N-glycan. The viability of strains carrying a disruption of *aglB* indicates that the N-linked glycosylation pathway is not essential in *M. voltae* [115].

Methanococcus maripaludis. With the development of advanced genetic tools [127], elucidation of the N-linked glycosylation pathway in methanogenic archaea has continued in *M. maripaludis,* where a tetrasaccharide glycan is N-linked to flagellin subunits [88]. The reported structure of the N-linked glycan was Sug-4-β-ManNAc3NAmA6Thr-4-β-GlcNAc3NAcA-3-β-GalNAc, where Sug was a previously unreported (5*S*)-2-acetamido-2,4-dideoxy-5-*O*-methyl-α-l-*erythro*-hexos-5-ulo-1,5-pyranose, representing the first example of a naturally occurring diglycoside of an aldulose [88]. Although the glycans of the two *Methanococcus* species are related, an obvious difference is that *M. maripaludis* uses N-acetylgalactosamine as the linking sugar, as compared to N-acetyglucosamine in *M. voltae*. In addition, the third sugar in both species is the same, except for a 3-acetamidino group addition in *M. maripaludis* that is carried out by the product of *MMP1081* (K. F. J., unpublished results). Interestingly, a strong homologue of MMP1081 is also found in the sequenced genome of *M. voltae* A3. Should this gene also be present in *M. voltae* PS; namely, that strain used for glycan structural study, it is unclear why an acetamidino group would not also be added here.

The genes *MMP1079*, *MMP1080* and *MMP1088*, designated *aglO*, *aglA* and *aglL*, respectively, have been implicated by deletion/complementation analysis and mass spectrometry as being the glycosyltransferases responsible for transfer of the second, third and fourth sugars to the glycan structure, respectively, [116]. As in *M. voltae,* but unlike the case in

Table 1: Effects of *Hfx. volcanii agl* deletions.

Gene	Role	Effect[1] of deletion on:						
		Hfx. volcanii Growth in high salt	S-layer Assembly	Shedding	Susceptibility to protease	S-layer glycoprotein SDS-PAGE migration	N-linked glycan structure	Reference
aglB	OTase[2]	Decreased	No effect	Increased	No effect	Increased	No glycan	[114, 123]
aglD	GTase[3] (sugar 5)	Decreased	Perturbed	Decreased	Decreased	Increased	Tetrasaccharide	[114, 123]
aglE	GTase (sugar 4)	No effect	n.d.	n.d.	No effect	No effect	Trisaccharide	[119]
aglF	Glucose-1-P uridyltransferase (sugar 3)	n.d.[4]	n.d.	n.d.	Increased	Increased	Disaccharide	[120]
aglG	GTase (sugar 2)	n.d.	n.d.	n.d.	Increased	Increased	Monosaccharide	[120]
aglI	GTase (sugar 3)	n.d.	n.d.	n.d.	Increased	Increased	Disaccharide	[120]
aglM	UDP-glucose dehydrogenase (sugars 2, 3 (4?))	n.d.	n.d.	n.d.	Increased	Increased	Monosaccharide	[122]
aglP	Methyltransferase (sugar 4)	n.d.	n.d.	n.d.	Increased	n.d.	Modified tetrasaccharide	[124]

[1] Relative to level detected in parent strain; increased, decreased or no effect.
[2] OTase: oligosaccharyltransferase.
[3] GTase: glycosyltransferase.
[4] n.d.: not determined.

Hfx. volcanii, where all but one of the *agl* genes are found in one large cluster, a*glB* is located elsewhere on the *M. maripaludis* chromosome. Its deletion leads to the appearance of nonglycosylated flagellins [116]. The glycosyltransferase responsible for the transfer of the first sugar residue has yet to be identified. Interestingly, mutants harboring deletions in genes that lead to a nonflagellated phenotype (i.e., *aglB* and *aglO*) initially synthesize normal levels of the flagellins and other cotranscribed *fla* gene products. However, upon continued laboratory sub-culturing, these strains appear to stop transcription of the entire *fla* operon. Other genes identified as involved in the glycan synthesis include *MMP0350,* the product of which is likely responsible for addition of one of the two acetyl groups found on the second sugar [117] and *MMP1085* which encodes a protein responsible for attachment of the methyl group to the terminal sugar (K. F. J. unpublished results). Available information on N-glycosylation in the two *Methanococcus* species is summarized in Table 2.

In addition to the genetic studies described, heterologous expression and in vitro biochemical and enzymatic studies of proteins predicted to be involved in the glycosylation pathway have helped described the biosynthesis of the acetamido sugar subunit precursors in methanococci [128].

6. O-Glycosylation

In addition to N-glycosylation, the *Hbt. salinarum* and *Hfx. volcanii* S-layer glycoproteins also undergo O-glycosylation. In each case, a Thr-rich region upstream of the predicted membrane-spanning domain of the protein is decorated at numerous positions by galactose-glucose disaccharides, linked through the galactose subunit [13, 129]. Essentially nothing is presently known of the archaeal O-glycosylation process.

7. Lipid Modification

In addition to signal peptide cleavage and glycosylation, haloarchaeal S-layer glycoproteins also experience covalent posttranslational attachment of lipids. This was first shown when *Hbt. salinarum* cells were incubated with [^{3}H]-mevalonate and other tritiated lipid tracers, leading to selective incorporation of the radiolabel into the S-layer glycoprotein [130]. The linked radioactive moiety was subsequently revealed by mass spectrometry to be a novel diphytanylglycerol phosphate. Although the precise location of the attached lipid has yet to be defined, a 28 kDa trypsin-generated fragment derived from the C-terminal region of the protein (residues 731–816) was shown to contain the linked group. In terms of attachment of the lipid, it is thought that phosphodiester-based linkage to either a S-layer glycoprotein Ser or Thr residue is responsible. Hence, it would appear that in addition to the single membrane-spanning domain located close to the C-terminus of the haloarchaeal S-layer glycoprotein, deduced from primary sequence analysis, a lipid moiety also anchors the protein to the membrane. Moreover, given sequence similarities in the same C-terminal region of the *Hbt. salinarum*, *Hfx. volcanii* and *Haloarcula japonica* S-layer glycoproteins [14, 129, 131], it is likely that the latter two similarly experience lipid modification [130]. Indeed, such lipid modification has been demonstrated in the case of the *Hfx. volcanii* S-layer glycoprotein [132, 133].

Table 2: Effects of *M. maripaludis* and *M. voltae agl* deletions.

Gene	Role	Effect[1] of deletion on:				
		Cell flagellation	Motility	Flagellin SDS-PAGE migration	N-linked glycan structure	Reference
M. maripaludis						
aglA	GTase[2] (sugar 3)	Present	Decreased	Increased	Disaccharide	[116]
aglB	OTase[3]	Absent	Non-motile	Increased	No glycan	[116]
aglL	GTase (sugar 4)	Present	Decreased	Increased	Modified trisaccharide	[116]
aglO	GTase (sugar 2)	Absent	Non-motile	Increased	Monosaccharide	[116]
MMP0350	Acetyltransferase (sugar 2)	Absent	Non-motile	Increased	Monosaccharide	[117]
MMP1081	Acetamidino transfer (sugar 3)	Absent	Decreased	Increased	Modified trisaccharide	unp
MMP1085	Methyltransferase (sugar 4)	Present	n.d.	Increased	Modified tetrasaccharide	unp
M. voltae						
aglA	GTase (sugar 3)	Present	n.d.	Increased	Disaccharide	[115]
aglB	OTase	Absent	Non-motile	Increased	No glycan	[115]
aglC	GTase (sugar 2)	Absent	n.d.	Increased	Monosaccharide	[125]
aglK	GTase (sugar 2)	Absent	n.d.	Increased	Monosaccharide	[125]

N. B.: *M. maripaludis* wild type N-linked glycan is a tetrasaccharide, *M. voltae* wild type N-linked glycan is a trisaccharide.
[1]Relative to level detected in parent strain.
[2]GTase: glycosyltransferase.
[3]OTase: oligosaccharyltransferase.
[4]n.d.: not determined.
[5]unp: unpublished data.

8. Importance of Flagellin and S-Layer Glycoprotein Posttranslational Modifications

The ability to generate deletion mutants of *Hfx. volcanii*, *M. voltae* and *M. maripaludis* as well as the availability of other molecular tools have allowed for the importance of posttranslational modifications of reporter proteins in these species to be addressed.

Haloferax volcanii. In *Hfx. volcanii*, the use of deletion strains has provided considerable insight into the importance of N-glycosylation to the cell. Strains lacking the ability to perform N-glycosylation, due to the absence of the oligosaccharyltransferase, AglB, or only able to partially recruit the N-glycosylation pathway, due to an absence of other Agl proteins, present various phenotypes, including an S-layer of modified architecture showing increased susceptibility to proteolytic digestion, enhanced S-layer glycoprotein release into the growth medium, and slower growth in medium of increasing salt [119, 120, 122, 123]. Indeed, differential transcription of the various Agl proteins in response to differing growth conditions, reflected by reverse transcription or real time PCR, points to N-glycosylation as being an adaptive process in *Hfx. volcanii*.

Studies addressing the biogenesis of the *Hfx. volcanii* S-layer glycoprotein have also provided insight into the importance of lipid modification. Metabolic [^{35}S] pulse-chase radiolabelling, together with the use of the ribosome-targeted antibiotic, anisomycin, revealed the S-layer glycoprotein to undergo a posttranslational maturation step on the outer surface of the plasma membrane, reflected as an increase in the hydrophobicity and apparent molecular weight of the protein [133]. Support for lipid modification as being responsible for S-layer glycoprotein maturation came from experiments showing that growth in the presence of [^{3}H] mevalonic acid led to radiolabel being incorporated into the S-layer glycoprotein and that mevinolin, an inhibitor of 3-HMG-CoA reductase (responsible for converting acetyl-CoA into mevalonic acid), prevented the maturation of the S-layer glycoprotein [132].

Moreover, such lipid modification-based maturation does not occur in the absence of Mg^{2+} [133], required for maintaining haloarchaeal S-layer integrity [13, 26]. As the *Hbt. salinarum* S-layer glycoprotein also undergoes a similar lipid-based maturation step [132], this posttranslational modification may be common to S-layer glycoprotein biogenesis in other haloarchaea.

Methanococcus voltae and Methanococcus maripaludis. The flagellins expressed by both *M. voltae* and *M. maripaludis* undergo two major posttranslational modifications necessary for their correct assembly into flagella, namely signal peptide cleavage and N-linked glycosylation. Disruption or deletion of the signal peptidase gene (*flaK)* results in cells that are no longer able to assemble flagella, with the unprocessed flagellins remaining in the cytoplasmic membrane [76]. In

the case of *M. maripaludis*, these cells are, however, still piliated, since the type IV pilin-like proteins are processed by a separate prepilin peptidase-like enzyme, EppA [105]. Deletion of *eppA* in a cell that is already deleted for *flaK* results in the appearance of nonflagellated and nonpiliated cells (K. F. J., unpublished results).

Early studies pointed to a critical role for glycosylation in flagella structure in *Methanococcus*. When incubated with bacitracin, a known inhibitor of glycosylation, *Methanococcus deltae* cells became nonflagellated, accompanied by a decrease in apparent molecular weight of the flagellins, suggestive of under-glycosylation [134]. More recently, in both *M. voltae* and *M. maripaludis*, it was shown that at least a two sugar-member glycan must be attached to flagellin subunits for proper assembly of the protein into flagella filaments [115, 116]. Deletion of genes involved in the N-glycosylation pathway of both *M. voltae* and *M. maripaludis* result in flagellin subunits that migrate faster on SDS-PAGE, with the enhanced migration corresponding incrementally to the degree of truncation of the glycan [115, 116]. Motility assays using semisolid agar demonstrated that strains harboring deletions of genes in this pathway that are still able to assemble flagella displayed impaired swimming capabilities, as compared to cells able to produce the native N-linked glycan [116]. Finally, deletion of a single gene, that is, *MMP0350*, assigned as an acetyltransferase necessary for the biosynthesis of the second sugar of the *M. maripaludis* N-linked glycan, resulted in defects in both flagellation and piliation. Since the glycan consisted of only a single sugar in this mutant, the fact that the cells were nonflagellated was not unexpected. However, further examination revealed that while these mutants were generally nonpiliated, apparently intact pili were found in the culture supernatants, indicating that a defect in pili anchoring had occurred [117].

9. Conclusions

As the research spotlight begins to shift from the genome to the proteome, it is becoming clear that numerous archaeal proteins experience posttranslational modifications [135–139]. As discussed here, the availability of well-characterized reporters of protein processing events, such as haloarchaeal and methanoarchaeal S-layer glycoproteins and flagellins, offer excellent models in studies attempting to dissect the pathways responsible for such modifications. Along with the identification of additional reporter proteins, it will be possible to consider questions related to the heterogeneity associated with a given modification, such as differential or modulated glycosylation. Moreover, with the development of appropriate *in vitro* assays for these novel reporters, future efforts can address the importance of posttranslational modifications to enzyme function, stability and other traits.

Acknowledgments

J. Eichler is supported by grants from the Israel Science Foundation (Grant 30/07) and the US Army Research Office (Grant W911NF-07-1-0260). K. F. Jarrell is supported by a Discovery Grant from the Natural Sciences and Engineering Research Council of Canada (NSERC).

References

[1] C. R. Woese, "The archaeal concept and the world it lives in: a retrospective," *Photosynthesis Research*, vol. 80, no. 1–3, pp. 361–372, 2004.

[2] G. E. Fox, E. Stackebrandt, R. B. Hespell, et al., "The phylogeny of prokaryotes," *Science*, vol. 209, no. 4455, pp. 457–463, 1980.

[3] C. R. Woese and G. E. Fox, "Phylogenetic structure of the prokaryotic domain: the primary kingdoms," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 74, no. 11, pp. 5088–5090, 1977.

[4] W. Zillig, "Comparative biochemistry of Archaea and Bacteria," *Current Opinion in Genetics and Development*, vol. 1, no. 4, pp. 544–551, 1991.

[5] K. F. Jarrell, D. J. VanDyke, and J. Wu, "Archaeal flagella and pili," in *Pili and Flagella: Current Research and Future Trends*, K. F. Jarrell, Ed., pp. 215–234, Caister Academic Press, Norfolk, UK, 2009.

[6] O. Kandler and H. Konig, "Cell envelopes of archaebacteria," in *The Bacteria. A Treatise on Structure and Function*, Academic Press, Orlando, Fla, USA, 1985.

[7] L. J. Magrum, K. R. Luehrsen, and C. R. Woese, "Are extreme halophiles actually "bacteria"?" *Journal of Molecular Evolution*, vol. 11, no. 1, pp. 1–8, 1978.

[8] S. Y. M. Ng, B. Zolghadr, A. J. M. Driessen, S.-V. Albers, and K. F. Jarrell, "Cell surface structures of archaea," *Journal of Bacteriology*, vol. 190, no. 18, pp. 6039–6047, 2008.

[9] S. Nickell, R. Hegerl, W. Baumeister, and R. Rachel, "*Pyrodictium* cannulae enter the periplasmic space but do not enter the cytoplasm, as revealed by cryo-electron tomography," *Journal of Structural Biology*, vol. 141, no. 1, pp. 34–42, 2003.

[10] C. Moissl, R. Rachel, A. Briegel, H. Engelhardt, and R. Huber, "The unique structure of archaeal 'hami', highly complex cell appendages with nano-grappling hooks," *Molecular Microbiology*, vol. 56, no. 2, pp. 361–370, 2005.

[11] Y. A. Wang, X. Yu, S. Y. M. Ng, K. F. Jarrell, and E. H. Egelman, "The structure of an archaeal pilus," *Journal of Molecular Biology*, vol. 381, no. 2, pp. 456–466, 2008.

[12] K. F. Jarrell and M. J. McBride, "The surprisingly diverse ways that prokaryotes move," *Nature Reviews Microbiology*, vol. 6, no. 6, pp. 466–476, 2008.

[13] M. F. Mescher and J. L. Strominger, "Purification and characterization of a prokaryotic glycoprotein from the cell envelope of *Halobacterium salinarium*," *The Journal of Biological Chemistry*, vol. 251, no. 7, pp. 2005–2014, 1976.

[14] J. Lechner and M. Sumper, "The primary structure of a procaryotic glycoprotein. Cloning and sequencing of the cell surface glycoprotein gene of halobacteria," *The Journal of Biological Chemistry*, vol. 262, no. 20, pp. 9724–9729, 1987.

[15] R. Mengele and M. Sumper, "Drastic differences in glycosylation of related S-layer glycoproteins from moderate and extreme halophiles," *The Journal of Biological Chemistry*, vol. 267, no. 12, pp. 8182–8185, 1992.

[16] M. Sumper, "Halobacterial glycoprotein biosynthesis," *Biochimica et Biophysica Acta*, vol. 906, no. 1, pp. 69–79, 1987.

[17] S. Yurist-Doutsch, B. Chaban, D. J. VanDyke, K. F. Jarrell, and J. Eichler, "Sweet to the extreme: protein glycosylation in Archaea," *Molecular Microbiology*, vol. 68, no. 5, pp. 1079–1084, 2008.

[18] S. Y. M. Ng, B. Chaban, D. J. VanDyke, and K. F. Jarrell, "Archaeal signal peptidases," *Microbiology*, vol. 153, no. 2, pp. 305–314, 2007.

[19] M. Abu-Qarn, J. Eichler, and N. Sharon, "Not just for Eukarya anymore: protein glycosylation in Bacteria and Archaea," *Current Opinion in Structural Biology*, vol. 18, no. 5, pp. 544–550, 2008.

[20] M. Pohlschröder, M. I. Giménez, and K. F. Jarrell, "Protein transport in Archaea: sec and twin arginine translocation pathways," *Current Opinion in Microbiology*, vol. 8, no. 6, pp. 713–719, 2005.

[21] A. L. Houwink, "Flagella, gas vacuoles and cell-wall structure in *Halobacterium halobium*; an electron microscope study," *Journal of General Microbiology*, vol. 15, pp. 146–150, 1956.

[22] W. Stoeckenius and R. Rowen, "A morphological study of *Halobacterium halobium* and its lysis in media of low salt concentration," *Journal of Cell Biology*, vol. 34, no. 1, pp. 365–393, 1967.

[23] H. Steensland and H. Larsen, "A study of the cell envelope of the halobacteria," *Journal of General Microbiology*, vol. 55, no. 3, pp. 325–336, 1969.

[24] R. G. Kirk and M. Ginzburg, "Ultrastructure of two species of *Halobacterium*," *Journal of Ultrasructure Research*, vol. 41, no. 1-2, pp. 80–94, 1972.

[25] A. E. Blaurock, W. Stoeckenius, D. Oesterhelt, and G. L. Scherphof, "Structure of the cell envelope of *Halobacterium halobium*," *Journal of Cell Biology*, vol. 71, no. 1, pp. 1–22, 1976.

[26] M. Kessel, I. Wildhaber, S. Cohen, and W. Baumeister, "Three-dimensional structure of the regular surface glycoprotein layer of *Halobacterium volcanii* from the Dead Sea," *The EMBO Journal*, vol. 7, pp. 1549–1554, 1988.

[27] S. Trachtenberg, B. Pinnick, and M. Kessel, "The cell surface glycoprotein layer of the extreme halophile *Halobacterium salinarum* and its relation to *Haloferax volcanii*: cryo-electron tomography of freeze-substituted cells and projection studies of negatively stained envelopes," *Journal of Structural Biology*, vol. 130, no. 1, pp. 10–26, 2000.

[28] O. Kandler and H. Konig, "Cell envelopes of archaea: structure and chemistry," in *The Biochemistry of Archaea*, M. Kates, D. J. Kusher, and A. T. Matheson, Eds., pp. 223–259, Elsevier, Amsterdam, The Netherlands, 1993.

[29] H. Konig, K. O. Stetter, W. Postulka, and F. Klink, "Studies on archaebacterial S-layers," *Systematic and Applied Microbiology*, vol. 7, pp. 300–309, 1986.

[30] E. Nusser and H. Konig, "S layer studies on three species of *Methanococcus* living at different temperatures," *Canadian Journal of Microbiology*, vol. 33, no. 3, pp. 256–261, 1987.

[31] E. Akca, H. Claus, N. Schultz et al., "Genes and derived amino acid sequences of S-layer proteins from mesophilic, thermophilic, and extremely thermophilic methanococci," *Extremophiles*, vol. 6, no. 5, pp. 351–358, 2002.

[32] S. F. Koval and K. F. Jarrell, "Ultrastructure and biochemistry of the cell wall of *Methanococcus voltae*," *Journal of Bacteriology*, vol. 169, no. 3, pp. 1298–1306, 1987.

[33] S. Voisin, R. S. Houliston, J. Kelly et al., "Identification and characterization of the unique N-linked glycan common to the flagellins and S-layer glycoprotein of *Methanococcus voltae*," *The Journal of Biological Chemistry*, vol. 280, no. 17, pp. 16586–16593, 2005.

[34] G. B. Patel, C. G. Choquet, J. H. E. Nash, and G. D. Sprott, "Formation and regeneration of *Methanococcus voltae* protoplasts," *Applied and Environmental Microbiology*, vol. 59, no. 1, pp. 27–33, 1993.

[35] G. B. Patel, J. H. E. Nash, B. J. Agnew, and G. D. Sprott, "Natural and electroporation-mediated transformation of *Methanococcus voltae* protoplasts," *Applied and Environmental Microbiology*, vol. 60, no. 3, pp. 903–907, 1994.

[36] U. Karcher, H. Schroder, E. Haslinger et al., "Primary structure of the heterosaccharide of the surface glycoprotein of *Methanothermus fervidus*," *The Journal of Biological Chemistry*, vol. 268, no. 36, pp. 26821–26826, 1993.

[37] K. R. Sowers, J. E. Boone, and R. P. Gunsalus, "Disaggregation of *Methanosarcina* spp. and growth as single cells at elevated osmolarity," *Applied and Environmental Microbiology*, vol. 59, no. 11, pp. 3832–3839, 1993.

[38] G.-W. Cheong, Z. Cejka, J. Peters, K. O. Stetter, and W. Baumeister, "The surface protein layer of *Methanoplanus limicola*: three-dimensional structure and chemical characterization," *Systematic and Applied Microbiology*, vol. 14, no. 3, pp. 209–217, 1991.

[39] G.-W. Cheong, R. Guckenberger, K.-H. Fuchs, H. Gross, and W. Baumeister, "The structure of the surface layer of *Methanoplanus limicola* obtained by a combined electron microscopy and scanning tunneling microscopy approach," *Journal of Structural Biology*, vol. 111, no. 2, pp. 125–134, 1993.

[40] M. Firtel, G. Southam, G. Harauz, and T. J. Beveridge, "Characterization of the cell wall of the sheathed methanogen *Methanospirillum hungatei* GP1 as an S layer," *Journal of Bacteriology*, vol. 175, no. 23, pp. 7550–7560, 1993.

[41] M. Stewart, T. J. Beveridge, and G. D. Sprott, "Crystalline order to high resolution in the sheath of *Methanospirillum hungatei*: a cross-beta structure," *Journal of Molecular Biology*, vol. 183, no. 3, pp. 509–515, 1985.

[42] T. J. Beveridge, G. D. Sprott, and P. Whippey, "Ultrastructure, inferred porosity, and gram-staining character of *Methanospirillum hungatei* filament termini describe a unique cell permeability for this archaeobacterium," *Journal of Bacteriology*, vol. 173, no. 1, pp. 130–140, 1991.

[43] D. Pum, P. Messner, and U. B. Sleytr, "Role of the S layer in morphogenesis and cell division of the archaebacterium *Methanocorpusculum sinense*," *Journal of Bacteriology*, vol. 173, no. 21, pp. 6865–6873, 1991.

[44] K. F. Jarrell, S. Y. Ng, and B. Chaban, "Flagellation and chemotaxis," in *Archaea: Molecular and Cellular Biology*, R. Cavicchioli, Ed., pp. 385–410, ASM Press, Washington, DC, USA, 2007.

[45] K. F. Jarrell, D. P. Bayley, and A. S. Kostyukova, "The archaeal flagellum: a unique motility structure," *Journal of Bacteriology*, vol. 178, no. 17, pp. 5057–5064, 1996.

[46] S. Y. M. Ng, B. Chaban, and K. F. Jarrell, "Archaeal flagella, bacterial flagella and type IV pili: a comparison of genes and posttranslational modifications," *Journal of Molecular Microbiology and Biotechnology*, vol. 11, no. 3-5, pp. 167–191, 2006.

[47] M. L. Kalmokoff and K. F. Jarrell, "Cloning and sequencing of a multigene family encoding the flagellins of *Methanococcus voltae*," *Journal of Bacteriology*, vol. 173, no. 22, pp. 7113–7125, 1991.

[48] S. L. Bardy and K. F. Jarrell, "FlaK of the archaeon *Methanococcus maripaludis* possesses preflagellin peptidase activity," *FEMS Microbiology Letters*, vol. 208, no. 1, pp. 53–59, 2002.

[49] S. L. Bardy, T. Mori, K. Komoriya, S.-I. Aizawa, and K. F. Jarrell, "Identification and localization of flagellins FlaA and FlaB3 within flagella of *Methanococcus voltae*," *Journal of Bacteriology*, vol. 184, no. 19, pp. 5223–5233, 2002.

[50] D. M. Faguy, S. F. Koval, and K. F. Jarrell, "Physical characterization of the flagella and flagellins from *Methanospirillum hungatei*," *Journal of Bacteriology*, vol. 176, no. 24, pp. 7491–7498, 1994.

[51] G. Southam, M. L. Kalmokoff, K. F. Jarrell, S. F. Koval, and T. J. Beveridge, "Isolation, characterization, and cellular insertion of the flagella from two strains of the archaebacterium *Methanospirillum hungatei*," *Journal of Bacteriology*, vol. 172, no. 6, pp. 3221–3228, 1990.

[52] D. Cruden, R. Sparling, and A. J. Markovetz, "Isolation and ultrastructure of the flagella of *Methanococcus thermolithotrophicus* and *Methanospirillum hungatei*," *Applied and Environmental Microbiology*, vol. 55, no. 6, pp. 1414–1419, 1989.

[53] M. Alam and D. Oesterhelt, "Morphology, function and isolation of halobacterial flagella," *Journal of Molecular Biology*, vol. 176, no. 4, pp. 459–475, 1984.

[54] M. Alam and D. Oesterhelt, "Purification, reconstitution and polymorphic transition of halobacterial flagella," *Journal of Molecular Biology*, vol. 194, no. 3, pp. 495–499, 1987.

[55] L. Gerl, R. Deutzmann, and M. Sumper, "Halobacterial flagellins are encoded by a multigene family Identification of all five gene products," *FEBS Letters*, vol. 244, no. 1, pp. 137–140, 1989.

[56] N. Patenge, A. Berendes, H. Engelhardt, S. C. Schuster, and D. Oesterhelt, "The *fla* gene cluster is involved in the biogenesis of flagella in *Halobacterium salinarum*," *Molecular Microbiology*, vol. 41, no. 3, pp. 653–663, 2001.

[57] M. G. Pyatibratov, S. N. Beznosov, R. Rachel et al., "Alternative flagellar filament types in the haloarchaeon *Haloarcula marismortui*," *Canadian Journal of Microbiology*, vol. 54, no. 10, pp. 835–844, 2008.

[58] M. Tripepi, S. Imam, and M Pohlschroder, "*Haloferax volcanii* flagella are required for motility but are not involved in PibD-dependent surface adhesion," *Journal of Bacteriology*, vol. 192, no. 12, pp. 3093–3102, 2010.

[59] Z. Szabó, M. Sani, M. Groeneveld et al., "Flagellar motility and structure in the hyperthermoacidophilic archaeon *Sulfolobus solfataricus*," *Journal of Bacteriology*, vol. 189, no. 11, pp. 4305–4309, 2007.

[60] M. G. Pyatibratov, K. Leonard, V. Y. Tarasov, and O. V. Fedorov, "Two immunologically distinct types of protofilaments can be identified in *Natrialba magadii* flagella," *FEMS Microbiology Letters*, vol. 212, no. 1, pp. 23–27, 2002.

[61] I. Serganova, V. Ksenzenko, A. Serganov et al., "Sequencing of flagellin genes from *Natrialba magadii* provides new insight into evolutionary aspects of archaeal flagellins," *Journal of Bacteriology*, vol. 184, no. 1, pp. 318–322, 2002.

[62] K. Nagahisa, S. Ezaki, S. Fujiwara, T. Imanaka, and M. Takagi, "Sequence and transcriptional studies of five clustered flagellin genes from hyperthermophilic archaeon *Pyrococcus kodakaraensis* KOD1," *FEMS Microbiology Letters*, vol. 178, no. 1, pp. 183–190, 1999.

[63] D. J. Näther, R. Rachel, G. Wanner, and R. Wirth, "Flagella of *Pyrococcus furiosus*: multifunctional organelles, made for swimming, adhesion to various surfaces, and cell-cell contacts," *Journal of Bacteriology*, vol. 188, no. 19, pp. 6915–6923, 2006.

[64] B. Chaban, S. Y. M. Ng, M. Kanbe et al., "Systematic deletion analyses of the *fla* genes in the flagella operon identify several genes essential for proper assembly and function of flagella in the archaeon, *Methanococcus maripaludis*," *Molecular Microbiology*, vol. 66, no. 3, pp. 596–609, 2007.

[65] A. F. Ellen, B. Zolghadr, A. J. M. Driessen, and S. V. Albers, "Shaping the Archaeal Cell Envelope," *Archaea*, vol. 2010, Article ID 608243, 2010.

[66] R. C. H. del Rosario, F. Diener, M. Diener, and D. Oesterhelt, "The steady-state phase distribution of the motor switch complex model of *Halobacterium salinarum*," *Mathematical Biosciences*, vol. 222, no. 2, pp. 117–126, 2009.

[67] T. Nutsch, D. Oesterhelt, E. D. Gilles, and W. Marwan, "A quantitative model of the switch cycle of an archaeal flagellar motor and its sensory control," *Biophysical Journal*, vol. 89, no. 4, pp. 2307–2323, 2005.

[68] W. Marwan, M. Alam, and D. Oesterhelt, "Rotation and switching of the flagellar motor assembly in *Halobacterium halobium*," *Journal of Bacteriology*, vol. 173, no. 6, pp. 1971–1977, 1991.

[69] D. M. Faguy and K. F. Jarrell, "A twisted tale: the origin and evolution of motility and chemotaxis in prokaryotes," *Microbiology*, vol. 145, no. 2, pp. 279–281, 1999.

[70] D. F. Blair, "Flagellar movement driven by proton translocation," *FEBS Letters*, vol. 545, no. 1, pp. 86–95, 2003.

[71] S. Streif, W. F. Staudinger, W. Marwan, and D. Oesterhelt, "Flagellar rotation in the archaeon *Halobacterium salinarum* depends on ATP," *Journal of Molecular Biology*, vol. 384, no. 1, pp. 1–8, 2008.

[72] D. E. Bradley, "A function of *Pseudomonas aeruginosa* PAO polar pili: twitching motility," *Canadian Journal of Microbiology*, vol. 26, no. 2, pp. 146–154, 1980.

[73] S. Cohen-Krausz and S. Trachtenberg, "The flagellar filament structure of the extreme acidothermophile *Sulfolobus shibatae* B12 suggests that archaeabacterial flagella have a unique and common symmetry and design," *Journal of Molecular Biology*, vol. 375, no. 4, pp. 1113–1124, 2008.

[74] S. Cohen-Krausz and S. Trachtenberg, "The structure of the archeabacterial flagellar filament of the extreme halophile *Halobacterium salinarum* R1M1 and its relation to eubacterial flagellar filaments and type IV pili," *Journal of Molecular Biology*, vol. 321, no. 3, pp. 383–395, 2002.

[75] S. Trachtenberg and S. Cohen-Krausz, "The archaeabacterial flagellar filament: a bacterial propeller with a pilus-like structure," *Journal of Molecular Microbiology and Biotechnology*, vol. 11, no. 3–5, pp. 208–220, 2006.

[76] S. L. Bardy and K. F. Jarrell, "Cleavage of preflagellins by an aspartic acid signal peptidase is essential for flagellation in the archaeon *Methanococcus voltae*," *Molecular Microbiology*, vol. 50, no. 4, pp. 1339–1347, 2003.

[77] D. P. Bayley and K. F. Jarrell, "Further evidence to suggest that archaeal flagella are related to bacterial type IV pili," *Journal of Molecular Evolution*, vol. 46, no. 3, pp. 370–373, 1998.

[78] C. R. Peabody, Y. J. Chung, M.-R. Yen, D. Vidal-Ingigliardi, A. P. Pugsley, and M. H. Saier Jr., "Type II protein secretion and its relationship to bacterial type IV pili and archaeal flagella," *Microbiology*, vol. 149, no. 11, pp. 3051–3072, 2003.

[79] R. M. Macnab, "How bacteria assemble flagella," *Annual Review of Microbiology*, vol. 57, pp. 77–100, 2003.

[80] K. F. Jarrell, D. P. Bayley, V. Florian, and A. Klein, "Isolation and characterization of insertional mutations in flagellin genes in the archaeon *Methanococcus voltae*," *Molecular Microbiology*, vol. 20, no. 3, pp. 657–666, 1996.

[81] V. Y. Tarasov, M. G. Pyatibratov, S.-L. Tang, M. Dyall-Smith, and O. V. Fedorov, "Role of flagellins from A and B loci in flagella formation of *Halobacterium salinarum*," *Molecular Microbiology*, vol. 35, no. 1, pp. 69–78, 2000.

[82] S. N. Beznosov, M. G. Pyatibratov, and O. V. Fedorov, "On the multicomponent nature of *Halobacterium salinarum* flagella," *Microbiology*, vol. 76, no. 4, pp. 435–441, 2007.

[83] S.-V. Albers and A. J. M. Driessen, "Analysis of ATPases of putative secretion operons in the thermoacidophilic archaeon *Sulfolobus solfataricus*," *Microbiology*, vol. 151, no. 3, pp. 763–773, 2005.

[84] N. A. Thomas and K. F. Jarrell, "Characterization of flagellum gene families of methanogenic archaea and localization of novel flagellum accessory proteins," *Journal of Bacteriology*, vol. 183, no. 24, pp. 7154–7164, 2001.

[85] N. A. Thomas, S. Mueller, A. Klein, and K. F. Jarrell, "Mutants in *fla*I and *fla*J of the archaeon *Methanococcus voltae* are deficient in flagellum assembly," *Molecular Microbiology*, vol. 46, no. 3, pp. 879–887, 2002.

[86] M. Schlesner, A. Miller, S. Streif et al., "Identification of Archaea-specific chemotaxis proteins which interact with the flagellar apparatus," *BMC Microbiology*, vol. 9, article 56, 2009.

[87] S. M. Logan, "Flagellar glycosylation—a new component of the motility repertoire?" *Microbiology*, vol. 152, no. 5, pp. 1249–1262, 2006.

[88] J. Kelly, S. M. Logan, K. F. Jarrell, D. J. VanDyke, and E. Vinogradov, "A novel N-linked flagellar glycan from *Methanococcus maripaludis*," *Carbohydrate Research*, vol. 344, no. 5, pp. 648–653, 2009.

[89] F. Wieland, G. Paul, and M. Sumper, "Halobacterial flagellins are sulfated glycoproteins," *The Journal of Biological Chemistry*, vol. 260, no. 28, pp. 15180–15185, 1985.

[90] R. M. Harshey, "Bacterial motility on a surface: many ways to a common goal," *Annual Review of Microbiology*, vol. 57, pp. 249–273, 2003.

[91] J. K. Anderson, T. G. Smith, and T. R. Hoover, "Sense and sensibility: flagellum-mediated gene regulation," *Trends in Microbiology*, vol. 18, no. 1, pp. 30–37, 2010.

[92] Q. Wang, A. Suzuki, S. Mariconda, S. Porwollik, and R. M. Harshey, "Sensing wetness: a new role for the bacterial flagellum," *EMBO Journal*, vol. 24, no. 11, pp. 2034–2042, 2005.

[93] S. Schopf, G. Wanner, R. Rachel, and R. Wirth, "An archaeal bi-species biofilm formed by *Pyrococcus furiosus* and *Methanopyrus kandleri*," *Archives of Microbiology*, vol. 190, no. 3, pp. 371–377, 2008.

[94] B. Zolghadr, A. Kling, A. Koerdt, A. J.M. Driessen, R. Rachel, and S.-V. Albers, "Appendage-mediated surface adherence of *Sulfolobus solfataricus*," *Journal of Bacteriology*, vol. 192, no. 1, pp. 104–110, 2010.

[95] M. Paetzel, A. Karla, N. C. J. Strynadka, and R. E. Dalbey, "Signal peptidases," *Chemical Reviews*, vol. 102, no. 12, pp. 4549–4580, 2002.

[96] S. Y. M. Ng and K. F. Jarrell, "Cloning and characterization of archaeal type I signal peptidase from *Methanococcus voltae*," *Journal of Bacteriology*, vol. 185, no. 20, pp. 5936–5942, 2003.

[97] S. L. Bardy, S. Y. M. Ng, D. S. Carnegie, and K. F. Jarrell, "Site-directed mutagenesis analysis of amino acids critical for activity of the type I signal peptidase of the archaeon *Methanococcus voltae*," *Journal of Bacteriology*, vol. 187, no. 3, pp. 1188–1191, 2005.

[98] E. Fink-Lavi and J. Eichler, "Identification of residues essential for the catalytic activity of Sec11b, one of the two type I signal peptidases of *Haloferax volcanii*," *FEMS Microbiology Letters*, vol. 278, no. 2, pp. 257–260, 2008.

[99] W. R. Tschantz, M. Sung, V. M. Delgado-Partin, and R. E. Dalbey, "A serine and a lysine residue implicated in the catalytic mechanism of the *Escherichia coli* leader peptidase," *The Journal of Biological Chemistry*, vol. 268, no. 36, pp. 27349–27354, 1993.

[100] M. Paetzel, R. E. Dalbey, and N. C. J. Strynadka, "Crystal structure of a bacterial signal peptidase apoenzyme. Implications for signal peptide binding and the Ser-Lys dyad mechanism," *The Journal of Biological Chemistry*, vol. 277, no. 11, pp. 9512–9519, 2002.

[101] R. E. Dalbey, M. O. Lively, S. Bron, and J. M. van Dijl, "The chemistry and enzymology of the type I signal peptidases," *Protein Science*, vol. 6, no. 6, pp. 1129–1138, 1997.

[102] C. VanValkenburgh, X. Chen, C. Mullins, H. Fang, and N. Green, "The catalytic mechanism of endoplasmic reticulum signal peptidase appears to be distinct from most eubacterial signal peptidases," *The Journal of Biological Chemistry*, vol. 274, no. 17, pp. 11519–11525, 1999.

[103] A. Fine, V. Irihimovitch, I. Dahan, Z. Konrad, and J. Eichler, "Cloning, expression, and purification of functional Sec11a and Sec11b, type I signal peptidases of the archaeon *Haloferax volcanii*," *Journal of Bacteriology*, vol. 188, no. 5, pp. 1911–1919, 2006.

[104] S.-V. Albers, Z. Szabó, and A. J. M. Driessen, "Archaeal homolog of bacterial type IV prepilin signal peptidases with broad substrate specificity," *Journal of Bacteriology*, vol. 185, no. 13, pp. 3918–3925, 2003.

[105] Z. Szabó, A. O. Stahl, S.-V. Albers, J. C. Kissinger, A. J. M. Driessen, and M. Pohlschröder, "Identification of diverse archaeal proteins with class III signal peptides cleaved by distinct archaeal prepilin peptidases," *Journal of Bacteriology*, vol. 189, no. 3, pp. 772–778, 2007.

[106] N. A. Thomas, E. D. Chao, and K. F. Jarrell, "Identification of amino acids in the leader peptide of *Methanococcus voltae* preflagellin that are important in posttranslational processing," *Archives of Microbiology*, vol. 175, no. 4, pp. 263–269, 2001.

[107] S. Y. M. Ng, D. J. VanDyke, B. Chaban et al., "Different minimal signal peptide lengths recognized by the archaeal prepilin-like peptidases FlaK and PibD," *Journal of Bacteriology*, vol. 191, no. 21, pp. 6732–6740, 2009.

[108] Z. Szabó, S.-V. Albers, and A. J. M. Driessen, "Active-site residues in the type IV prepilin peptidase homologue PibD from the archaeon *Sulfolobus solfataricus*," *Journal of Bacteriology*, vol. 188, no. 4, pp. 1437–1443, 2006.

[109] S.-V. Albers and M. Pohlschröder, "Diversity of archaeal type IV pilin-like structures," *Extremophiles*, vol. 13, pp. 403–410, 2009.

[110] D. W. Müller, C. Meyer, S. Gürster et al., "The Iho670 fibers of *Ignicoccus hospitalis*: a new type of archaeal cell surface appendage," *Journal of Bacteriology*, vol. 191, no. 20, pp. 6465–6468, 2009.

[111] M. A. Koncewicz, "Glycoproteins in the cell envelope of *Halobacterium halobium*," *Biochemical Journal*, vol. 128, no. 4, p. 124, 1972.

[112] M. F. Mescher, J. L. Strominger, and S. W. Watson, "Protein and carbohydrate composition of the cell envelope of *Halobacterium salinarium*," *Journal of Bacteriology*, vol. 120, no. 2, pp. 945–954, 1974.

[113] J. Eichler and M. W. W. Adams, "Posttranslational protein modification in Archaea," *Microbiology and Molecular Biology Reviews*, vol. 69, no. 3, pp. 393–425, 2005.

[114] M. Abu-Qarn and J. Eichler, "Protein N-glycosylation in Archaea: defining *Haloferax volcanii* genes involved in S-layer

glycoprotein glycosylation," *Molecular Microbiology*, vol. 61, no. 2, pp. 511–525, 2006.

[115] B. Chaban, S. Voisin, J. Kelly, S. M. Logan, and K. F. Jarrell, "Identification of genes involved in the biosynthesis and attachment of *Methanococcus voltae* N-linked glycans: insight into N-linked glycosylation pathways in Archaea," *Molecular Microbiology*, vol. 61, no. 1, pp. 259–268, 2006.

[116] D. J. Vandyke, J. Wu, S. M. Logan et al., "Identification of genes involved in the assembly and attachment of a novel flagellin N-linked tetrasaccharide important for motility in the archaeon *Methanococcus maripaludis*," *Molecular Microbiology*, vol. 72, no. 3, pp. 633–644, 2009.

[117] D. J. VanDyke, J. Wu, S. Y. M. Ng et al., "Identification of a putative acetyltransferase gene, MMP0350, which affects proper assembly of both flagella and pili in the archaeon *Methanococcus maripaludis*," *Journal of Bacteriology*, vol. 190, no. 15, pp. 5300–5307, 2008.

[118] C. M. Szymanski and B. W. Wren, "Protein glycosylation in bacterial mucosal pathogens," *Nature Reviews Microbiology*, vol. 3, no. 3, pp. 225–237, 2005.

[119] M. Abu-Qarn, A. Giordano, F. Battaglia et al., "Identification of AglE, a second glycosyltransferase involved in N glycosylation of the *Haloferax volcanii* S-layer glycoprotein," *Journal of Bacteriology*, vol. 190, no. 9, pp. 3140–3146, 2008.

[120] S. Yurist-Doutsch, M. Abu-Qarn, F. Battaglia et al., "AglF, aglG and aglI, novel members of a gene island involved in the N-glycosylation of the *Haloferax volcanii* S-layer glycoprotein," *Molecular Microbiology*, vol. 69, no. 5, pp. 1234–1245, 2008.

[121] S. Yurist-Doutsch and J. Eichler, "Manual annotation, transcriptional analysis, and protein expression studies reveal novel genes in the agl cluster responsible for N glycosylation in the halophilic archaeon *Haloferax volcanii*," *Journal of Bacteriology*, vol. 191, no. 9, pp. 3068–3075, 2009.

[122] S. Yurist-Doutsch, H. Magidovich, V. V. Ventura, P. G. Hitchen, A. Dell, and J. Eichler, "N-glycosylation in Archaea: on the coordinated actions of *Haloferax volcanii* AglF and AglM," *Molecular Microbiology*, vol. 75, no. 4, pp. 1047–1058, 2010.

[123] M. Abu-Qarn, S. Yurist-Doutsch, A. Giordano et al., "*Haloferax volcanii* AglB and AglD are involved in N-glycosylation of the S-layer glycoprotein and proper assembly of the surface layer," *Journal of Molecular Biology*, vol. 374, no. 5, pp. 1224–1236, 2007.

[124] H. Magidovich, S. Yurist-Doutsch, Z. Konrad et al., "AglP is a S-adenosyl-L-methionine-dependent methyltransferase that participates in the N-glycosylation pathway of *Haloferax volcanii*," *Molecular Microbiology*, vol. 76, no. 1, pp. 190–199, 2010.

[125] B. Chaban, S. M. Logan, J. F. Kelly, and K. F. Jarrell, "AglC and AglK are involved in biosynthesis and attachment of diacetylated glucuronic acid to the N-glycan in *Methanococcus voltae*," *Journal of Bacteriology*, vol. 91, no. 1, pp. 187–195, 2009.

[126] H. Shams-Eldin, B. Chaban, S. Niehus, R. T. Schwarz, and K. F. Jarrell, "Identification of the archaeal *alg*7 gene homolog (encoding N-acetylglucosamine-1-phosphate transferase) of the N-linked glycosylation system by cross-domain complementation in *Saccharomyces cerevisiae*," *Journal of Bacteriology*, vol. 190, no. 6, pp. 2217–2220, 2008.

[127] B. C. Moore and J. A. Leigh, "Markerless mutagenesis in *Methanococcus maripaludis* demonstrates roles for alanine dehydrogenase, alanine racemase, and alanine permease," *Journal of Bacteriology*, vol. 187, no. 3, pp. 972–979, 2005.

[128] S. C. Namboori and D. E. Graham, "Acetamido sugar biosynthesis in the Euryarchaea," *Journal of Bacteriology*, vol. 190, no. 8, pp. 2987–2996, 2008.

[129] M. Sumper, E. Berg, R. Mengele, and I. Strobel, "Primary structure and glycosylation of the S-layer protein of *Haloferax volcanii*," *Journal of Bacteriology*, vol. 172, no. 12, pp. 7111–7118, 1990.

[130] A. Kikuchi, H. Sagami, and K. Ogura, "Evidence for covalent attachment of diphytanylglyceryl phosphate to the cell-surface glycoprotein of *Halobacterium halobium*," *The Journal of Biological Chemistry*, vol. 274, no. 25, pp. 18011–18016, 1999.

[131] H. Wakai, S. Nakamura, H. Kawasaki et al., "Cloning and sequencing of the gene encoding the cell surface glycoprotein of *Haloarcula japonica* strain TR-1," *Extremophiles*, vol. 1, no. 1, pp. 29–35, 1997.

[132] Z. Konrad and J. Eichler, "Lipid modification of proteins in Archaea: attachment of a mevalonic acid-based lipid moiety to the surface-layer glycoprotein of *Haloferax volcanii* follows protein translocation," *Biochemical Journal*, vol. 366, no. 3, pp. 959–964, 2002.

[133] J. Eichler, "Post-translational modification of the S-layer glycoprotein occurs following translocation across the plasma membrane of the haloarchaeon *Haloferax volcanii*," *European Journal of Biochemistry*, vol. 268, no. 15, pp. 4366–4373, 2001.

[134] D. P. Bayley, M. L. Kalmokoff, and K. F. Jarrell, "Effect of bacitracin on flagellar assembly and presumed glycosylation of the flagellins of *Methanococcus deltae*," *Archives of Microbiology*, vol. 160, no. 3, pp. 179–185, 1993.

[135] M. Falb, M. Aivaliotis, C. Garcia-Rizo et al., "Archaeal N-terminal protein maturation commonly involves N-terminal acetylation: a large-scale proteomics survey," *Journal of Molecular Biology*, vol. 362, no. 5, pp. 915–924, 2006.

[136] M. Aivaliotis, K. Gevaert, M. Falb et al., "Large-scale identification of N-terminal peptides in the halophilic Archaea *Halobacterium salinarum* and *Natronomonas pharaonis*," *Journal of Proteome Research*, vol. 6, no. 6, pp. 2195–2204, 2007.

[137] N. F. W. Saunders, C. Ng, M. Raftery, M. Guilhaus, A. Goodchild, and R. Cavicchioli, "Proteomic and computational analysis of secreted proteins with type I signal peptides from the antarctic archaeon *Methanococcoides burtonii*," *Journal of Proteome Research*, vol. 5, no. 9, pp. 2457–2464, 2006.

[138] P. A. Kirkland, M. A. Gil, I. M. Karadzic, and J. A. Maupin-Furlow, "Genetic and proteomic analyses of a proteasome-activating nucleotidase a mutant of the haloarchaeon *Haloferax volcanii*," *Journal of Bacteriology*, vol. 190, no. 1, pp. 193–205, 2008.

[139] D. R. Francoleon, P. Boontheung, Y. Yang et al., "S-layer, surface-accessible, and concanavalin a binding proteins of *Methanosarcina acetivorans* and *Methanosarcina mazei*," *Journal of Proteome Research*, vol. 8, no. 4, pp. 1972–1982, 2009.

Hindawi Publishing Corporation
Archaea
Volume 2010, Article ID 754101, 10 pages
doi:10.1155/2010/754101

Research Article

The S-Layer Glycoprotein of the Crenarchaeote *Sulfolobus acidocaldarius* Is Glycosylated at Multiple Sites with Chitobiose-Linked *N*-Glycans

Elham Peyfoon,[1] Benjamin Meyer,[2] Paul G. Hitchen,[1,3] Maria Panico,[1] Howard R. Morris,[1] Stuart M. Haslam,[1] Sonja-Verena Albers,[2] and Anne Dell[1]

[1] *Division of Molecular Biosciences, Faculty of Natural Sciences, Imperial College London, London SW7 2AZ, UK*
[2] *Molecular Biology of Archaea, Max Planck Institute for Terrestrial Microbiology, Karl-von-Frisch-Straße 10, 35043 Marburg, Germany*
[3] *Centre for Integrative Systems Biology, Faculty of Natural Sciences, Imperial College London, London SW7 2AZ, UK*

Correspondence should be addressed to Anne Dell, a.dell@imperial.ac.uk

Received 14 May 2010; Accepted 26 July 2010

Academic Editor: Jerry Eichler

Glycosylation of the S-layer of the crenarchaea *Sulfolobus acidocaldarius* has been investigated using glycoproteomic methodologies. The mature protein is predicted to contain 31 N-glycosylation consensus sites with approximately one third being found in the C-terminal domain spanning residues L_{1004}-Q_{1395}. Since this domain is rich in Lys and Arg and therefore relatively tractable to glycoproteomic analysis, this study has focused on mapping its N-glycosylation. Our analysis identified nine of the 11 consensus sequence sites, and all were found to be glycosylated. This constitutes a remarkably high glycosylation density in the C-terminal domain averaging one site for each stretch of 30–40 residues. Each of the glycosylation sites observed was shown to be modified with a heterogeneous family of glycans, with the largest having a composition $Glc_1Man_2GlcNAc_2$ plus 6-sulfoquinovose (QuiS), consistent with the tribranched hexasaccharide previously reported in the cytochrome $b_{558/566}$ of *S. acidocaldarius*. *S. acidocaldarius* is the only archaeal species whose N-glycans are known to be linked via the chitobiose core disaccharide that characterises the N-linked glycans of *Eukarya*.

1. Introduction

In many *Archaea* the surface layer (S-layer) proteins are the sole cell wall component [1]. These S-layer proteins assemble into a natural 2-D crystal structure with very strong self interactions. In *Archaea*, which do not possess other cell wall components, the S-layer has to maintain the cell integrity and stabilize as well as to protect the cell against mechanical and osmotic stresses or extreme pH conditions. It is also predicted that the S-layer has to maintain or even determine the cell shape [2–6].

In *Sulfolobus* spp. the S-layer is composed of two proteins: a small protein of approximately 45 kD, SlaB, and a large protein, SlaA, of approximately 120 kDa. SlaB is an integral membrane protein and its strong interaction with SlaA, which covers the whole cell surface, tethers the S-layer to the membrane [8, 9]. Taking into account the harsh growth condition of the thermoacidophilic *Sulfolobus* spp. (pH 2-3 and 75–80°C), the S-layer proteins will play an important role in maintaining cell integrity and must be adapted to be functional under these conditions.

One possible posttranslational modification proteins can undergo is glycosylation, which has a major effect on stability and half-life [10]. Indeed, all archaeal S-layer proteins which have been structurally studied to date, have been found to carry N-glycans [11–17].

Although *Eukarya*, *Bacteria*, and *Archaea* all share certain characteristics of the N-glycosylation pathway, the resulting glycan structures in *Bacteria* and *Archaea* are more diverse than in *Eukarya* [18, 19]. Notably a far greater variety of monosaccharides is used, many of which carry functionalities such as sulfate and methyl groups, or even amino acids such as threonine [12]. In the last two years, substantial progress in describing the enzymes involved in

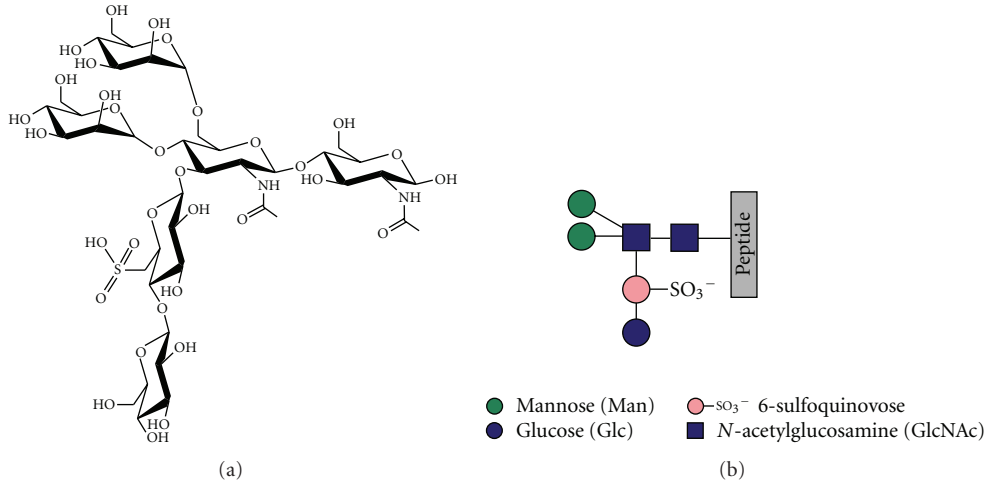


FIGURE 1: (a) Structure of the 6-sulfoquinovose-containing tribranched hexasaccharide identified in cytochrome $b_{558/566}$ [7]. (b) This shows a symbolic representation of the glycan which is used in the annotations of Figures 4–6.

archaeal N-glycosylation pathways has been made [19–23]. The archaeal N-glycosylation machinery combines aspects of both the eukaryal and bacterial pathways. For instance *Archaea* and *Eukarya* use dolichol as the lipid carrier whereas *Bacteria* use undecaprenyl. On the other hand, in *Archaea* and *Bacteria* the oligosaccharyltransferase is comprised of a single subunit whereas in *Eukarya* it is a multimeric complex.

So far proteins from about 25 species of *Archaea* have been reported to be glycosylated and about ten have had their N-glycans partially or fully characterized [19, 24]. The best understood are S-layers of the halophiles *Halobacterium salinarum* and *Haloferax volcanii*, and S-layers and flagellins of the methanogens *Methanothermus fervidus*, *Methanococcus voltae,* and *M. maripaludis*. In contrast to the N-glycans of *Eukarya,* which are almost always branched and usually considerably greater than six residues in size, these archaeal glycoproteins contain unbranched glycans most of which have fewer than six sugar residues. The exception is *Hbt. salinarum* which, in addition to bearing a trisaccharide composed of monosulfated glucuronic acid linked via glucose at about ten consensus sites, has one N-linked site occupied by a GalNAc-linked polysaccharide comprised of multiple repeats of a sulfated pentasaccharide (composed of GlcNAc, GalNAc, Gal, GalA, and 3-O-methyl-GalA) [25]. In *H. volcanii* the S-layer glycoprotein is modified by the attachment of a pentasaccharide, composed of two hexoses, two hexuronic acids, and a methylester of hexuronic acid [26, 27]. The N-glycans attached to the S-layer of *Methanothermus fervidus* are hexasaccharides containing 3-O-methylmannose, mannose, and GalNAc [28]. In *M. voltae* the flagellins and S-layer proteins are glycosylated with a complex trisaccharide composed of GlcNAc, GlcdiNAcA, and a threonine-substituted ManNAcA [12]. A second strain of *M. voltae* has been recently found to carry tetrasaccharides which share this trisaccharide sequence capped by an uncharacterized sugar [29]. An even more unusual tetrasaccharide has been found on the pilins and flagellins of *M. maripaludis*. This has the sequence Sug-4-β-ManNAc3NAmA6Thr-4-β-GlcNAc3NAcA-3-β-GalNAc where Sug is a highly unusual aldolase diglycoside [30].

Glycosylation of extreme thermophile members of the *Archaea* domain is quite poorly understood despite the fact that one member of this class, *Thermoplasma acidophilum*, from the Euryarchaeota kingdom of *Archaea*, was amongst the first of the archaea to have its glycoproteins studied by biophysical methods [31]. This early study, which reported the presence of branched mannose-rich glycans linked via GlcNAc to Asn, has not, however, been followed up with more rigorous structure analysis. A second extremophile member of the Euryarchaeota kingdom to have its glycosylation studied is *Pyrococcus furiosus*, which, interestingly, has also been shown to biosynthesis branched glycans [32]. The oligosaccharyltransferase from this species has been purified, and its ability to glycosylate a fluorescently labeled peptide containing a consensus sequence has been assayed in the presence and absence of lipid-linked oligosaccharide (LLO) prepared from *Pyrococcus furiosus* cells. In the presence of the LLO a glycopeptide was produced which was shown by mass spectrometry to be a branched heptasaccharide having a pentose sugar attached to each of the second and third residues of a pentasaccharide of sequence HexNAc-HexA-Hex-Hex-HexNAc [32].

Branching is also a feature of the only glycan so far determined from a member of the Crenarchaeota kingdom. Thus cytochrome $b_{558/566}$ of *Sulfolobus acidocaldarius*, which grows optimally at 75–80°C and pH 2-3, was shown to be modified with a tribranched hexasaccharide of composition $Glc_1Man_2GlcNAc_2$ plus 6-sulfoquinovose, an unusual sugar which is characteristic of chloroplasts and photosynthetic bacteria (see Figure 1 for structure) [7]). As, to date, this is the only characterised glycan structure from a crenarchaeal species our objective is to determine the glycan composition of other extracellular proteins of the *Sulfolobales*. It is known that nearly all extracellular proteins found in these organisms

are glycosylated [8, 33, 34]. As a first model protein, the S-layer protein of *S. acidocaldarius* was isolated and its glycosylation investigated using glycoproteomic methodologies.

2. Materials and Methods

2.1. Strains and Growth Conditions. S. acidocaldarius (DSM639) was grown in Brock medium at pH 3 and 76°C [35] and the medium was supplemented with 0.1(w/v) % of tryptone as sole carbon and energy source. Growth of cells was monitored by measuring the optical density at 600 nm.

2.2. S-Layer Isolation. Fresh cells or frozen cell pellets from a 50 ml culture were resuspended in 40 ml buffer A (10 mM NaCl, 1 mM PMSF, 0.5% Na-Lauroylsarcosine) with the addition of a little bit of DNAse. The samples were shaken for 45 minutes at 37°C and centrifuged for 20 min in an Optima Max-XU Ultracentrifuge (Beckman Coulter) at 16.000 rcf, yielding a brownish tan pellet. The pellet was resuspended in 1,5 ml buffer A and incubated for 30 min at 37°C. After centrifugation in a tabletop centrifuge at 14.000 rpm the pellet was purified by repeatable washes in buffer B (10 mM NaCl, 0,5 mM $MgSO_4$, 0.5% SDS), incubation for 20 min at 37°C and subsequent centrifugation, until a translucent tan pellet was obtained. Once the pellet was translucent the S-layer proteins were once washed with water ad then stored in water at 4°C.

2.3. Proteolytic Digestion for Glycoproteomic Analysis. Purified S-layer samples of *S. acidocaldarius* were run on a 2–8% precast gel (Invitrogen, Paisley, UK) and stained with Novex Colloidal blue stain (Invitrogen). The S-layer was observed as a broadband between 116 and 160 kDa. The band was then excised and cut into pieces, destained using 400 μl of 50% (v/v) acetonitrile in 0.1 M ammonium bicarbonate (pH 8.4) and dried in a SpeedVac. Reduction/carboxymethylation was carried out by swelling and incubating the dried gel pieced in Dithiothreitol (10 mM) (200 μl) in ammonium bicarbonate (AMBIC) (50 mM, pH 8.4) (Roche, West Sussex, UK) at 56°C for 30 min. The DTT solution was then removed and the gel pieces were washed with acetonitrile (200 μl) and dried. The gel pieces were incubated in dark at room temperature in 50 mM iodoacetic acid (IAA) (200 μl) (Sigma-Aldrich Dorset, UK) which was dissolved in ammonium bicarbonate (50 mM, pH 8.4). The IAA was then removed and gel pieces were washed in AMBIC buffer (500 μl) for 15 min. The AMBIC was removed and gel pieces were shrunk in acetonitrile (200 μl) for 5 min. The gel pieces were dried in SpeedVac and then subjected to digestion with trypsin or chymotrypsin. For the tryptic digest the dried gel pieces were reswelled in AMBIC solution and incubated at 37°C with 25ng/μl trypsin (20 μl) (Promega cat V5111) overnight. The supernatant was removed and placed in a clean eppendorf. The gel pieces were then incubated in 0.1% TFA (50 μl) at 37°C for 10 min. Acetonitrile (100 μl) was added to the mixture which was incubated at 37°C for 15 min. The supernatant was then pooled with the previous supernatant and the process was repeated twice. The supernatant volume was then reduced (to about 35 μl) in preparation for LC-MS. For the chymotryptic digest re-swelled gel pieces were incubated at 37°C in 25 ng/μl chymotrypsin (Sigma C-3142) dissolved in Tris-HCl (100 mM, pH 7.8), overnight. The remainder of the experiment was carried out as for tryptic digestion.

2.4. LC-MS Analysis. The extracted peptides/glycopeptides from the gel pieces were analyzed a nano-LC-ES-MS/MS employing a quadrupole TOF mass spectrometer (Q-STAR Pulsar I, MDS Sciex). Separation of the peptides/glycopeptides was carried out by using a nano-LC gradient method generated by an Ultimate pump fitted with a Famos autosampler and a Switchos microcolumn switching module (LC Packings, Amsterdam, The Netherlands). The system was coupled to an analytical C_{18} nanocapillary (75 m inside diameter × 15 cm, PepMap) and a microprecolumn C_{18} cartridge for online peptide/glycopeptide separation. The digest was first loaded onto the precolumn and eluted with 0.1% formic acid (Sigma) in water (HPLC grade, Purite) for 4 min. The eluant was then transferred onto the column and eluted at a flow rate of 150 nL/min using the following gradient of solvent A [0.05% (v/v) formic acid in a 95:5 (v/v) water/acetonitrile mixture] and solvent B [0.04% formic acid in a 95:5 (v/v) acetonitrile/water mixture]: 99% A from 0 to 5 min, 99 to 90% A from 5 to 10 min, 90 to 60% A from 10 to 70 min, 60 to 50% A from 70 to 71 min, 50 to 5% A from 71 to 75 min, 5% A from 75 to 85 min, 5 to 95% A from 85 to 86 min, and 95% A from 86 to 90 min. Data acquisition was performed using Analyst QS software with an automatic information-dependent-acquisition (IDA) function.

2.5. Sugar Composition Analysis. Samples were hydrolysed in 1 M methanolic hydrogen chloride at 80°C for 16 h and the reagent was removed under a stream of nitrogen. Hexosamines were re-*N*-acetylated in 500 μl of methanol/pyridine/acetic anhydride (500:1:5, v/v/v) for 15 min at room temperature, then dried under nitrogen. Trimethylsilyl derivatisation was performed in 200 μL of Tri-Sil "Z" (Pierce) at room temperature for 30 min, after which the reagent was removed under nitrogen. Derivatized monosaccharides were resuspended in 1 ml of hexanes, centrifuged at 3000 rpm for 10 min, and the supernatant transferred and dried under nitrogen for analysis by gas chromatography-mass spectrometry (GC-MS).

2.6. GC-MS Analysis. This was carried out using a Perkin Elmer Clarus 500 instrument, fitted with a RTX-5 (30 m × 0.25 mm internal diameter, Restek Corp.). Temperature program: the oven was held at 65°C for 1 min before being increased to 140°C at a rate of 25°C/min, then to 200°C at a rate of 5°C/min and finally to a temperature of 300°C at a rate of 10°C/min.

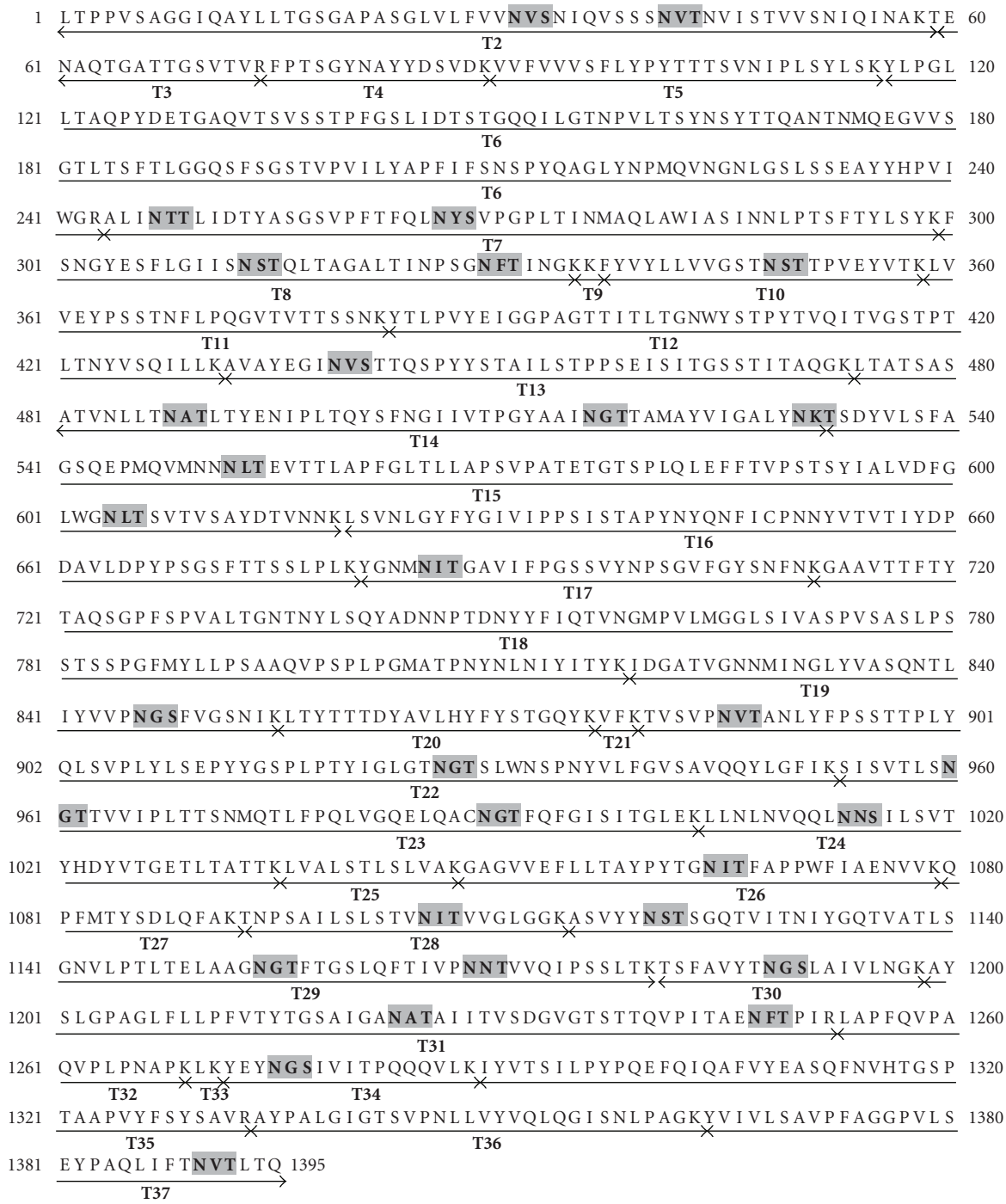


Figure 2: Polypeptide sequence of the *S. acidocaldarius* S-layer. The N-terminal signal sequence has been omitted. Consensus sequences for N-glycosylation are shaded and the predicted products of tryptic digestion are shown by underlining.

3. Results

3.1. Strategy for Glycoproteomic Analysis of the S. acidocaldarius S-Layer. The polypeptide sequence of the *S. acidocaldarius* S-layer SlaA protein (Saci_2355) is shown in Figure 2 [36]. The mature protein is predicted to comprise 1,395 amino acids and to contain 31 consensus sites for N-glycosylation. These are scattered throughout the S-layer with the greatest density being in the C-terminal domain where about 25% of the polypeptide contains one third of the consensus sequences. This domain is rich in Lys and Arg, in contrast to the remainder of the S-layer where these residues are quite sparse. As shown in Figure 2, the majority of the predicted tryptic peptides from Leu_{1003} onwards are smaller than about 40 residues, and are thus well suited to electrospray tandem mass spectrometry (ES-MS/MS) whereas many of those in

Table 1: Glycopeptides observed in LC-MS/MS analyses of proteolytic digests of the S-layer from *S. acidocaldarius*. Tryptic glycopeptides are designated T24, T26 and so forth based on the *in silico* digest shown in Figure 2. Two chymotryptic glycopeptides derived from the C-terminal domain of the S-layer are designated C-1 and C-2.

Glycopeptides designation	Peptide Residues	Theoretical Peptide $[M + H]^+$	Observed glycopeptide (m/z)	Deduced glycopeptide $[M + H]^+$	Glycan Modification
T24	1004–1035	3566.00	$[1561.5]^{3+}$	4682.5	$Hex_3HexNAc_2Qui\ (SO_3^-)$
			$[1507.5]^{3+}$	4520.5	$Hex_2HexNAc_2Qui\ (SO_3^-)$
			$[1377.7]^{3+}$	4131.2	$Hex_1HexNAc_2$
			$[1256.4]^{3+}$	3767.2	$HexNAc_1$
T26	1048–1079	3484.82	$[1151.6]^{4+}$	4603.4	$Hex_3HexNAc_2Qui\ (SO_3^-)$
			$[1111.3]^{4+}$	4442.2	$Hex_2HexNAc_2Qui\ (SO_3^-)$
			$[1014.1]^{4+}$	4053.0	$Hex_1HexNAc_2$
T28	1093–1114	2141.22	$[1087.3]^{3+}$	3259.9	$Hex_3HexNAc_2Qui\ (SO_3^-)$
			$[1033.2]^{3+}$	3097.7	$Hex_2HexNAc_2Qui\ (SO_3^-)$
			$[904.0]^{3+}$	2710.0	$Hex_1HexNAc_2$
T30	1181–1198	1855.00	$[992.2]^{3+}$	2974.6	$Hex_3HexNAc_2Qui\ (SO_3^-)$
			$[938.1]^{3+}$	2811.0	$Hex_2HexNAc_2Qui\ (SO_3^-)$
			$[809.0]^{3+}$	2425.0	$Hex_1HexNAc_2$
T31	1228–1281	5445.19	$[1537.0]^{5+}$	7681.0	$[Hex_3HexNAc_2Qui\ (SO_3^-)]_2$
T34	1272–1288	1980.04	$[1033.5]^{3+}$	3098.6	$Hex_3HexNAc_2Qui\ (SO_3^-)$
			$[979.5]^{3+}$	2936.5	$Hex_2HexNAc_2Qui\ (SO_3^-)$
			$[904.2]^{3+}$	2709.5	$Hex_2HexNAc_2$
			$[850.1]^{3+}$	2548.4	$Hex_1HexNAc_2$
T37	1364–1395	3407.85	$[1509.7]^{3+}$	4526.3	$Hex_3HexNAc_2Qui\ (SO_3^-)$
			$[1325.7]^{3+}$	3976.2	$Hex_1HexNAc_2$
C-1	1374–1395	2320.09	$[1146.6]^{3+}$	3437.8	$Hex_3HexNAc_2Qui\ (SO_3^-)$
C-2	1164–1178	1582.71	$[1350.6]^{2+}$	2700.2	$Hex_3HexNAc_2Qui\ (SO_3^-)$
			$[1075.6]^{2+}$	2150.2	$Hex_1HexNAc_2$

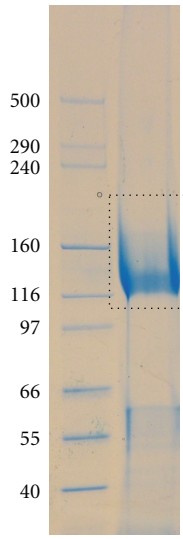


Figure 3: SDS PAGE gel of the *S. acidocaldarius* S-layer showing the band that was cut out for in-gel digestion.

the N-terminal domain are much larger and are therefore expected to be far less tractable to proteomic analysis. We decided, therefore, to focus our efforts on defining glycosylation in the C-terminal domain by performing nano-LC-MS/MS analyses of in-gel tryptic digests (Figure 3) of the S-layer and manually searching the resulting MS/MS data for potential C-terminal glycopeptides. First, we identified spectra containing fragment ions suggesting the presence of sugars. Then, promising MS/MS data were examined for the presence of peptide sequence ions that could be attributed to predicted tryptic peptides in the C-terminal domain. Finally, glycopeptide structures were deduced taking into account likely peptide and glycan compositions, the latter assignment being assisted by knowledge of the glycan on the cytochrome $b_{558/566}$ of *S. acidocaldarius* (Figure 1). Additionally, all the mass spectra acquired in, and adjacent to, the elution windows of the identified glycopeptides were examined for evidence of molecular ions consistent with glycoforms whose abundance and/or m/z values had precluded their selection for MS/MS analysis by the automatic software. The results of these analyses are presented in the following sections.

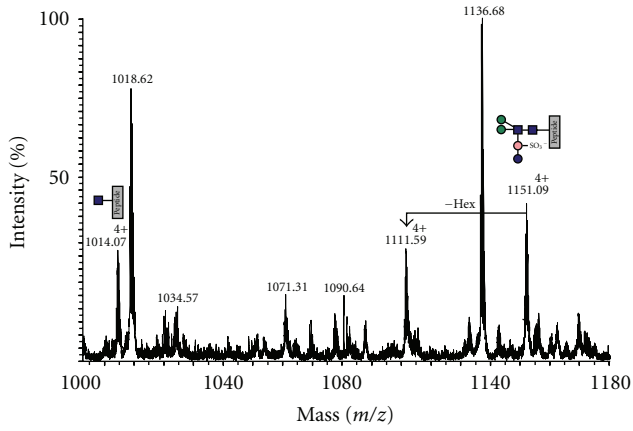


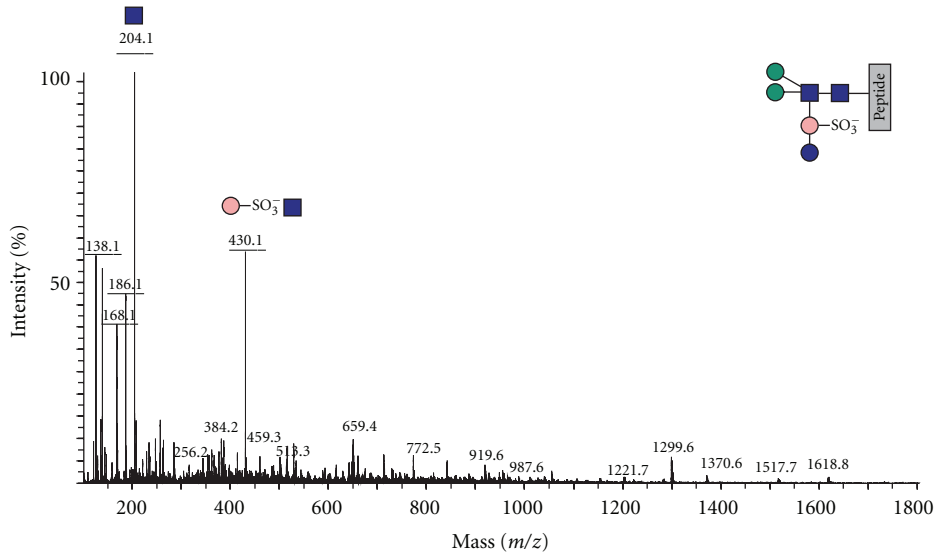
Figure 4: Summed mass spectra recorded between 46 and 52 min from an on-line nanoLC-ES-MS/MS analysis of an in-gel tryptic digest of the S-layer. The quadruply charged signals at *m/z* 1014.07, 1111.59 and 1151.09 are molecular ions of glycopeptides of sequence GAGVVEFLLTAYPYTGNITFAPPWFIAENVVK carrying the glycans shown in the annotations. Unassigned signals are molecular ions of peptides from elsewhere in the S-layer.

3.2. Evidence That T-26 and T-24 Are Glycosylated. Following on-line nano-LC-ES-MS analysis of the S-layer tryptic digest and manual interpretation of the resulting data, a number of multiply charged molecular ions were observed whose product ion spectra were indicative of peptide glycosylation. Once recognised, related glycoforms were identified by summation across the appropriate nano-LC elution time. The summed mass spectrum of components eluting between 46 and 52 min is shown in Figure 4. The signals labelled with their charge states (4+) are glycopeptides corresponding to the tryptic peptide T-26 of sequence GAGVVEFLLTAYPYTGNITFAPPWFIAENVVK (see Figure 2 for designation of tryptic peptides). The $[M + 4H]^{4+}$ signals at m/z 1151.09 and 1014.07 were automatically selected for MS/MS, and the resulting spectra are shown in Figure 5. Both are rich in peptide fragment ions that confirm the sequence. Both also are dominated by sugar fragment ions in the low mass region which are diagnostic for HexNAc (m/z 138, 168, 186, and 204) and for HexNAc (204) plus a 226 increment (m/z 430). The latter increment is identical to the mass of the unusual sugar in the *S. acidocaldarius* cytochrome $b_{558/566}$ glycan (Qui-sulfonate, Figure 1). The molecular weights of the quadruply charged signals at m/z 1151.09 and 1014.07 (4600.36 Da and 4052.28 Da) correspond to the molecular weights of the peptide (3483.82 Da) plus a $Hex_3HexNAc_2QuiS$ hexasaccharide and a $Hex_1HexNAc_2$ trisaccharide, respectively. The former is consistent with the glycan composition characterised previously for *S. acidocaldarius* cytochrome $b_{558/566}$ (Figure 1). The signal at m/z 1111.59 was not selected for MS/MS but its m/z value is consistent with this component having one fewer hexose than m/z 1151.09 (Figure 4). Without MS/MS data, we are not able to determine whether Glc or Man are absent from the glycan. It is interesting that we do not observe a molecular ion corresponding to $Hex_2HexNAc_2$.

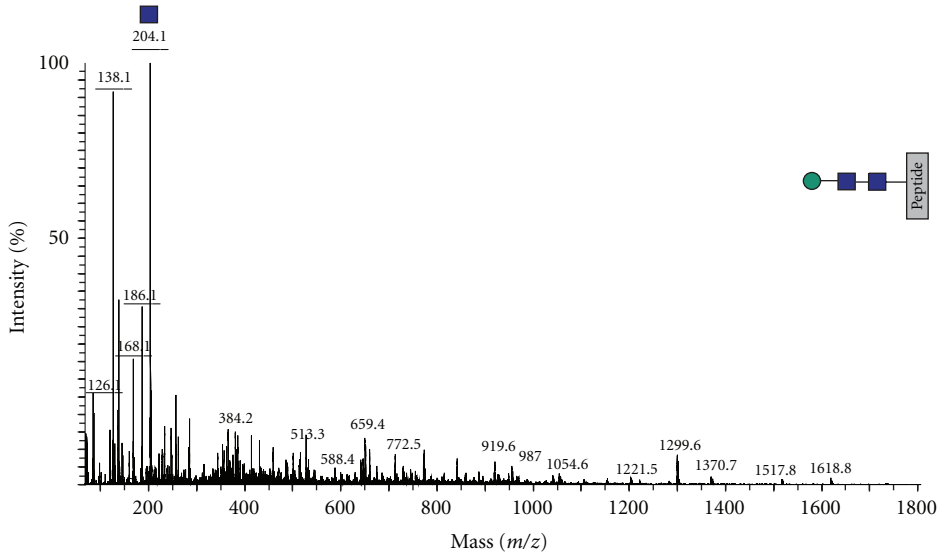
We observed similar patterns for all the other tryptic glycopeptides, although there was some variation in relative abundances, particularly of minor components. Some sites had significant amounts of glycans that were truncated to a single GlcNAc. An example is shown in Figure 6, which is the tryptic glycopeptide T-24 having the sequence LLNLNVQQLNNSILSVTYHDYVTGETLTATTK. Triply charged $[M + 3H]^{3+}$ molecular ions are observed at m/z 1256.38, 1378.09, 1507.44 and 1561.45 corresponding to glycans of composition HexNAc, $Hex_1HexNAc_2$, $Hex_2HexNAc_2QuiS$, and $Hex_3HexNAc_2QuiS$, respectively. The MS/MS data confirmed the identity of the peptide T-24 (data not shown). For this glycopeptide a relatively abundant ion for the glycopeptides truncated to a single HexNAc was observed (m/z 1256.4). This was not a significant glycoform in the case of T-26.

3.3. All Observed Consensus Sequences in the C-Terminal Domain Are Glycosylated. Using the same logic as applied to T-26 and T-24 (see above) we unambiguously identified all the remaining tryptic peptides in the C-terminal domain with the exception of T-29. This is a 66 residue peptide having three consensus sequences (A_{1115}SVYY...SSLTK$_{1180}$) and is likely to be too large for the ES-MS experiment. The identified glycopeptides are shown in Table 1 which summarises the m/z values and compositions for glycoforms observed by ES-MS. In an attempt to identify the consensus sites falling within the T-29 tryptic glycopeptide we carried out a chymotryptic digestion. This is not an ideal enzyme for glycoproteomics because it cleaves relatively nonspecifically at large hydrophobic residues as well as at aromatic residues. Hence it yields a very complex digest and consequently the MS and MS/MS data are an enormous challenge for manual interpretation. Fortunately, however, two useful sets of glycopeptide data from the C-terminal domain of the S-layer were revealed by manual inspection. Firstly high quality spectra were found for the C-terminus itself (AGGPVLSEYPAQLIFTNVTLTQ; designated C-1 in Table 1), which includes the consensus sequence at Asn_{1390}. MS/MS results (not shown) confirmed the major glycoform was the same as that assigned in the tryptic digest (Table 1) where our evidence had been confined to MS data only. More usefully the chymotryptic digest gave data corresponding to a glycopeptide of sequence TIVPNNTVVQIPSSL which spans the consensus site at Asn_{1168} within T-29 (designated C-2). Once again the glycan profile was similar to other observed sites with the hexa- and trisaccharides being the most abundant components (Table 1).

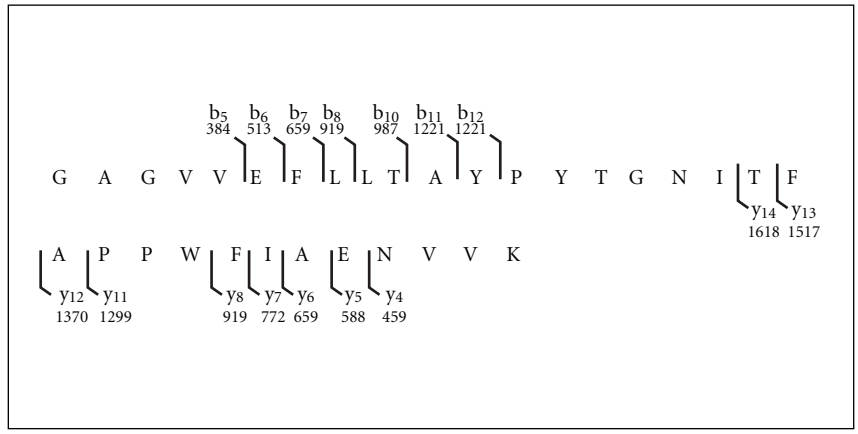
3.4. Sugar Analysis Confirms Man, Glc, and GlcNAc Content. The *S. acidocaldarius* S-layer was analysed for its sugar composition by GC-MS of trimethylsilyl (TMS) methyl glycoside derivatives and the data are shown in Table 2. Consistent with both the MS data described earlier, and the structure reported by Zähringer and colleagues [7], the analysis showed Man, Glc, and GlcNAc as the only observable sugars. No GalNAc was present, confirming the chitobiose core sequence. The Man:Glc ratio is consistent with the



(a)



(b)



(c)

Figure 5: MS/MS of m/z 1151.09 (a) and m/z 1014.07 (b) (see mass spectrum in Figure 4). Note the diagnostic fragment ions for HexNAc in both spectra and for HexNAc-QuiS in (a). The peptide sequence ions are identical in (a) and (b), and their assignments are shown in (c).

Table 2: GC-MS analysis of TMS methyl glycoside sugar derivatives obtained from *S. acidocaldarius* S-layer glycoprotein. Note that for experimental reasons GlcNAc recoveries are always poor in sugar analysis experiments. We consider it likely that this is the reason for the GlcNAc:Mannose ratio being much lower than expected from the LC-ES/MS/MS data, although we cannot rule out the possibility that other mannose-containing polymers are present in the sample.

Elution time (min)	Characteristic fragment ions	Assignment	Relative amounts
12.85	204, 217, 305, 361	Man	1
14.62	204, 217, 305, 361	Glc	0.38
18.04	173, 259, 314	GlcNAc	0.22

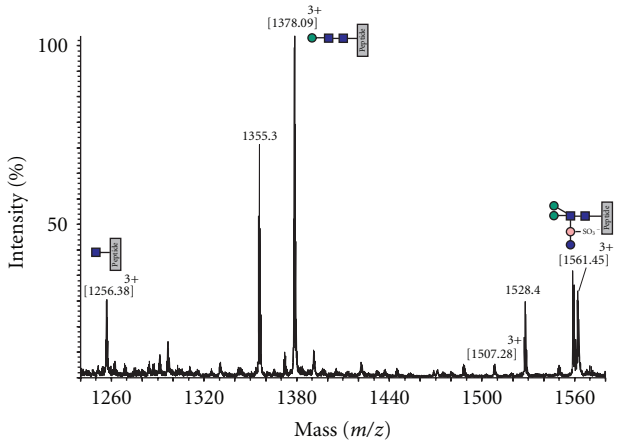


Figure 6: Summed mass spectra of T24 recorded between 48.5 and 50.6 min from an on-line nanoLC-ES-MS/MS analysis of an in-gel tryptic digest of the S-layer. The triply charged signals at m/z 1256.38, 1378.09, 1507.28, and 1561.45 are molecular ions of glycopeptides of sequence LLNLNVQQLNNSILSVTYHDYVTGETLTATTK carrying the glycans shown in the annotations. Unassigned signals are molecular ions of peptides from elsewhere in the S-layer.

mixture of glycans that we found in the glycoproteomics experiments. The unusual quinovose sugar was not observed because sulfonated sugars are not recovered under the experimental conditions used for sugar analysis.

4. Discussion

In this study we have employed glycoproteomic methodologies to map the C-terminal domain of the *S. acidocaldarius* S-layer, spanning residues L_{1004}-Q_{1395}. This domain has eleven potential N-linked glycosylation sites, nine of which we have shown to be glycosylated (Figure 7), including Asn_{1390} which is only six residues from the C-terminus. The two sites which were not identified, Asn_{1120} and Asn_{1154}, fall within a 66 residue tryptic peptide (T-29) that has three consensus sequences, the third of which (Asn_{1168}) was found to be occupied via analysis of a chymotryptic digest of the S-layer. Unfortunately the other two sites were not revealed by this digest. It is noteworthy that we were successful in obtaining good quality data on T-31, which is a 54 residue tryptic glycopeptide carrying two N-glycans (see Figure 7 and Table 1) whose size is not very different from the predicted value for monoglycosylated T-29. We think it is likely, therefore, that our failure to detect signals attributable to the T-29 glycopeptide is probably due to either or both Asn_{1120} and Asn_{1154} being glycosylated, in addition to Asn_{1168}, thus moving this glycopeptide outside the observable m/z range of the glycoproteomics experiment. Irrespective of whether these two sites are indeed occupied, the glycosylation density in the C-terminal domain is quite remarkable, averaging one glycosylation site for each stretch of 30–40 residues.

Each of the observed glycosylation sites was found to be heterogeneously glycosylated with a family of glycans, the largest of which has a composition consistent with it having the sequence of the tribranched hexasaccharide found in cytochrome $b_{558/566}$ of *S. acidocaldarius* (Figure 1, [7]). The other members of the family appear to be biosynthetic precursors of this glycan. The most abundant is $Man_1GlcNAc_2$ (see annotations on Figures 4 and 6). In addition, a nonextended GlcNAc is a minor component at some sites (Figure 6), and a pentasaccharide lacking one of the hexoses of the mature glycan was also observed (Figures 4 and 6). Interestingly $Man_2GlcNAc_2$ was only observed as a very minor component at a single site (T-34, Table 1), suggesting that either the second mannose is added after the 6-sulfoquinovose, or addition of the latter is rapid compared with the second mannosylation. It is not surprising that the S-layer and the cytochrome share a glycan sequence because conservation of N-glycosylation within a species appears to be characteristic of *Archaea* [24]. Moreover, the presence of a family of biosynthetically related glycans is not unexpected since the S-layer of *H. volcani* exhibits a similar phenomenon [27].

With the exception of the two halophiles described in the Introduction, whose glycans are linked via glucose, all known archaeal glycans are attached to Asn via GalNAc or GlcNAc. Interestingly, *S. acidocaldarius* is the only species characterized so far whose glycans are linked via chitobiose (GlcNAcβ1-4GlcNAc), the core disaccharide shared by the N-linked glycans of *Eukarya*. Moreover the tribranched topology of the *Sulfolobus* glycan is reminiscent of eukaryotic glycans which are usually multiantennary.

Sulfolobus spp. are developing into model organisms for studies on *Eukarya*-like mechanisms of transcription, translation, and cell division, and the vast amount of recently established genetic tools now make these organisms a prime choice to study these phenomena. The glycan

961 GTTVVIPLTTSNMQTLFPQLVGQELQACNGTFQFGISITGLEKLLNLNVQQL**NNS**ILSVT 1020

1021 YHDYVTGETLTATTKLVALSTLSLVAKGAGVVEFLLTAYPYTG**NIT**FAPPWFIAENVVKQ 1080

1081 PFMTYSDLQFAKTNPSAILSLSTV**NIT**VVGLGGKASVYY**NST**SGQTVITNIYGQTVATLS 1140

1141 GNVLPTLTELAAG**NGT**FTGSLQFTIVP**NNT**VVQIPSSLTKTSFAVYT**NGS**LAIVLNGKAY 1200

1201 SLGPAGLFLLPFVTYTGSAIGA**NAT**AIITVSDGVGTSTTQVPITAE**NFT**PIRLAPFQVPA 1260

1261 QVPLPNAPKLKYEY**NGS**IVITPQQQVLKIYVTSILPYPQEFQIQAFVYEASQFNVHTGSP 1320

1321 TAAPVYFSYSAVRAYPALGIGTSVPNLLVYVQLQGISNLPAGKYVIVLSAVPFAGGPVLS 1380

1381 EYPAQLIFT**NVT**LTQ 1395

Figure 7: Sequence of the C-terminal domain of the S-layer which was mapped in the glycoproteomics experiments. Consensus sites (in bold) which were shown to be glycosylated are underlined.

structural information reported in this paper will facilitate the application of genetic tools to the elucidation of the N-glycosylation pathway in *S. acidocaldarius*. It will be very interesting to establish whether the commonalities in core structure between the glycans of *S. acidocaldarius* and those of *Eukarya* are mirrored in the biosynthetic pathways. Moreover, identification of the glycosylation enzymes of *S. acidocaldarius* could lead to interesting biotechnological applications.

Acknowledgments

B. Meyer and S.-V. Albers were supported by a VIDI grant of the Dutch Science foundation (NWO) and S.-V. Albers received intramural funds from the Max Planck Society. The Imperial College authors were supported by the Biotechnology and Biological Sciences Research Council (BBF0083091, BBC5196701) and the UK Research Councils' Basic Technology Initiative Translational Grant (EP/G037604/1).

References

[1] H. Koenig, "Archaeobacterial cell envelopes," *Canadian Journal of Microbiology*, vol. 34, pp. 395–406, 1988.

[2] M. Sára and U. B. Sleytr, "S-layer proteins," *Journal of Bacteriology*, vol. 182, no. 4, pp. 859–868, 2000.

[3] H. Engelhardt and J. Peters, "Structural research on surface layers: a focus on stability, surface layer homology domains, and surface layer-cell wall interactions," *Journal of Structural Biology*, vol. 124, no. 2-3, pp. 276–302, 1998.

[4] H. Engelhardt, "Are S-layers exoskeletons? The basic function of protein surface layers revisited," *Journal of Structural Biology*, vol. 160, no. 2, pp. 115–124, 2007.

[5] H. Claus, E. Aka, T. Debaerdemaeker, C. Evrard, J.-P. Declercq, and H. König, "Primary structure of selected archaeal mesophilic and extremely thermophilic outer surface layer proteins," *Systematic and Applied Microbiology*, vol. 25, no. 1, pp. 3–12, 2002.

[6] T. J. Beveridge, P. H. Pouwels, M. Sára et al., "Functions of S-layers," *FEMS Microbiology Reviews*, vol. 20, no. 1-2, pp. 99–149, 1997.

[7] U. Zähringer, H. Moll, T. Hettmann, Y. A. Knirel, and G. Schäfer, "Cytochrome $b_{558/566}$ from the archaeon sulfolobus racidocaldarius has a unique Asn-linked highly branched hexasaccharide chain cotaining 6- sulfoquinovose," *European Journal of Biochemistry*, vol. 267, no. 13, pp. 4144–4149, 2000.

[8] D. W. Grogan, "Organization and interactions of cell envelope proteins of the extreme thermoacidophile Sulfolobus acidocaldarius," *Canadian Journal of Microbiology*, vol. 42, no. 11, pp. 1163–1171, 1996.

[9] A. Veith, A. Klingl, B. Zolghadr et al., "Acidianus, Sulfolobus and Metallosphaera surface layers: structure, composition and gene expression," *Molecular Microbiology*, vol. 73, no. 1, pp. 58–72, 2009.

[10] A. Varki, "Biological roles of oligosaccharides: all of the theories are correct," *Glycobiology*, vol. 3, no. 2, pp. 97–130, 1993.

[11] R. Zeitler, E. Hochmuth, R. Deutzmann, and M. Sumper, "Exchange of Ser-4 for Val, Leu or Asn in the sequon Asn-Ala-Ser does not prevent N-glycosylation of the cell surface glycoprotein from Halobacterium halobium," *Glycobiology*, vol. 8, no. 12, pp. 1157–1164, 1998.

[12] S. Voisin, R. S. Houliston, J. Kelly et al., "Identification and characterization of the unique N-linked glycan common to the flagellins and S-layer glycoprotein of Methanococcus voltae," *The Journal of Biological Chemistry*, vol. 280, no. 17, pp. 16586–16593, 2005.

[13] M. Sumper, E. Berg, R. Mengele, and I. Strobel, "Primary structure and glycosylation of the S-layer protein of Haloferax volcanii," *Journal of Bacteriology*, vol. 172, no. 12, pp. 7111–7118, 1990.

[14] G. Paul, F. Lottspeich, and F. Wieland, "Asparaginyl-N-acetylgalactosamine. Linkage unit of halobacterial glycosaminoglycan," *Journal of Biological Chemistry*, vol. 261, no. 3, pp. 1020–1024, 1986.

[15] M. F. Mescher and J. L. Strominger, "Purification and characterization of a prokaryotic glycoprotein from the cell envelope of Halobacterium salinarium," *Journal of Biological Chemistry*, vol. 251, no. 7, pp. 2005–2014, 1976.

[16] M. Kessel, "Three-dimensional structure of the regular surface glycoprotein layer of Halobacterium volcanii from the Dead Sea," *The EMBO Journal*, vol. 7, no. 5, pp. 1549–1554, 1988.

[17] D. R. Francoleon, P. Boontheung, Y. Yang et al., "S-layer, surface-accessible, and concanavalin a binding proteins of Methanosarcina acetivorans and methanosarcina mazei," *Journal of Proteome Research*, vol. 8, no. 4, pp. 1972–1982, 2009.

[18] C. M. Szymanski and B. W. Wren, "Protein glycosylation in bacterial mucosal pathogens," *Nature Reviews Microbiology*, vol. 3, no. 3, pp. 225–237, 2005.

[19] M. Abu-Qarn, J. Eichler, and N. Sharon, "Not just for Eukarya anymore: protein glycosylation in Bacteria and Archaea," *Current Opinion in Structural Biology*, vol. 18, no. 5, pp. 544–550, 2008.

[20] E. Weerapana and B. Imperiali, "Asparagine-linked protein glycosylation: from eukaryotic to prokaryotic systems," *Glycobiology*, vol. 16, no. 6, pp. 91R–101R, 2006.

[21] S. Yurist-Doutsch, H. Magidovich, V. V. Ventura, P. G. Hitchen, A. Dell, and J. Eichler, "N-glycosylation in Archaea: on the coordinated actions of Haloferax volcanii AglF and AglM," *Molecular Microbiology*, vol. 75, no. 4, pp. 1047–1058, 2010.

[22] D. J. Vandyke, J. Wu, S. M. Logan et al., "Identification of genes involved in the assembly and attachment of a novel flagellin N-linked tetrasaccharide important for motility in the archaeon Methanococcus maripaludis," *Molecular Microbiology*, vol. 72, no. 3, pp. 633–644, 2009.

[23] B. Chaban, S. Voisin, J. Kelly, S. M. Logan, and K. F. Jarrell, "Identification of genes involved in the biosynthesis and attachment of Methanococcus voltae N-linked glycans: insight into N-linked glycosylation pathways in Archaea," *Molecular Microbiology*, vol. 61, no. 1, pp. 259–268, 2006.

[24] D. Calo, L. Kaminski, and J. Eichler, "Protein glycosylation in Archaea: sweet and extreme," *Glycobiology*, vol. 20, no. 9, pp. 1065–1076, 2010.

[25] G. Paul and F. Wieland, "Sequence of the halobacterial glycosaminoglycan," *Journal of Biological Chemistry*, vol. 262, no. 20, pp. 9587–9593, 1987.

[26] M. Abu-Qarn, A. Giordano, F. Battaglia et al., "Identification of AglE, a second glycosyltransferase involved in N glycosylation of the Haloferax volcanii S-layer glycoprotein," *Journal of Bacteriology*, vol. 190, no. 9, pp. 3140–3146, 2008.

[27] M. Abu-Qarn, S. Yurist-Doutsch, A. Giordano et al., "Haloferax volcanii AglB and AglD are involved in N-glycosylation of the S-layer glycoprotein and proper assembly of the surface layer," *Journal of Molecular Biology*, vol. 374, no. 5, pp. 1224–1236, 2007.

[28] U. Karcher, H. Schroder, E. Haslinger et al., "Primary structure of the heterosaccharide of the surface glycoprotein of Methanothermus fervidus," *Journal of Biological Chemistry*, vol. 268, no. 36, pp. 26821–26826, 1993.

[29] B. Chaban, S. M. Logan, J. F. Kelly, and K. F. Jarrell, "AglC and AglK are involved in biosynthesis and attachment of diacetylated glucuronic acid to the N-glycan in Methanococcus voltae," *Journal of Bacteriology*, vol. 91, no. 1, pp. 187–195, 2009.

[30] J. Kelly, S. M. Logan, K. F. Jarrell, D. J. VanDyke, and E. Vinogradov, "A novel N-linked flagellar glycan from Methanococcus maripaludis," *Carbohydrate Research*, vol. 344, no. 5, pp. 648–653, 2009.

[31] L. L. Yang and A. Haug, "Purification and partial characterization of a procaryotic glycoprotein from the plasma membrane of Thermoplasma acidophilum," *Biochimica et Biophysica Acta*, vol. 556, no. 2, pp. 265–277, 1979.

[32] M. Igura, N. Maita, J. Kamishikiryo et al., "Structure-guided identification of a new catalytic motif of oligosaccharyltransferase," *The EMBO Journal*, vol. 27, no. 1, pp. 234–243, 2008.

[33] S.-V. Albers, M. G. L. Elferink, R. L. Charlebois, C. W. Sensen, A. J. M. Driessen, and W. N. Konings, "Glucose transport in the extremely thermoacidophilic Sulfolobus solfataricus involves a high-affinity membrane-integrated binding protein," *Journal of Bacteriology*, vol. 181, no. 14, pp. 4285–4291, 1999.

[34] M. G. L. Elferink, S.-V. Albers, W. N. Konings, and A. J. M. Driessen, "Sugar transport in Sulfolobus solfataricus is mediated by two families of binding protein-dependent ABC transporters," *Molecular Microbiology*, vol. 39, no. 6, pp. 1494–1503, 2001.

[35] T. D. Brock, K. M. Brock, R. T. Belly, and R. L. Weiss, "Sulfolobus: a new genus of sulfur-oxidizing bacteria living at low pH and high temperature," *Archiv für Mikrobiologie*, vol. 84, no. 1, pp. 54–68, 1972.

[36] The UniProt Consortium, "The Universal Protein Resource (UniProt) in 2010," *Nucleic Acids Research*, vol. 38 (Database issue), pp. D142–D148, 2010.

Hindawi Publishing Corporation
Archaea
Volume 2010, Article ID 315108, 9 pages
doi:10.1155/2010/315108

Research Article

Identification of Residues Important for the Activity of *Haloferax volcanii* AglD, a Component of the Archaeal N-Glycosylation Pathway

Lina Kaminski and Jerry Eichler

Department of Life Sciences, Ben Gurion University, P.O. Box 653, Beersheva 84105, Israel

Correspondence should be addressed to Jerry Eichler, jeichler@bgu.ac.il

Received 11 January 2010; Accepted 10 February 2010

Academic Editor: Julie Maupin-Furlow

In *Haloferax volcanii*, AglD adds the final hexose to the N-linked pentasaccharide decorating the S-layer glycoprotein. Not knowing the natural substrate of the glycosyltransferase, together with the challenge of designing assays compatible with hypersalinity, has frustrated efforts at biochemical characterization of AglD activity. To circumvent these obstacles, an in vivo assay designed to identify amino acid residues important for AglD activity is described. In the assay, restoration of AglD function in an *Hfx. volcanii aglD* deletion strain transformed to express plasmid-encoded versions of AglD, generated through site-directed mutagenesis at positions encoding residues conserved in archaeal homologues of AglD, is reflected in the behavior of a readily detectable reporter of N-glycosylation. As such Asp110 and Asp112 were designated as elements of the DXD motif of AglD, a motif that interacts with metal cations associated with nucleotide-activated sugar donors, while Asp201 was predicted to be the catalytic base of the enzyme.

1. Introduction

Although the presence of N-glycosylated proteins in Archaea has been known for over 30 years [1], the pathways responsible for this posttranslational modification have only recently been addressed. In *Methanococcus voltae*, *Methanococcus maripaludis,* and *Haloferax volcanii*, products of the *agl* genes have been shown to participate in the assembly of oligosaccharides decorating various glycoproteins in these species [2–4]. At present, however, apart from the oligosaccharyltransferase, AglB [5–7], virtually nothing is known of the catalytic workings of the different Agl proteins. Of the *Hfx. volcanii* Agl proteins identified to date, at least five (i.e., AglD, AglE, AglG, AglI, and AglJ) are predicted to act as glycosyltransferases (GTs), enzymes that catalyze the formation of glycosidic bonds through the transfer of the sugar moieties from nucleotide-activated saccharides to appropriate targets [8].

Based on their amino acid similarities, GTs can be classified into 91 family groups (http://www.cazy.org/fam/acc_GT.html; January, 2009), varying in size and number of functions fulfilled by family members [9, 10]. Furthermore, the different GT families can be clustered based on whether the canonical GT-A or GT-B fold is employed and whether sugar stereochemistry is retained or inverted upon addition of a glycosyl donor [11]. Still, the ability to predict the function of a given GT or to define its catalytic mechanism remains a challenge. This is particularly true in the case of the GT2 family, an ancient group of GT-A fold-bearing GTs containing over 10,000 members derived from various sources and serving at least 12 distinct functions [11, 12]. Like all GT-A fold-bearing GTs, GT2 family members contain a DXD signature motif, shown to interact with a divalent cation (usually Mg^{2+} or Mn^{2+}) that facilitates the leaving of the nucleoside diphosphate group of a nucleotide-activated sugar donor as part of the S_N2-like displacement mechanism believed to be employed by these enzymes [11, 13–17]. The DXD motif also serves to divide the GT-A fold into two portions. The N-terminal portion, containing the sequence that assigns the protein to the GT2 family [18], binds the nucleotide-activated sugar donor [19–22]. By contrast, the C-terminal portion is highly variable and generally serves

to recognize the acceptor [17]. Despite such variability, the C-terminal portion of GT2 family members includes a conserved Asp or Glu residue that presumably serves as the catalytic base, thought to assist in the protonation of the nucleophilic hydroxyl group of the acceptor saccharide [11, 21, 23–25].

The GT2 glycosyltransferase family includes *Hfx. volcanii* AglD, previously shown to participate in adding the final hexose to the pentasaccharide comprising two hexoses, two hexuronic acids, and a methylated ester of hexuronic acid decorating at least two sequons of the S-layer glycoprotein [26, 27]. However, due to the fact that its natural substrates have yet to be defined and the challenge of devising in vitro assays for haloarchaeal enzymes due to their hypersaline requirements, little is known of the catalytic workings of AglD. Towards remedying the situation, an in vivo approach has been developed in which the ability of plasmid-encoded versions of AglD, modified through site-directed mutagenesis, to restore the absent function to an *aglD* deletion strain, was tested. Results obtained employing this novel assay point to Asp110-Thr111-Asp112 as corresponding to the DXD motif and Asp201 as corresponding to the catalytic base of *Hfx. volcanii* AglD.

2. Methods

2.1. Strains and Growth Conditions. The *Hfx. volcanii* background strain WR536 (H53) and the same strain deleted for *aglD* were grown in complete medium containing 3.4 M NaCl, 0.15 M $MgSO_4 \cdot 7H_20$, 1 mM $MnCl_2$, 4 mM KCl, 3 mM $CaCl_2$, 0.3% (w/v) yeast extract, 0.5% (w/v) tryptone and 50 mM Tris-HCl, pH 7.2, at 42°C [28]. A complete description of the *aglD* deletion strain and the protocol used to delete the gene have been previously published [5].

2.2. In Vivo AglD Assay. To assay AglD activity, *Hfx. volcanii* cells deleted of *aglD* [5] were transformed to express plasmid-encoded versions of AglD that included an N-terminally fused *Clostridium thermocellum* cellulose-binding domain (CBD) [29]. To introduce nonnative residues into AglD, the plasmid-encoded version of *aglD* (GenBank accession number CAM91696.1) was modified by site-directed mutagenesis. Restoration of AglD function lost as a result of deletion of the genomic copy of the encoding gene was determined by the ability of the transformed cells to reverse the enhanced SDS-PAGE migration of the S-layer glycoprotein and loss of PAS glycostaining of the same reporter, that is, novel traits of the S-layer glycoprotein that appeared in cells lacking AglD.

2.3. Site-Directed Mutagenesis. Mutated versions of *aglD* were generated by site-directed mutagenesis using the Quikchange (Stratagene) protocol, performed according to the manufacturer's instructions, with plasmid pWL-CBD-AglD, encoding CBD-AglD [29], serving as template. Oligonucleotide primers used to introduce the various mutations are listed in Supplementary Table 1 (available online at doi:10.1155/2010/315108). The introduction of mutations was confirmed by sequencing, performed both before and following introduction of plasmid-encoded mutated *aglD* into *Hfx. volcanii*.

2.4. Other Methods. Periodic acid-Schiff (PAS) reagent glycoprotein staining was performed as described previously [30]. Immunoblots were performed using polyclonal antibodies raised against the *C. thermocellum* CBD (obtained from Ed Bayer, Weizmann Institute of Science; 1 : 10,000). Antibody binding was detected using goat antirabbit horseradish peroxidase-(HRP-) conjugated antibodies (1 : 4000, BioRad, Hercules, CA) and an ECL-enhanced chemiluminescence kit (Amersham, Buckingham, UK).

3. Results

3.1. AglD Activity in AglD-Deleted Hfx. volcanii Cells Is Restored Upon Complementation with Plasmid-Encoded AglD. As a first step towards describing AglD function, efforts were directed at creating an assay to allow for characterization of the activity of the enzyme. Accordingly, cells deleted of the encoding gene were transformed with a plasmid encoding a version of the protein designed to include an N-terminally fused *Clostridium thermocellum* cellulose-binding domain (CBD) [29]. As previously reported [29], an 85 kDa band, corresponding to the predicted molecular mass of the 17 kDa CBD moiety and the 68 kDa AglD protein, was expressed in the transformed cells and recognized in an immunoblot using antiCBD antibodies (not shown).

Deletion of *aglD* results in the absence of the final hexose of the pentasaccharide decorating the *Hfx. volcanii* S-layer glycoprotein [26]. As such, the S-layer glycoprotein in the deletion strain migrates faster in SDS-PAGE than does the native protein in the background strain [5]. Moreover, the S-layer glycoprotein is not recognized by PAS glycostain in the mutant strain. However, as reflected in Figure 1, the S-layer glycoprotein from cells of the *aglD*-deleted strain transformed to express CBD-tagged AglD migrated to the same position as did the protein from the background strain and was similarly PAS-stained. As such, complementation of *Hfx. volcanii* cells lacking *aglD* with an AglD-encoding plasmid restores the absent activity to the deletion strain.

To demonstrate the involvement of a given AglD residue in the activity of the enzyme, the return of AglD activity to the *aglD* deletion strain, upon introduction of a plasmid-encoded version of AglD modified at the amino acid position in question, was assessed. In these experiments, the SDS-PAGE migration of the S-layer glycoprotein from cells of the background strain, from cells deleted of *aglD*, and from AglD-lacking cells transformed to express select mutant AglD proteins was addressed. In addition, the S-layer glycoprotein in each of the three populations of *Hfx. volcanii* cells was subjected to PAS glycostaining. While this in vivo approach cannot distinguish between residues necessary for catalytic activity from those important for proper AglD folding, it, nonetheless, offers a facile route for identifying important AglD residues until such time as AglD activity can be directly assayed in vitro.

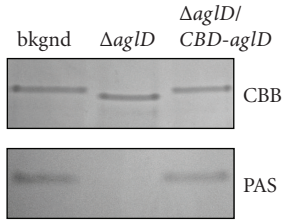


Figure 1: *aglD*-complemented *Hfx. volcanii* cells regain the ability to properly glycosylate the S-layer glycoprotein. The protein contents of cells of the WR536 background strain (bkgnd), the same strain deleted of *aglD* (Δ*aglD*) or the AglD-lacking strain transformed with a plasmid encoding CBD-AglD (Δ*aglD/CBD-aglD*) were separated by 5% SDS-PAGE and the S-layer glycoprotein was detected by Coomassie stain (CBB) or periodic acid-Schiff (PAS) reagent. In the presence of CBD-AglD, the migration and positive glycostaining of the S-layer glycoprotein are as observed in the background strain.

3.2. Identification of Conserved AglD Residues. To select candidate residues for site-directed mutagenesis, the *Hfx. volcanii* AglD sequence was aligned with selected homologous archaeal sequences using ClustalW (http://www.ebi.ac.uk/Tools/clustalw2/index.html). It should be noted, however, that it is not yet known whether the various homologues considered indeed catalyze the same reaction as does *Hfx. volcanii* AglD. Indeed, it remains to be confirmed that N-glycosylation occurs in all of the species listed. Nonetheless, such alignment revealed the presence of a stretch of amino acids in the N-terminal region of AglD showing substantial overlap with similarly situated regions in the various archaeal homologues considered (Figure 2). In the *Hfx. volcanii* protein, this stretch corresponds to the region between Asp110 and Glu203, a portion of the protein previously localized to the cytoplasm [29] and which includes seven residues absolutely conserved in the sequences considered, namely, Asp110, Asp112, Asp133, Arg139, Arg152, Asp173, and Gly177. Between Asp133 and Arg139, between Trp198 and Glu203, and in the region surrounding Asp173 and Gly177, several highly conserved residues were also detected. To determine whether any of these residues contribute to AglD function, the corresponding *aglD* codons were modified by site-directed mutagenesis using the primer pairs listed in Supplementary Table 1, and the ability of plasmid-encoded versions of the mutant proteins to restore AglD function in the *aglD* deletion strain was considered.

3.3. The DXD Catalytic Motif of AglD Likely Comprises Asp110 and Asp 112. The GT-A fold found in GT2 family members includes a DXD motif that contributes to the catalytic activity of the enzyme [11, 13, 14, 16, 17]. Sequence alignment-based examination of the *Hfx. volcanii* AglD sequence points to Asp110-Thr111-Asp112 as comprising this motif (Figure 2). To directly test this hypothesis, the site-directed mutagenesis approach described above was enlisted. Figure 3 addresses the effects of replacing either AglD Asp110 or Asp112 with other residues. Transformation of AglD-lacking cells to express AglD D110A resulted in both the failure of plasmid-encoded AglD to restore S-layer glycoprotein migration to the position of this reporter in the background strain as well as the lost ability of PAS glycostain to label the S-layer glycoprotein. The same was true in cells transformed to express AglD D110E, a mutation that retains the negative charge at this position (Figure 3), or upon introduction of an Asn residue at this position (not shown).

When AglD Asp112 of the plasmid-encoded protein was replaced with an Asn, no recovery of AglD function in the transformed *aglD* deletion strain was realized, reflected in the inability of cells expressing the mutated version of AglD to restore SDS-PAGE migration and PAS glycostaining of the S-layer glycoprotein, as realized in the background strain (Figure 3). By contrast, transformation of the deletion strain to express AglD D112E led to a restoration of SDS-PAGE migration of the S-layer glycoprotein to the position seen in background cells but only a partial recovery (6% ± 0.5% (standard deviation), $n = 3$) of PAS glycostaining (Figure 3).

The importance of Asp110 and Asp112 for *Hfx. volcanii* AglD activity points to these two residues as comprising the DXD motif found in GT-A fold-bearing GTs. However, while the Asp residue at position 110 is apparently essential for activity, the presence of Glu at position 112 yields a functional enzyme that apparently acts differently from the native enzyme, as reflected in the limited PAS staining detected with this mutant.

3.4. Asp201 Is Likely the Catalytic Base of AglD. In addition to the DXD motif considered above, the activity of GT2 family members also relies on an Asp or Glu residue found in the acceptor-binding domain of the protein. First identified in the solved three-dimensional structure of *Bacillus subtilis* SpsA as Asp191 [20], this residue and its equivalents in other GTs are thought to serve as the base catalyst in the direct displacement mechanism apparently employed by these enzymes [11, 20]. To identify the *Hfx. volcanii* AglD equivalent of *B. subtilis* SpsA Asp191, the sequence of the soluble region of AglD (residues 1–259) was aligned with the sequence of *B. subtilis* SpsA, as well as with those of the other GT2 enzymes where the functional equivalent of *B. subtilis* SpsA Asp191 is known, namely, *Sinorhizobium meliloti* ExoM and *Salmonella enterica* WbbE. Earlier site-directed mutagenesis efforts had revealed ExoM Asp187 and WbbE Glu180 to serve the same role as SpsA Asp191 [21, 22]. Tcoffee (http://tcoffee.vital-it.ch/cgi-bin/Tcoffee/tcoffee_cgi/index.cgi?stage1=1anddaction=TCOFFEE::Regular) aligned AglD Asp201 with SpsA Asp191, ExoM Asp187 and WbbE Glu180. The MAFFT program (v6.531b; http://www.ebi.ac.uk/Tools/mafft/index.html) also aligned AglD Asp201 with the same SpsA, ExoM and WbbE residues. On the other hand, alignment of archaeal homologues of *Hfx. volcanii* AglD using the ClustalW (Figure 2), Tcoffee or MAFFT programs revealed that AglD Asp173 but not Asp201 is conserved. In all cases, the programs consulted were used with their default settings.

As a next step towards identifying the AglD equivalent of SpsA Asp191, the importance of Asp173, Asp195, and

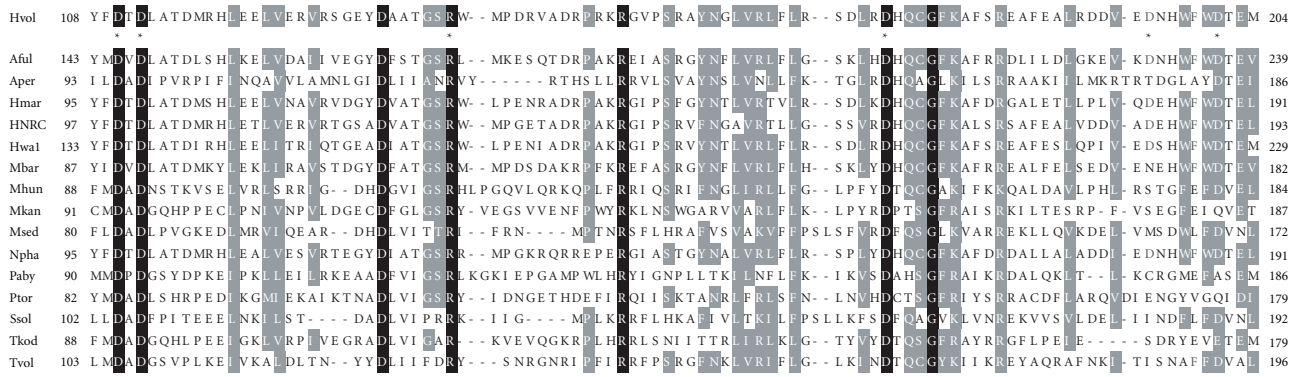


Figure 2: Conserved residues in archaeal AglD proteins. The sequences of *Hfx. Volcanii* AglD (accession number AM698042) and AglD homologues in *Archaeoglobus fulgidus* (NP_069415.1; Aful), *Aeropyrum pernix* (NP_147774.1; Aper), *Haloarcula marismortui* (YP_136461.1; Hmar), *Halobacterium* sp. NRC-1 (NP_279416.1; HNRC), *Haloquadratum walsbyi* (YP_657261.1; Hwal), *Methanosarcina acetivorans* (NP_618739.1; Mace), *Methanosarcina barkeri* (YP_304067.1; Mbar), *Methanospirillum hungatei* (YP_503949.1; Mhun), *Methanopyrus kandleri* (NP_614163.1; Mkan), *Metallosphaera sedula* (YP_001191894; Msed), *Natronomonas pharaonis* (YP_326773.1; Npha), *Pyrococcus abyssi* (NP_127133.1; Paby), *Picrophilus torridus* (YP_024256.1; Ptor), *Sulfolobus solfataricus* (NP_342803.1; Ssol), *Thermococcus kodakarensis* (YP_182777.1; Tkod), and *Thermoplasma volcanium* (NP_111403.1; Tvol) were aligned by ClustalW2 (www.ebi.ac.uk/Tools/clustalw2/index.html), using the default settings. The region of the highest similarity is shown. Completely conserved residues are shown against a black background, while largely conserved residues (i.e., similar residues conserved in at least 11 sequences) are shown against a grey background. Amino acid numbers are shown at the start and end of each sequence. Asterisks are placed under *Hfx. volcanii* AglD D110, D112, R139, D173, D195, and D201 (see text for details).

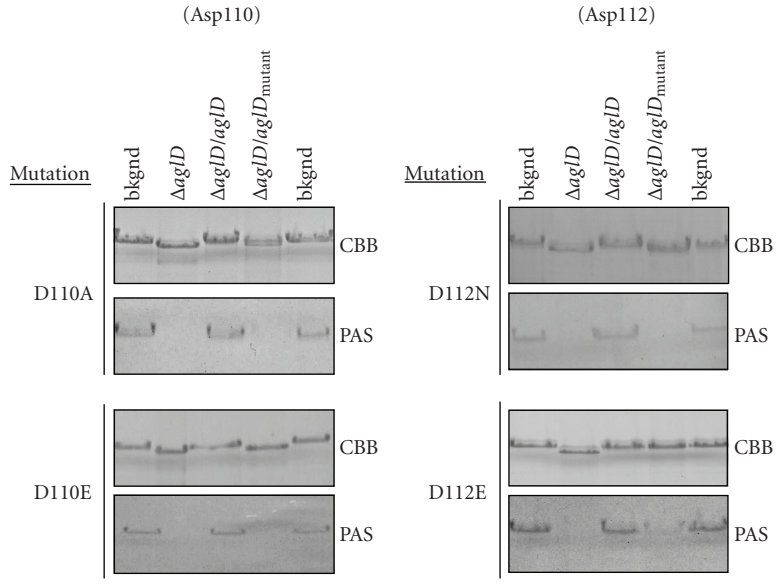


Figure 3: *Hfx. volcanii* AglD Asp110 and Asp112 residues likely participate in the GT2 DXD motif involved in the catalytic activity of the enzyme. Site-directed mutagenesis was performed to generate CBD-AglD containing mutations of Asp110 (left column) or Asp112 (right column), as listed on the left of each panel. For each mutant, the upper and lower panels, respectively, show the Coomassie- and PAS-stained S-layer glycoprotein from the background strain (lanes 1 and 5), from the *aglD* deletion strain (lane 2), from the *aglD* deletion strain complemented with a plasmid encoding CBD-AglD (lane 3), or from the *aglD* deletion strain complemented with a plasmid encoding CBD fused to mutated AglD (lane 4).

Asp201 for AglD activity was considered by site-directed mutagenesis.

Transformation of *Hfx. volcanii* Δ*aglD* cells with a plasmid encoding AglD D173E (Figure 4) did not restore the SDS-PAGE migration or glycostaining of the S-layer glycoprotein. However, as these S-layer glycoprotein traits were fully restored in cells expressing AglD D173N, it would appear that the conserved Asp173 does not serve as the catalytic base of the enzyme but is likely important for AglD structure. This idea may be supported by the observation that the D173A mutant could not be expressed (not shown). The importance of Asp195, another somewhat conserved

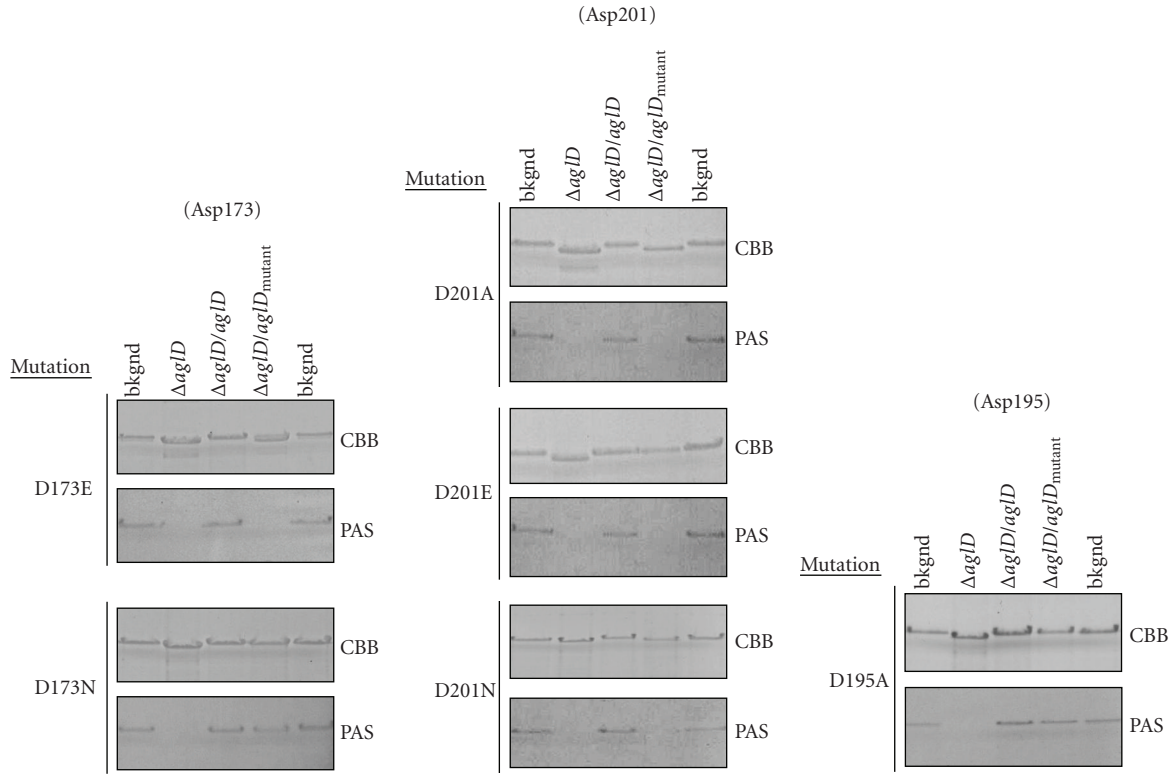


Figure 4: *Hfx. volcanii* AglD Asp201 is apparently the catalytic base of the enzyme. Site-directed mutagenesis was performed to generate CBD-AglD containing mutations of Asp173 (upper left column), Asp195 (lower left column), and Asp201 (right column), as listed on the left of each panel. For each mutant, the upper and lower panels, respectively, show the Coomassie- and PAS-stained S-layer glycoprotein from the background strain (lanes 1 and 5), from the *aglD* deletion strain (lane 2), from the *aglD* deletion strain complemented with a plasmid encoding CBD-AglD (lane 3), or from the *aglD* deletion strain complemented with a plasmid encoding CBD fused to mutated AglD (lane 4).

Asp residue in this region, was also considered. *Hfx. volcanii* Δ*aglD* cells transformed to express AglD D195A (Figure 4) or D195E (not shown) readily replaced the actions of the missing enzyme, showing that the Asp at this position is not necessary for AglD activity. By contrast, if the *aglD*-deleted strain was transformed to express AglD D201A or D201N, SDS-PAGE migration and glycostaining of the S-layer glycoprotein were as observed in cells lacking native AglD (Figure 4). When, however, *Hfx. volcanii* Δ*aglD* cells were transformed to express the D201E mutant, the S-layer glycoprotein behaved as in the background strain (Figure 4). These results thus point to Asp201 as being the catalytic base of AglD and the functional equivalent of SpsA Asp191. Furthermore, as is the case with other GT2 family members [22], Asp201 can be replaced by Glu.

Homology modeling of *Hfx. volcanii* AglD residues Asp110-Asp112, Asp173, Asp195, and Asp201, based on the available three-dimensional structural of *B. subtilis* SpsA [20], further supports the assignment of AglD D201 as being equivalent to subtilis SpsA Asp191. As reflected in Figure 5, considerable overlap in term of both position and orientation exists between AglD D201 and *B. subtilis* SpsA Asp191. The same cannot be said for either AglD Asp173 or Asp195.

3.5. The Conserved Arg139 Residue Is Needed for AglD Activity. *Hfx. volcanii* AglD and its archaeal homologues also contain several other fully conserved residues in that part of the soluble N-terminal region under consideration in this study. The contribution of these residues, as well as that of their neighbors, was next considered. Complementation of Δ*aglD* *Hfx. volcanii* cells with plasmid-encoded AglD R139A failed to restore either S-layer glycoprotein migration in SDS-PAGE or the ability of PAS glycostain to label this reporter (Figure 6). The same was true in cells expressing AglD R139E, R139K or R139M (not shown). Thus, AglD Arg139 is apparently essential for enzyme activity. By contrast, complementation of AglD-lacking *Hfx. volcanii* cells with plasmid-encoded AglD D133A or G177A restored S-layer glycoprotein SDS-PAGE migration to that observed for the native protein, although less S-layer glycoprotein is detected in the cells expressing AglD D133A. The significance of this observation is not clear. In addition, the S-layer glycoprotein in both AglD D133A- and G177A-expressing cells could be glycostained (Figure 6). As such, although conserved in *Hfx. volcanii* AglD and its archaeal homologues, Asp133 and Gly177 do not appear to be essential for the catalytic workings of the enzyme. Similarly, the introduction of CBD-tagged AglD G137A, S138A, Q175A, C176A, F178A or K179A mutants into *Hfx. volcanii* Δ*aglD* cells led to a

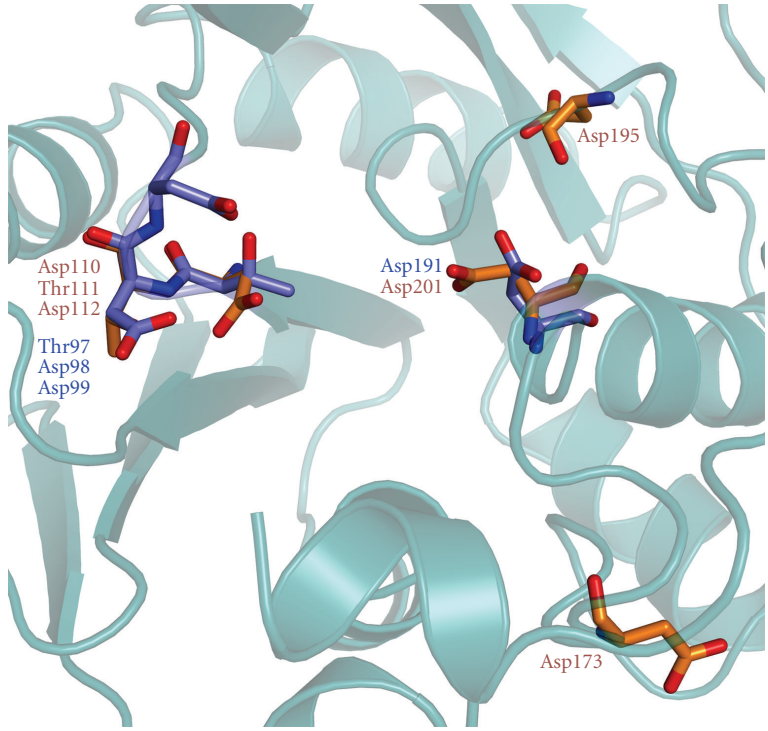


Figure 5: Homology modeling of *Hfx. volcanii* AglD residues based on the available three-dimensional structure of *B. subtilis* SpsA. Structural modeling was performed by using the SWISS-MODEL program (http://swissmodel.expasy.org/) and visualized using PyMol (http://www.pymol.org/). *B. subtilis* SpsA Thr97, Asp98, Asp99, and Asp191 are shown in blue, while *Hfx. volcanii* AglD Asp110, Thr111, Asp112, Asp173, Asp195, and Asp 201 are shown in brown. The ribbon diagram in the background corresponds to the three-dimensional structure of SpsA [20]. The RMS value, reflecting the quality of the homology modeling, is 0.61 angstroms.

restoration of AglD activity, indicating that none of these residues contribute to the reaction catalyzed by the enzyme (not shown).

Finally, to eliminate the possibility that the inability of certain plasmid-encoded versions of the protein to restore absent AglD activity was due to poor or no expression, the level of each CBD-AglD considered in this study was assessed by immunoblot using antiCBD antibodies (Figure 7).

4. Discussion

When one considers that *Nanoarchaeum equitans*, the archaeon containing the smallest genome identified to date [31, 32], encodes only 3 GTs, namely, one member of the GT2 family and two members of the GT4 family [33], it is fair to say that analysis of archaeal GTs can provide unique insight into the evolution of such enzymes, as well as adding to our comprehension of protein processing in extreme conditions. Despite such promise, only limited experimental data on archaeal GTs involved in protein glycosylation is presently available. The crystal structure of Stt3/AglB from *Pyrococcus furiosus*, the sole component of the archaeal oligosaccharyltransferase [5, 6], has been solved [7], shedding new light on the workings of this central component of the N-glycosylation machinery. Still, although *P. furiosus* has been reported to contain glycoproteins [34], the oligosaccharyltransferase in this species has been thus far only demonstrated to modify an artificial substrate [7, 35]. Similarly, while both *Thermoplasma acidophilum* [36] and *Pyrococcus horikoshii* [37] have been reported to contain glycoproteins, the participation of biochemically characterized dolichyl phosphomannose synthases from these species [37, 38] in protein glycosylation has yet to be demonstrated. As such, the present analysis of *Hfx. volcanii* AglD represents the first examination of a glycosyltransferase experimentally verified as participating in the modification of an identified archaeal glycoprotein, namely, the S-layer glycoprotein.

In the present study, sequence alignment was combined with site-directed mutagenesis to identify AglD residues important for the function of the enzyme, as reflected in AglD-mediated modulation of the SDS-PAGE migration and glycostaining of a reporter glycoprotein, the S-layer glycoprotein. This approach assigned AglD Asp110, Thr111, and Asp112 as the DXD motif typical of inverting GT-A fold-bearing GTs. AglD Asp110 was shown to be essential for catalytic activity, while a negative charge at position 112 was deemed necessary. In the case of the AglD D112E mutant, recovery of S-layer glycoprotein SDS-PAGE migration was noted, yet the loss of PAS glycostaining associated with the deletion strain was largely not restored. This could reflect the generation of an enzyme possessing different activity than that of the native protein, one that adds a different sugar to the final position of S-layer glycoprotein-bound pentasaccharide. Indeed, the failure of PAS glycostain to label

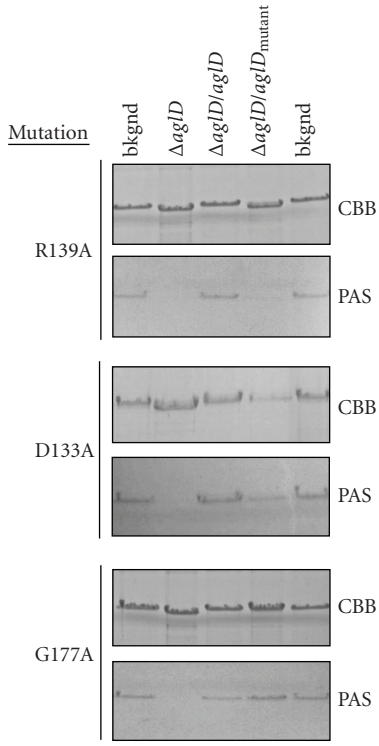


Figure 6: The conserved Arg139 residue is needed for AglD activity, unlike the conserved Asp133 or Gly177 residues. Site-directed mutagenesis was performed to generate CBD-AglD containing mutations of Asp133, Arg139, and Gly177, as listed on the left of each panel. For each mutant, the upper and lower panels respectively show the Coomassie- and PAS-stained S-layer glycoprotein from the background strain (lanes 1 and 5), from the *aglD* deletion strain (lane 2), from the *aglD* deletion strain complemented with a plasmid encoding CBD-AglD (lane 3), or from the *aglD* deletion strain complemented with a plasmid encoding CBD fused to mutated AglD (lane 4).

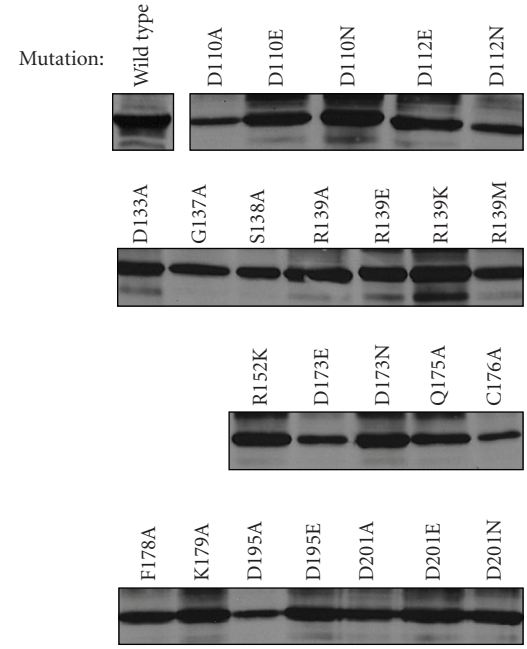


Figure 7: Expression levels of the various versions of CBD-AglD. *Hfx. volcanii* cells expressing the various AglD mutants considered in this study were grown to OD_{550} 1.0 and their protein contents were separated on 10% SDS-PAGE. The CBD-AglD content of each strain was subsequently assessed by immunoblot using polyclonal antiCBD antibodies. Antibody binding was detected using HRP-conjugated secondary antibodies and an enhanced chemiluminescence kit.

the tetrasaccharide N-linked to the S-layer glycoprotein in cells lacking AglD points to inability of this labeling reagent to interact with certain sugar subunits. Moreover, within the GT2 family (whose members include *Hfx. volcanii* AglD), differences in the organization and importance of DXD motif constituents exist. In the case of *S. meliloti* ExoM, where the DXD motif includes Asp96 and Asp98, it was shown that replacing the former with Ala completely eliminated enzymatic activity, whereas the same replacement at position 98 only led to a 70% loss of activity [21]. In *S. enterica* WbbE, where the DXD motif is expanded to include Asp93, Asp95, and Asp96, it was shown that exchanging either Asp93 or Asp96 with Ala abolished enzyme activity, while the same replacement at Asp95 only reduced that activity [22].

The site-directed mutagenesis approach developed here, along with sequence alignment and homology modeling, also indicate Asp201 as likely serving as the AglD catalytic base. Just as the corresponding residue in *S. enterica* WbbE, that is, Glu180, could be functionally replaced by Asp [22], AglD D201E was also active. By contrast, replacing the Asp187 catalytic base in *S. mililoti* ExoM led to a complete loss of function [21]. Such nuances may be indicative of differences in the donors and/or acceptors employed by each enzyme or point to unique mechanistic traits. In addition, although conserved in the archaeal AglD homologues examined in this study, *Hfx. volcanii* AglD Asp173 was not assigned as the catalytic base of the enzyme, given its functional replacement by a similarly sized Asn but not a similarly charged Glu. Hence, it would appear that Asp173 is of structural, rather than catalytic, importance to AglD activity. In addition to these residues, at least another *Hfx. volcanii* AglD amino acid seems to be needed for enzyme function, that is, Arg139. The AglD counterparts of other residues shown to be important for the catalytic activity of GT2 family members, such as *S. meliloti* ExoM Asp44 and Asp96 [21], may also play a role in the activity of the archaeal enzyme.

In conclusion, this paper describes an in vivo assay designed to consider the contribution of various AglD residues to the activity of the enzyme. In the assay, the ability of plasmid-encoded versions of AglD, selectively mutated at positions suspected of being important for enzyme function, to restore both S-layer glycoprotein SDS-PAGE migration and glycostaining affected in Δ*aglD* cells is assessed. In this manner, Asp110, Asp112, and Asp201 were all determined as being important for AglD activity, as was Asp139.

Acknowledgment

The authors thank Dr. Raz Zarivach for assistance with the homology modeling. Support came from the Israel Science Foundation (Grant 30/07) and the US Army Research Office (Grant W911NF-07-1-0260).

References

[1] M. F. Mescher and J. L. Strominger, "Purification and characterization of a prokaryotic glycoprotein from the cell envelope of *Halobacterium salinarium*," *The Journal of Biological Chemistry*, vol. 251, no. 7, pp. 2005–2014, 1976.

[2] M. Abu-Qarn, J. Eichler, and N. Sharon, "Not just for Eukarya anymore: N-glycosylation in Bacteria and Archaea," *Current Opinion in Structural Biology*, vol. 18, no. 5, pp. 544–550, 2008.

[3] S. Yurist-Doutsch, B. Chaban, D. J. VanDyke, K. F. Jarrell, and J. Eichler, "Sweet to the extreme: protein glycosylation in Archaea," *Molecular Microbiology*, vol. 68, no. 5, pp. 1079–1084, 2008.

[4] D. J. VanDyke, J. Wu, and S. M. Logan, "Identification of genes involved in the assembly and attachment of a novel flagellin N-linked tetrasaccharide important for motility in the archaeon *Methanococcus maripaludis*," *Molecular Microbiology*, vol. 72, no. 3, pp. 633–644, 2009.

[5] M. Abu-Qarn and J. Eichler, "Protein N-glycosylation in Archaea: defining *Haloferax volcanii* genes involved in S-layer glycoprotein glycosylation," *Molecular Microbiology*, vol. 61, pp. 511–525, 2006.

[6] B. Chaban, S. Voisin, J. Kelly, S. M. Logan, and K. F. Jarrell, "Identification of genes involved in the biosynthesis and attachment of *Methanococcus voltae* N-linked glycans: insight into N-linked glycosylation pathways in Archaea," *Molecular Microbiology*, vol. 61, no. 1, pp. 259–268, 2006.

[7] M. Igura, N. Maita, J. Kamishikiryo, et al., "Structure-guided identification of a new catalytic motif of oligosaccharyltransferase," *EMBO Journal*, vol. 27, no. 1, pp. 234–243, 2008.

[8] N. Taniguchi, A. Ekuni, J. H. Ko, et al., "A glycomic approach to the identification and characterization of glycoprotein function in cells transfected with glycosyltransferase genes," *Proteomics*, vol. 1, no. 2, pp. 239–247, 2001.

[9] P. M. Coutinho, E. Deleury, G. J. Davies, and B. Henrissat, "An evolving hierarchical family classification for glycosyltransferases," *Journal of Molecular Biology*, vol. 328, no. 2, pp. 307–317, 2003.

[10] B. I. Cantarel, P. M. Coutinho, C. Rancurel, T. Bernard, V. Lombard, and B. Henrissat, "The Carbohydrate-Active EnZymes database (CAZy): an expert resource for glycogenomics," *Nucleic Acids Research*, vol. 37, supplement 1, pp. D233–D238, 2009.

[11] L. L. Lairson, B. Henrissat, G. J. Davies, and S. G. Withers, "Glycosyl transferases: structures, functions, and mechanisms," *Annual Review of Biochemistry*, vol. 77, pp. 521–555, 2008.

[12] B. Henrissat, G. Sulzenbacher, and Y. Bourne, "Glycosyltransferases, glycoside hydrolases: surprise, surprise!," *Current Opinion in Structural Biology*, vol. 18, no. 5, pp. 527–533, 2008.

[13] C. Breton, E. Bettler, D. H. Joziasse, R. A. Geremia, and A. Imberty, "Sequence-function relationships of prokaryotic and eukaryotic galactosyltransferases," *Journal of Biochemistry*, vol. 123, no. 6, pp. 1000–1009, 1998.

[14] C. A. Wiggins and S. Munro, "Activity of the yeast MNN1 alpha-1,3-mannosyltransferase requires a motif conserved in many other families of glycosyltransferases," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 95, pp. 7945–7950, 1998.

[15] C. Breton and A. Imberty, "Structure/function studies of glycosyltransferases," *Current Opinion in Structural Biology*, vol. 9, no. 5, pp. 563–571, 1999.

[16] U. M. Unligil and J. M. Rini, "Glycosyltransferase structure and mechanism," *Current Opinion in Structural Biology*, vol. 10, no. 5, pp. 510–517, 2000.

[17] C. Breton, L. Šnajdrová, C. Jeanneau, J. Koča, and A. Imberty, "Structures and mechanisms of glycosyltransferases," *Glycobiology*, vol. 16, no. 2, pp. 29R–37R, 2006.

[18] J. A. Campbell, G. J. Davies, V. Bulone, and B. Henrissat, "A classification of nucleotide-diphospho-sugar glycosyltransferases based on amino acid sequence similarities," *Biochemical Journal*, vol. 326, no. 3, pp. 929–939, 1997.

[19] I. M. Saxena, R. M. Brown Jr., M. Fevre, R. A. Geremia, and B. Henrissat, "Multidomain architecture of β-glycosyl transferases: implications for mechanism of action," *Journal of Bacteriology*, vol. 177, no. 6, pp. 1419–1424, 1995.

[20] S. J. Charnock and G. J. Davies, "Structure of the nucleotide-diphospho-sugar transferase, SpsA from Bacillus subtilis, in native and nucleotide-complexed forms," *Biochemistry*, vol. 38, no. 20, pp. 6380–6385, 1999.

[21] C. Garinot-Schneider, A. C. Lellouch, and R. A. Geremia, "Identification of essential amino acid residues in the *Sinorhizobium meliloti* glucosyltransferase ExoM," *The Journal of Biological Chemistry*, vol. 275, no. 40, pp. 31407–31413, 2000.

[22] W. J. Keenleyside, A. J. Clarke, and C. Whitfield, "Identification of residues involved in catalytic activity of the inverting glycosyl transferase WbbE from Salmonella enterica serovar borreze," *Journal of Bacteriology*, vol. 183, no. 1, pp. 77–85, 2001.

[23] B. W. Murray, S. Takayama, J. Schultz, and C.-H. Wong, "Mechanism and specificity of human α-1,3-fucosyltransferase V," *Biochemistry*, vol. 35, no. 34, pp. 11183–11195, 1996.

[24] L. C. Pedersen, T. A. Darden, and M. Negishi, "Crystal structure of β1,3-glucuronyltransferase I in complex with active donor substrate UDP-GlcUA," *The Journal of Biological Chemistry*, vol. 277, no. 24, pp. 21869–21873, 2002.

[25] S. Kakuda, T. Shiba, M. Ishiguro, et al., "Structural basis for acceptor substrate recognition of a human glucuronyltransferase, GlcAT-P, an enzyme critical in the biosynthesis of the carbohydrate epitope HNK-1," *The Journal of Biological Chemistry*, vol. 279, no. 21, pp. 22693–22703, 2004.

[26] M. Abu-Qarn, S. Yurist-Doutsch, A. Giordano, et al., "*Haloferax volcanii* AglB and AglD are involved in N-glycosylation of the S-layer glycoprotein and proper assembly of the surface layer," *Journal of Molecular Biology*, vol. 374, no. 5, pp. 1224–1236, 2007.

[27] H. Magidovich, S. Yurist-Doutsch, Z. Konrad, et al., "AglP is a S-adenosyl-L-methionine-dependent methyltransferase that participates in the N-glycosylation pathway of *Haloferax volcanii*," *Molecular Microbiology*, vol. 76, pp. 190–199, 2010.

[28] M. Mevarech and R. Werczberger, "Genetic transfer in *Halobacterium volcanii*," *Journal of Bacteriology*, vol. 162, pp. 461–462, 1985.

[29] N. Plavner and J. Eichler, "Defining the topology of the N-glycosylation pathway in the halophilic archaeon *Haloferax*

volcanii," *Journal of Bacteriology*, vol. 190, no. 24, pp. 8045–8052, 2008.

[30] G. Dubray and G. Bezard, "A highly sensitive periodic acid-silver stain for 1,2-diol groups of glycoproteins and polysaccharides in polyacrylamide gels," *Analytical Biochemistry*, vol. 119, no. 2, pp. 325–329, 1982.

[31] H. Huber, M. J. Hohn, R. Rachel, T. Fuchs, V. C. Wimmer, and K. O. Stetter, "A new phylum of Archaea represented by a nanosized hyperthermophilic symbiont," *Nature*, vol. 417, no. 6884, pp. 63–67, 2002.

[32] E. Waters, M. J. Hohn, and I. Ahel, "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*, vol. 100, pp. 12984–12988, 2003.

[33] H. Magidovich and J. Eichler, "Glycosyltransferases and oligosaccharyltransferases in Archaea: putative components of the N-glycosylation pathway in the third domain of life," *FEMS Microbiology Letters*, vol. 300, pp. 120–130, 2009.

[34] M. V. Weinberg, G. J. Schut, S. Brehm, S. Datta, and M. W. W. Adams, "Cold shock of a hyperthermophilic archaeon: *Pyrococcus furiosus* exhibits multiple responses to a suboptimal growth temperature with a key role for membrane-bound glycoproteins," *Journal of Bacteriology*, vol. 187, no. 1, pp. 336–348, 2005.

[35] D. Kohda, M. Yamada, M. Igura, J Kamishikiryo, and K. Maenaka, "New oligosaccharyltransferase assay method," *Glycobiology*, vol. 17, pp. 1175–1182, 2007.

[36] L. L. Yang and A. Haug, "Purification and partial characterization of a procaryotic glycoprotein from the plasma membrane of *Thermoplasma acidophilum*," *Biochimica et Biophysica Acta*, vol. 556, no. 2, pp. 265–277, 1979.

[37] Y. Urushibata, S. Ebisu, and I. Matsui, "A thermostable dolichol phosphoryl mannose synthase responsible for glycoconjugate synthesis of the hyperthermophilic archaeon *Pyrococcus horikoshii*," *Extremophiles*, vol. 12, no. 5, pp. 665–676, 2008.

[38] B. C. R. Zhu and R. A. Laine, "Dolichyl-phosphomannose synthase from the Archae *Thermoplasma acidophilum*," *Glycobiology*, vol. 6, no. 8, pp. 811–816, 1996.