Chapter 6

Summary and outlook

The domain of Archaea

The domain of archaea represents a group of organisms which are only distantly related to
the bacteria and eukaryotes. The classification of living organism into the three domains
bacteria, archaea and eukaryotes has been introduced by Carl Woese and was based on
comparison of 16S rRNA sequences (Woese et al., 1990, Woese & Fox, 1977). Sequencing
and comparison of the genetic information of organisms determines the degree of
differences between them and therefore is an indication of the amount of changes
undergone by each species from a shared evolutionary ancestor. The ribosome is an
essential part of the protein synthesis machinery and therefore the comparison of the 16S
rRNA components of ribosomes gives a reliable picture of evolution and classification of
species. Furthermore, 16S rRNA is extracted relatively simple and from every cell in high
amounts. However, the acceptance of archaea as the third domain of life has been
challenging due to the fact that both bacteria and archaea lack a nucleus and the
morphology of both organismic groups appears similar. Despite the minimal morphological
differences, bacteria and archaea do differ in information processing, DNA transcription,
post translation modification of proteins and the assembly of flagella. Many of these
processes are more eukarya-like.

Archaea have been isolated from oceans, soil and unique environmental conditions
such as extreme salt concentrations, high temperatures, extreme pH levels as low as 0.5
or as high as 12. Many of the extreme archaea have been successfully isolated and
cultured in laboratories, however, studies on these cultivated strains have been slow due
to the lack of efficient genetic tools for targeted gene deletion and protein expression
systems. Only recently genetic tools have been developed for three Sulfolobales species S.
solfataricus, S. acidocaldarius and S. islandicus (Albers et al., 2006a, Berkner & Lipps,
2008, She et al., 2009, Wagner et al., 2009, Albers & Driessen, 2007). S. solfataricus was
among the first crenarchaeotes from which the genome was sequenced (She et al., 2001)
and an autotrophy based gene deletion system was developed (Schelert et al., 2004,
remains time consuming and is applicable only in one of the strains, PBL2025. The gene
expression system in S. solfataricus is based on the SSV virus vector (Prangishvili et al.,
2001). The SSV virus transfects S. solfataricus cells and integrates site specifically into its
genome. Newly synthesized virus particles are then secreted into the medium without
seriously damaging the host cell. This harmless transfection mechanism is utilized
effectively for expressing recombinant proteins in S. solfataricus (Albers et al., 2006a).
Chapter 6

Recently, marker-less gene deletion methods and plasmid based expression systems have been developed for *S. acidocaldaius* and *S. islandicus* (She et al., 2009, Wagner et al., 2009). *Sulfolobales* represent the only genetically accessible members of the Crenarchaeota while in the Euryarchaeota genetic systems for different organisms exist. For two other hyperthermophiles, *Pyrococcus furiosus* and *Thermococcus kodakaraensis*, and a few species of the methanogens and halophiles genetic tools have been developed (Sato et al., 2005, Hammelmann & Soppa, 2008, Allers et al., 2004, Thomas et al., 2001, Waege et al., 2010). These recent advancements in genetic tools, together with the availability of the genome sequences of many archaea, has resulted in a massive expansion of the amount of *in silico* data, and has promoted *in vivo* and *in vitro* analysis leading to an advanced understanding of Archaeal biology and their unique biochemistry.

**Protein secretion and assembly of extracellular structures**

Microorganisms secrete proteins for the assembly of their cell wall and extracellular filament structures. Through electron microscopy and a detailed visualization of these microorganisms, long filament and pili structures have been observed that abundantly cover the cell surface. These structures perform crucial tasks such as interaction of microorganisms with their environment including motility, intercellular interaction, colonization of surfaces and the formation of biofilm communities. The advances in genetic and biochemical tools have made it possible to identify and characterize a large number of filament structures and their assembly systems in bacteria and archaea. These can be classified in different surface structure assembly systems.

The assembly and function of the flagellum and UV induced pili (Ups) in *S. solfataricus* has been studied using the gene deletion methods described previously (Szabo et al., 2007a, Froels et al., 2008). The flagellum was demonstrated to be essential for swimming motility of the cells on Gelrite plates and an increased flagellin expression was observed at low carbon source concentration, indicating a strategy of *S. solfataricus* to escape substrate poor environments (Szabo et al., 2007a). The flagella are also required for the transition of cells from the planktonic phase to a surface attached state (Chapter 4). Flagella seem only to be necessary for initial surface attachment as the expression of FlaB is drastically reduced in the surface attached cells (Chapter 4). When subjected to UV-radiation stress, *S. solfataricus* expresses Ups pili on its surface causing the cells to aggregate. The Ups pili possibly facilitate the conjugational exchange of DNA between cells in DNA repair. However, the Ups pili are also required for surface attachment since strains deficient in Ups pili assembly are unable to colonize different surfaces (Chapter 4). Therefore the *ups* system seems to have a dual role in protection of cells against DNA damage by UV stress and the formation of communities.

With the identification of many new surface structure assembly systems in bacteria and archaea, the classification of these systems appears somewhat confusing. The terms
‘pili’, ‘filament’, ‘fibers’, and ‘flagellum’ have often been used interchangeably. In archaea, the distinction between the flagella and pili structures is not entirely clear. The archaeal flagella and pili in many aspects resemble bacterial type IV pili. Moreover, the archaeal flagellum performs many functions similar to the bacterial flagellum but its structure and assembly system is different and no homology between both systems is apparent. In Gram-negative bacteria, the structure and mechanism of assembly of the flagellum is related to type III secretion systems.

In conclusion, the flagellum of \textit{S. solfataricus} evidently has a dual function in cell motility and surface attachment. The flagellum and Ups pili are assembled by homologous systems and they are distantly related to bacterial type IV pili systems (Szabo et al., 2007a, Froels et al., 2008). However, it remains unclear why these two different filament assembly systems are required for surface adhesion and their exact function in this process needs to be investigated.

\textbf{Glycosylation of extracellular structures of \textit{S. solfataricus}.} Attachment of \textit{S. solfataricus} to surfaces is the initial step for the colonization of surfaces and subsequent biofilm formation. Biofilm formation is initiated by the production of extracellular material (EPS) by the attached cells. Fluorescent conjugated lectin labeling demonstrated that the EPS contains glucose, mannose, \(\alpha\)-D-galactose and N-acetyl-D-glucosamine (Zolghadr \textit{et al.}, 2010). The sugar composition of the EPS matched the sugars found in extracellular glycoproteins of \textit{S. solfataricus} and EPS isolated in earlier experiments (Elferink \textit{et al.}, 2001a, Nicolaus \textit{et al.}, 2003). The EPS is possibly required to coat hydrophilic surfaces like mica and glass as on carbon-coated and hydrophobic grids similar EPS structures were absent. EPS coating of smooth and “abiotic” surfaces such as mica may increase the ability cells to adhere to the surface and allow for colonization. Interestingly, more \textit{S. solfataricus} wild type cells were able to adhere to glass by producing less EPS when compared to mica. Possibly this is because the surface of mica is much smoother than glass and therefore the cells need to produce more EPS for colonization of mica. The structure of EPS on mica produced by PBL2025 was more bulky when compared to those structures produced by \textit{S. solfataricus} P2. PBL2025 lacks a large genomic region of 50 genes of which many are predicted to be involved in sugar degradation and sugar transport. Therefore these genes might be involved in the degradation or modification of the secreted EPS. The application of fluorescent conjugated lectins for visualization of extracellular glycosylated structures appears to be useful for analysis of EPS within biofilms of \textit{S. solfataricus}, \textit{S. acidocaldarius} and \textit{S. tokodaii} and recent data indicate that the formation of EPS within the biofilm is dynamic (Koerd and Albers, submitted). This method can be utilized to investigate the enzymes involved in assembly and reorganization of EPS structures.
S-layer protein of *S. solfataricus*

The cell wall of almost all archaea consists of a two dimensional crystalline layer of protein termed the S-layer. The S-layer is highly stable and able to resist many different harsh ecological conditions. In *Sulfolobales*, two surface layer proteins, SlaA and SlaB, form the main components of the S-layer (Veith et al., 2009). SlaA and SlaB have been extracted from *S. solfataricus*, *A. ambivalens* and *M. sedula*, and identified by mass spectrometry. The domain architecture of SlaB was analyzed with molecular modeling and is predicted to contain 2–3 beta sandwich domains followed by a coiled-coil domain (15-17 nm) and a C-terminal transmembrane helix. The primary and secondary structure of SlaB is relatively well conserved among Crenarcheota, whereas SlaA is rather distinct with little sequence homology between the different species. SlaB is inserted in the membrane via its C-terminal transmembrane helix and anchors the S-layer to the membrane. SlaA and SlaB form a mushroom-shaped unit consisting of three SlaB proteins anchoring the complex in the membrane and six SlaA proteins forming the outer crystalline layer. Therefore, an extensive interaction between SlaA and SlaB is predicted. SlaA is arranged into crystalline arrays with triangular and hexagonal pores. Each unit cell is composed of three SlaA dimers forming one triangular pore while the repeated triangular pores form hexagonal pores. In contrast to SlaB, SlaA is poorly conserved among *Sulfolobales*. The identification of SlaA homologs in other archaea by a homology search analysis is not possible and suggests that these proteins have low sequence similarity. Therefore the identification of SlaA proteins from other archaea requires extraction and mass spectrometry analysis as discussed in chapter 5.

**Figure 1.** A schematic representation of the P6 symmetry of *Sulfolobales* S-layer. The SlaA proteins form a strong dimer (A) that is arranged by tetragonal (P3) symmetry (B). The interaction between the trimers creates hexagonal pore-like structures (C).
Although the main component of the S-layer is highly stable, the extraction of the S-layer by detergent possibly removes many biological components associated to the surface layer. The model presented in figure 1, (Veith et al., 2009) only represents the ‘skeleton’ structure of the S-layer formed by the crystalline arrangement of the main component SlaA. The analysis of the cell envelope of *S. solfataricus* by ultra-thin sectioning prepared by high pressure-freezing (Chapter 3, Figure 6) showed that the space in between the S-layer and the cytoplasmic membrane is composed of dense biological materials, likely protein, indicating the presence of many extracellular components.

### The Bindosome

The bindosome is a proposed structure consisting of sugar binding proteins in the cell envelope of *S. solfataricus* (Zolghadr et al., 2007). Bindosome assembly likely leads to the localization of the sugar binding proteins near the external surface of the cells and probably allows for a very efficient scavenging of sugars from the environment. Since the habitat of *Sulfolobus* species is relatively poor in carbon sources, assembly of the bindosome might be beneficial for the growth of these organisms. The suggestion for the existence of the bindosome has been initiated by the observation that a class of sugar binding proteins contain a class III signal peptide similar to those of archaeal flagellins, pilins and bacterial type IV pilins (Albers et al., 1999b). The precursors of these sugar binding proteins are substrates for PibD, the membrane-bound archaeal type IV signal peptidase (Albers et al., 2003). *In vitro* cleavage assays demonstrated that the flagellin FlaB and the glucose binding protein GlcS are both processed by PibD. Subsequently, the bindosome assembly system (*bas*) mediates the assembly of the sugar binding proteins into the bindosome structure. This might correspond to a positioning of the binding proteins in the S-layer. The subunits of the *bas* system are homologous to proteins of bacterial type II secretion and type IV pilin assembly systems. Deletion of *basEF*, the core of the Bas system that encode an ATPase and an integral membrane protein, interferes with growth of *S. solfataricus* on sugars such as glucose that are transported into the cell via binding protein dependent ABC transporters (Chapter 2). In this respect, it is remarkable that the Bas system is absolutely required for the functional localization of some sugar binding proteins but not all binding proteins present in the membrane of *S. solfataricus*.

In chapter 3, the periplasmic space of *S. solfataricus* is shown to be composed of dense material (Chapter 3, Figure 6). The high density of e.g. sugar chains and proteins in this compartment may limit the diffusion of extracellular solutes such as sugars. The binding proteins containing a class III signal sequence tend to form complexes that can be partially purified from the cell surface. Complex formation might be necessary for the correct localization of these binding proteins in the surface layer envelope. One may speculate that
this is close to the S-layer pores, although the exact structure and composition of the bindosome complex remains unclear.

The bas system

The bindosome assembly (bas) system is composed of 5 genes, i.e., \( \text{basEFABC} \) (Zolghadr et al., 2007). The \( \text{basE} \) gene encodes a cytosolic ATPase while the \( \text{basF} \) gene specifies a membrane protein. The purification of tagged BasF lead to the co-isolation of BasE demonstrating that these are interacting proteins (Chapter 3). Extraction of BasEF from the cytoplasmic membrane and solubilisation, however, remained a challenge as the complex was not stable for a longer period of time. Optimizing the conditions of BasEF solubilisation did lead to higher yields of BasEF complex, but the degradation of the complex by proteases remained a major factor of its instability during subsequent purification steps. The size of the BasEF complex and the stochiometry of the Bas components remains unknown. Further optimization of the conditions for expression and purification are essential for future experiments with the BasEF complex. Possibly, a membrane-associated protease is responsible for the degradation of BasEF as this problem was encountered after the membrane isolation stage when the protein complex was solubilized. A search for candidate membrane proteases in \( \text{S. solfataricus} \) and subsequent gene deletion may improve the stability of the complex and enable purification of the complex.

The hydrophobicity profile of the membrane protein BasF is very similar to FlaJ and UpsF and other membrane protein components of archaeal pilus assembly systems. However, these proteins exhibit little sequence identity. The high conservation of the hydrophobicity profile suggests structural homology and a conserved role in pilus assembly. Similar to FlaJ and its homologues, BasF contains two large cytoplasmic domains of about 35 and 15 kDa (Figure 2). The cytoplasmic domains are thought to be the sites of interaction with the ATPase BasE. With the type II secretion system of \( \text{Vibrio cholerae} \), both biochemical and structural evidence has been provided that demonstrates an interaction between EpsE and EspN, the ATPase and the membrane protein components, respectively. The structure of the cytoplasmic loop of EpsN with the N-domain of EpsE suggests that hydrophobic and electrostatic interactions contribute to complex formation (Abendroth et al., 2005). The first cytoplasmic region (Loop 1) is the most conserved among the membrane subunits of the archaeal assembly systems (Figure 3). This region contains several conserved hydrophobic amino acids and amino acids that may be involved in ionic interactions and hydrogen-bond formation. These are likely involved in binding the ATPase subunit. Interestingly, this region also contains several conserved glycine and proline residues that likely fulfill a structural role. The second cytoplasmic region is much less conserved.
Summary and outlook

Figure 2. Predicted membrane topology of BasF and homologs using the programs MTHMM (Krogh et al., 2001) and TopPred (Claros & Heijne, 1994).

<table>
<thead>
<tr>
<th>Protein</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>UpsF</td>
<td>LITQIRKDEQRKLHNEMPVDEYVVSNLGLRNLYKTFEEIRNSK--V</td>
</tr>
<tr>
<td>Hbut1100</td>
<td>SKASSRAERDEEFPEFAAYMTAMAAYGVPFNLERLABIR--M</td>
</tr>
<tr>
<td>Tpen0445</td>
<td>VYPKIKLSRTSDFLEVPFYYHVTATGSGPISPTFPERALAPKVL</td>
</tr>
<tr>
<td>BasF</td>
<td>FFPLTFIFYPEQKYKSRAREIRDDIQDEIPFFVTLITIINASGTTIYEGMRKILQFP--I</td>
</tr>
<tr>
<td>Msed2105</td>
<td>SKMEYLSHLLELFSILLYNLSLGLKGLYETFNDVNSS--L</td>
</tr>
<tr>
<td>Msed1198</td>
<td>AIVLVTYFLPWMRYDIRSTLKVQLEVGLVSLSWGLLSEIGYDVMR114AKETS-E</td>
</tr>
</tbody>
</table>

Figure 3. Amino acid sequence alignment between the cytosolic loop1 regions of the membrane component of archaeal assembly systems. BasF and UpsF are from *S. solfataricus*. Hbut1100 is from conserved hypothetical protein from *Hyperthermus butylicus*, Tpen0445 is a conserved hypothetical protein from *Thermofilum pendens*, Msed2105 and Msed1198 are from *Metallosphaera sedula*.

Type II secretion and type IV pili assembly ATPases homologous to BasE have a similar domain architectures and likely also form hexametric ring complexes (Savvides et al., 2003, Krause et al., 2000). These proteins contain two domains, an N- and C-domain (Yeo et al., 2000, Savvides et al., 2003, Yamagata & Tainer, 2007). The conserved Walker A, Walker B, Asp and His boxes are located in the C-terminal (CTD) domain and they form the ATP binding and hydrolysis site (Possot & Pugsley, 1994, Sandkvist, 2001). Crystal structure analysis and molecular modeling suggested that the C-domain is involved in the hexametric oligomerization required to form the inner ring of the complex while the N-domain forms the outer ring (Yeo et al., 2000, Yamagata & Tainer, 2007). So far, the co-purification of BasE and BasF is the only biochemical evidence for complex formation of core components of archaeal assembly systems. However, the measured size of the BasEF complex is approximately 150 kDa which is smaller than expected. Considering that the size of
monomeric BasE is 60 kDa, a hexamer would be around 360 kDa. The size of monomeric BasF is 75 kDa. The minimal complex size should therefore be around 435 kDa. It is unknown if protein degradation during size-exclusion chromatography or non-specific interaction with the chromatography column cause the dissociation of the expected complex into smaller units.

The basABC genes encode small proteins with a predicted class III signal peptide. Database searches failed to identify other homologous proteins, but by analogy they may resemble pseudo pilin subunits of type II secretion systems. Deletion of basABC genes leads to a defect in cell growth and sugar uptake, but the effect was less severe as with the basEF deletion strain (Zolghadr et al., 2007).

![Phylogenetic relationship of archaeal secretion ATPases those are homologous to BasE. The FlaI and UpsE ATPases form separate subfamilies.](image-url)
Summary and outlook

Figure 5. Examples of gene operons containing the uncharacterized non-FlaI ATPases homologous to BasE. BasE and basF like genes are indicated in red and yellow, respectively. Many of the putative operons include genes encoding for small unknown protein with class III signal peptides and S-layer like proteins similar to SlaB from S. solfataricus. The selected ATPases were from Metallosphaera sedula (B), Methanocorpusculum labreanum (C), uncultured methanogenic archaeon RC-I (D), Methanocaldococcus jannaschii (E), Desulfurococcus kamchatkensis (F) and Halorubrum lacusprofundi (G).

The exact role of BasABC is unclear. Apparently, these proteins are not essential for the correct assembly of the ABC sugar transporter. What are these small proteins, and how unique is the bas system or are there other assembly operons that fulfill roles other than flagellum assembly? To address this question, a preliminary BLAST search for other non-fla-operons was initiated by selecting 26 archaeal secretion ATPases homologous to BasE (supplementary data). Alignment analysis shows that these ATPases are divided into two subfamilies, one containing FlaI homologs and another subfamily containing mostly
Chapter 6

ATPases of unknown function (Figure 4). The latter group contains UpsE from the Sulfolobales that is part of a further subfamily of Ups assembly systems. BasE and UpsE are more closely related than to FlaI. The fla operon is highly conserved among archaea (Introduction, Figure 8). Such clear distinction between FlaI and other uncharacterized ATPases suggests that these proteins are the core component of other assembly operons with functions different from the fla-operon. For a number of assembly ATPases, the flanking genes were examined and compared to the bas system of S. solfataricus (Figure 5). In all cases, the genes encoding the ATPase are accompanied by a gene encoding for a membrane protein homologous to basF. These two proteins form the core of the putative secretion/assembly system. Interestingly, most of the operons contain in additions genes that specify small proteins with a putative class III signal peptide. This analysis does not ascertain that these genes indeed are organized in an operon, but the similarity in gene organization with the BasABCEF cluster suggests that these may be related genes. Importantly, the small protein encoding genes exhibit no significant homology with the basABC genes. Interestingly, one of the clusters contains a gene encoding a protein similar to SlaB of S. solfataricus (Figures 5C, D, and F), FlaK, the archaeal pilin signal peptidase (Figure 5E), or a gene encoding for a putative sugar transport permease (Figure 5F). The function of these systems remains largely unknown but they are likely involved in the assembly of extracellular structures others than the flagellum. This may concerns functions like motility, surface adhesion, assembly and modification of the components of biofilm and/or biogenesis of the S-layer and possible associated proteins. It will be a challenge for future work to determine the function of the gene clusters and to determine if they specify novel surface appendages.